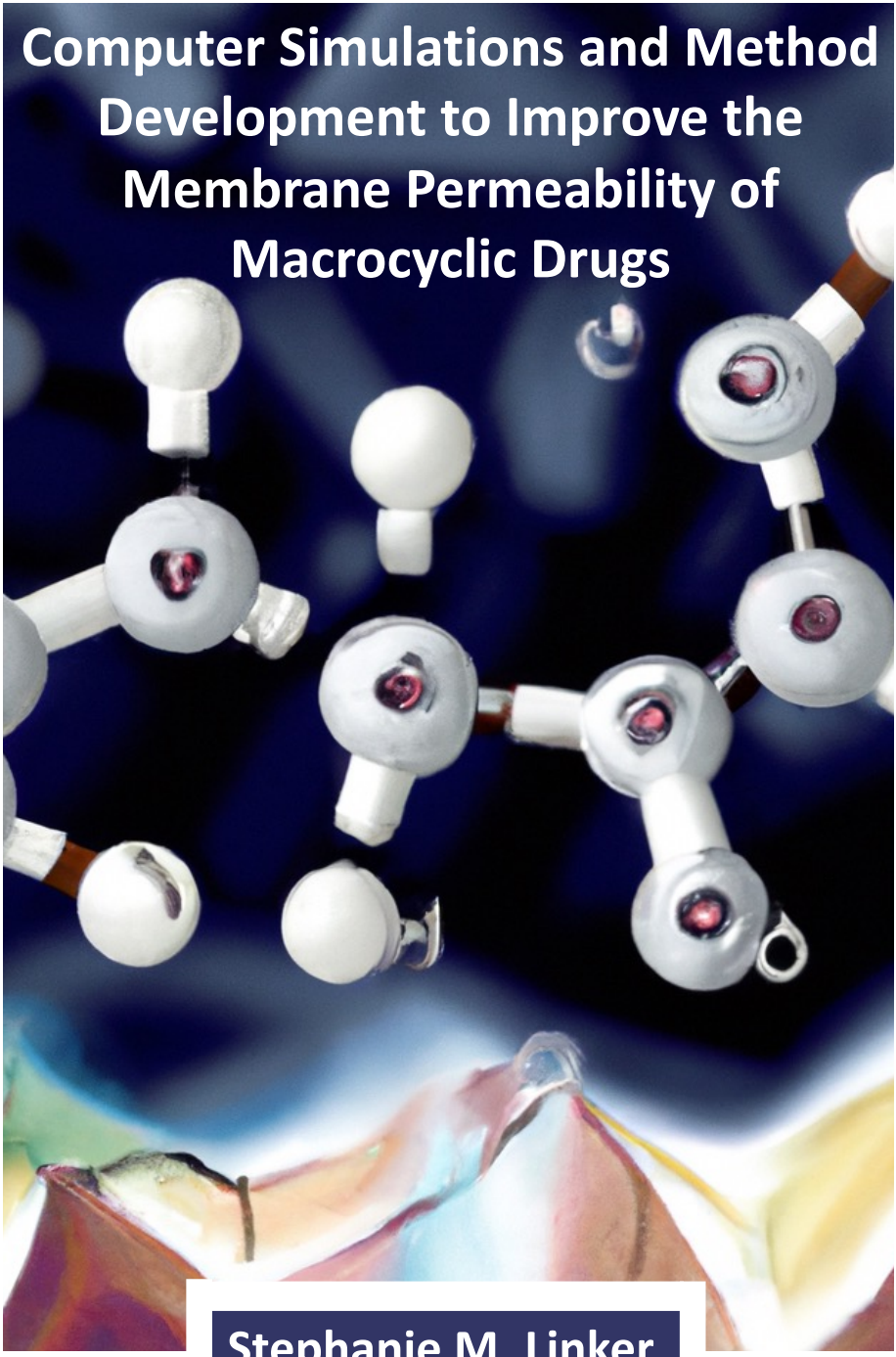
DISS. ETH NO. 28971

Computer Simulations and Method Development to Improve the Membrane Permeability of Macrocyclic Drugs



Stephanie M. Linker

2023

DISS. ETH NO. 28971

Computer Simulations and Method Development to Improve the Membrane Permeability of Macrocyclic Drugs

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

Stephanie Maria Linker

MSc in Biochemistry,
Goethe-University Frankfurt am Main

born on 17.07.1993

accepted on the recommendation of

Prof. Dr. Sereina Riniker, examiner
Prof. Dr. Philippe H. Hünenberger, co-examiner
Prof. DDr. Klaus R. Liedl, co-examiner

2023

For Frederik and Robin

Acknowledgements

I am deeply grateful for the many people who have supported me and my projects over the last few years. This work would not have been possible without you!

A big appraisal goes to my supervisor Prof. Sereina Riniker. Thank you for trusting these fascinating projects and bright students to me. I am deeply thankful for all the support in the bureaucratic jungle we entered for my scholarship recognition and the maternity leave. Thank you for teaching me about concise and well-structured science and improving my ability focus on what matters most. I would like to honor Prof. Philippe Hünenberger for his enthusiasm for straight and clear science and creative method development. I want to thank Claudia Hilty, for the fastest and most competent support one could imaging.

I want to thank the combined groups of computational chemistry (CCG) and Computer Simulation of Molecular Systems (CSMS) for all the great shared memories and our shared passion for science. Specifically, I would like to thank the following people (chronological order): David Hahn, for his bright and outgoing personality and the great collaboration on the Ensember project. Patrick Bleiziffer for insights into QM/MM and for inheriting his plant Uwe. Dominik Sidler for organizing the group ski event and very helpful discussions about physical based reweighting. Gerhard König, for a lot of fun stories, many enjoyable coffee meetings, a brave jump into the Limmat, and for a quick dive into Zurich's cake culture. Sadra Kashef ol Gheta for his calm

attitude and our shared passion for food. R. Gregor Weiss for teaching me how to write accurate and beautiful equations like a 'real' physicist and for sharing his deep view on the world. Carmen Esposito, for a lot of fun we had in WiNS together and all the good vibes and support. Shuzhe Wang for sending me interesting videos about the Chinese food culture and our discussions about working in the pharmaceutical industry. Lennard Böselt for introducing the rowing classes to me, many deep talks about the purpose of life and a phD and dinners to remember. Alzbeta Kubincova for being the best possible companion in the difficult and lonely lockdown period and many beautiful Sunday walks. Benjamin Ries for more thinks than I can express here: for fun collaborations, building up the SCS together, rowing together, being cheerful during difficult times, and many career talks. Marina Pereira Oliviera for always being ready to participate in games and joining group hikes. Albert Hofstetter for many insightful ML questions during the group seminar. Thomas Stadelmann for many discussions about music and being the NMR expert of our group. Anna Albertini for a lot of uplifting talks about parenthood in academia, the best interview coaching and being a genuine and loving 'people person'. Salomé Rieder for being an accurate and very competent colleague. Moritz Thürliman for his dry humour. Marc Lehner for a wonderful weekend at Hasliberg, many fun DnD sessions and his helpful and kind attitude. Candide Champion for long Sunday walks during the lockdown, his bravery for not giving up when facing difficulties, and very competently managing the country machine team. Emilia P. Barros for her cheering character and lots of tasty stories and samples on Brazilian food. Greg Landrum for sharing an office with me and your deep knowledge about chemoinformatics and drug discovery. Jessica Braun for being a programming enthusiast and

her contagious humor. Felix Pultar, I admire how quickly you adjusted to the world of computational chemists. Paul Katzberger for your humorous update on the political status of Austria and the pleasant atmosphere you radiate. Marta Brucker for your curious mind and the interest in Swiss culture we share. Domen Pregeljc for all the insights into the college live in England. I also would like to thank the members of the Richardson group for fun after-work activities and our shared SOLA team.

I want to express my gratitude to the bright students Alix Frischknecht, Christian Schellhaas, Robin Wolf, and Mac Veldhuizen that I supervised. Alix, your work was essential to wrap-up the glycogen project. Christian, you impressed me with your organized, independent, and smart work style. Without you, the project of cyclic peptides at interfaces and membranes would not have been possible. Robin, I especially remember your artistic skills and the fun we had in the student game afternoons. Thanks for the valuable contributions to the Ensembler project. Mac, you have the great ability to spot even small inconsistencies and flaws in a theory and you input helped to improve the project a lot. Keep your curiosity and positivity.

In addition, I want to thank Dr. Daniel Seeliger and Prof. Gerhard Hummer for introducing me into the world of MD simulations and being valuable mentors ever since.

Last but not least, I want to thank Frederik Zwilling, Inga Rückert, Iryna Stasiuk, Leoni Weyrauch, Linda An, Miriam Weber, and Oliver Stach for their continuing support and friendship. I'm sincerely grateful for the support of my family, friends and partner.

Ihre Wort' und Werke
Merkt' ich und den Brauch,
Und mit Geistesstärke
Tu' ich Wunder auch.

---

frei nach
Johann Wolfgang von Goethe

Contents

Acknowledgements *i*

Summary *xi*

Zusammenfassung *xv*

Publications *xix*

1 Introduction 3
- 1.1 General overview 4
- 1.2 Molecular Dynamics Simulations 6
  - 1.2.1 Integration Schemes 7
  - 1.2.2 Simulation Ensembles 8
- 1.3 Markov State Models 10
  - 1.3.1 Mathematical Foundation of Markov State Models 10
  - 1.3.2 Discrete Markov State Processes 15

2 Passing the Barrier - How Computer Simulations Can Help to Understand and Improve the Passive Membrane Permeability of Cyclic Peptides 19
- 2.1 Introduction 21
- 2.2 Computational Approaches to Design Cyclic Peptides 24
  - 2.2.1 Molecular Dynamics Simulations of Cyclic Peptides 25

2.2.2 Kinetic Models of Cyclic Peptides 28
2.2.3 Simulations of Cyclic Peptides at Membranes and Interfaces 30
2.3 Conclusion 31

3 Polar/Apolar Interfaces Modulate the Conformational Behavior of Cyclic Peptides with Impact on Their Passive Membrane Permeability 35
3.1 Introduction 39
3.2 Methods 43
3.2.1 MD Simulations 44
3.3 Results and Discussion 50
3.3.1 3D Hydrophobicity Profile 50
3.3.2 Peptide Orientations at Interfaces 54
3.3.3 Composition Effects on the Orientation Preference at the Interface 58
3.3.4 Comparison with Lipid Membranes 61
3.3.5 Interface Catalyzes the Closing of Cyclic Peptides 67
3.3.6 Permeable Conformations 84
3.4 Conclusion 87

4 Lessons for Oral Bioavailability: How Conformationally Flexible Cyclic Peptides Enter and Cross Lipid Membranes 93
4.1 Introduction 95
4.2 Methods 99
4.2.1 MD Simulations 99
4.2.2 Biased simulations 100
4.2.3 Data Analysis 102
4.2.4 Biasing Simulation for the Phenylalanine Lock 103

4.2.5 Distribution of Head-Group Gap Size . . . 103
4.2.6 Construction of Markov State Models . . . 104
4.2.7 Data and Software Availability 105
4.3 Results and Discussion 105
4.3.1 Cyclic Peptides Enter Lipid Membranes Using 'Anchor' Residues 108
4.3.2 Effect of Membrane Composition on Head-Group Gaps 116
4.3.3 Cyclic Peptides Occupy Two Distinct Orientations at Lipid Membranes 118
4.3.4 Amino-Acid Composition Influences the Preferred Peptide Orientation 124
4.3.5 'open' Cyclic Peptides Can Close Inside the Lipid Membrane 128
4.3.6 Phenylalanine Residues Can Act as a 'Lock' That Prevents Closing 133
4.3.7 Crossing From the Upper to the Lower Leaflet Requires Anchoring and Flipping 136
4.3.8 Four Steps of Membrane Permeation 141
4.4 Conclusion 145

5 Connecting Dynamic Reweighting Algorithms:Derivation of the Dynamic Reweighting Family Tree 149
5.1 Introduction 151
5.2 Theory 154
5.2.1 Markov State Models 154
5.2.2 DHAM Reweighting 155
5.2.3 Girsanov Reweighting 156
5.2.4 Weber-Pande Reweighting 159
5.3 Connection between Weber-Pande and Girsanov Reweighting 162

5.4 Differences between Weber-Pande and Girsanov Reweighting . . . 165
5.5 An integrator independent reweighting scheme and the connection to DHAM . . . 169
5.6 Methods . . . 173
5.6.1 Data and Software Availability . . . 175
5.7 Results and Discussion . . . 175
5.7.1 Transition Matrix Comparison Between Weber-Pande and Girsanov Reweighting . . . 175
5.7.2 Reweighting Results for Weber-Pande, Girsanov, and DHAM . . . 178
5.7.3 Path Correction Expands DHAM to Long Lagtimes . . . 181
5.8 Conclusion and Outlook . . . 186

6 Revisiting the Kinetics of Cyclosporin A 191
6.1 Introduction . . . 193
6.2 Computational Details . . . 196
6.2.1 Starting Conformations . . . 196
6.2.2 MD Simulations . . . 197
6.2.3 Markov State Model and Analysis . . . 198
6.2.4 Dynamic Reweighting . . . 199
6.3 Results and Discussion . . . 200
6.4 Conclusion . . . 222

7 Outlook 229
7.1 Design of Permeable Cyclic Peptides . . . 230
7.2 Application of Dynamic Reweighting Methods and Kinetic Modes . . . 233

Abbreviations 235

References 270

Summary

Macrocyclic drugs are able to bind to flat binding sides and modulate protein-protein interactions. Thus, they have the potential to lead to new therapeutics for currently non-druggable diseases. However, their application is limited by their often poor solubility and membrane permeability.

In this PhD thesis, I aim to shed light on the permeation process of large and flexible cyclic peptides and their internal conformational dynamics. This knowlege is used to the development of new design strategies to increase the membrane permeability of macrocyclic drugs. To gain atomistic and kinetic insights into the permeation processes, we use molecular dynamics simulations in combinations with Markov state modeling. A major challenge for studying the permeability process at lipid membranes are the slow kinetics dominating these processes, termed rare events. Despite their importance, it is often not feasible to computationally investigate such rare events, as their typical time scale exceeds the accessible simulation time. In this thesis, I present two approaches to address this problem. The first approach aims to mimic the membrane permeation event using a simpler polar/apolar interface system. Reducing the system size and complexity decreases the computational costs. The second approach focuses on enhanced sampling and dynamic reweighting techniques. Enhanced sampling methods speed up rare events by applying a bias to the system. Dynamic reweighting is afterwards used to recover the underlying equilibrium and kinetic properties.

In Chapters 1 - 7 of this thesis, I report our advances in understanding the structure-permeation relationship of macrocyclic drugs applying the strategies outlined above. In addition, I describe the unbiased simulation results on cyclic peptides at lipid membranes. These results shed light on the permeation pathway of flexible cyclic peptides and open up new design strategies for permeable cyclic peptides.

Chapter 1 gives an introduction into the simulations techniques fundamental for this thesis.

Chapter 2 provides a comprehensive review on the state of the art and current challenges for macrocyclic drug discovery. A focus is set on how the combined power of computer simulations and experimental techniques has led to new insights into the rational design for permeable macrocycles.

In Chapter 3, we investigate the conformational and kinetic behavior of cyclic decapeptides at a polar/apolar interface. The interface mimics the unique environment inside the lipid membrane at the head group and tail group region. We discovered that in this unique interface environment the peptides can position themselves in two different orientation with different dynamic properties. One of these orientations facilitates the adoption of the permeable 'closed' conformation.

Continuing on the results of Chapter 3, in Chapter 4 we track the pathway of cyclic decapeptides through a lipid membrane. We analysed how the cyclic peptides establish the first contact with the membrane, and were able to collect closing and leaflet crossing events. The discovered pathway highlights the role of molecular anchors for the permeability of cyclic peptides.

In order to speed up the occurrence of rare events, one can use enhanced sampling methods. However, enhanced sampling comes at the cost of losing the underlying kinetics. Thus, in Chap-

ter 5 we describe the current approaches in dynamic reweighting that were developed to recover the system's kinetics. Using the underlying physical principles, we derived the connection between the existing reweighting methods and suggested an improved reweighting approach.

In Chapter 6, we revisit the conformational dynamics of cyclosporin A and apply dynamic reweighting techniques. By improving and correcting previous kinetic models for cyclosporin A, we gain new insights into its permeable states and highlight challenges and pitfalls associated with kinetic models. Our results show that a vast amount of simulations are necessary to sample the rare cis/trans isomerization in the peptide backbone of cyclosporin A. As an alternative to this massive sampling, we show how enhanced sampling and dynamic reweighting can be applied to this system. We compare the reweighted equilibrium and kinetic information as well as the model's robustness to our massive-sampling reference. In addition, we give guidelines on best practices and highlight obstacles for enhanced sampling and dynamic reweighting.

In the final chapter (Chapter 7), an outlook is given on the implications of the results covered in this thesis, and future directions of macrocyclic drug discovery and rare event kinetics are discussed.

Zusammenfassung

Makrozyklische Wirkstoffe sind in der Lage, an flache Bindungsstellen zu binden sowie Protein-Protein-Wechselwirkungen zu modulieren. Daher haben sie das Potenzial zu neuen Therapeutika für derzeit nicht heilbare Krankheiten zu führen. Ihre oft schlechte Löslichkeit und Membranpermeabilität begrenzt jedoch noch ihre Anwendbarkeit.

Ziel dieser Doktorarbeit ist, den Permeationsprozess sowie die interne Konformationsdynamik grosser und flexibler zyklischer Peptide zu beleuchten. Dies soll zur Entwicklung neuer Designstrategien zur Erhöhung der Membranpermeabilität makrozyklischer Wirkstoffe beitragen. Um atomistische und kinetische Einblicke in den Permeationsprozess zu gewinnen, verwenden wir Molekulardynamiksimulationen in Kombination mit Markov-Zustandsmodellierung. Eine grosse Herausforderung für die Untersuchung des Permeabilitätsprozesses durch Lipidmembranen sind die langsamen Kinetiken, welche diese Prozesse dominieren und als seltene Dynamiken bezeichnet werden. Trotz ihrer grossen Bedeutung für den Permeationsprozess ist es oft nicht möglich, diese seltenen Ereignisse rechnerisch zu untersuchen, da ihre typische Zeitskala die verfügbare Simulationszeit übersteigt. In dieser Arbeit stelle ich zwei Ansätze vor, um diese Problematik anzugehen. Der erste Ansatz zielt darauf ab, den Membranpermeationsprozess mit einem einfacheren polaren/apolaren Grenzflächensystem nachzuahmen. Das Reduzieren der Grösse und Komplexität des Systems verringert seinen Rechenbedarf. Der zweite Ansatz konzentriert

sich auf *enhanced sampling* und *dynamic reweighting* Techniken. *Enhanced sampling* Methoden beschleunigen seltene Ereignisse, indem sie einen Bias auf das System anwenden. Anschliessend wird *dynamic reweighting* verwendet, um das zugrunde liegende Gleichgewicht und die kinetischen Eigenschaften wiederherzustellen.

In den Kapiteln 1 - 7 dieser Dissertation berichte ich über unsere Fortschritte beim Verständnis der Struktur-Permeations-Beziehung von makrozyklischen Arzneimitteln, unter Anwendung dieser Strategien. Zusätzlich berichte ich über unsere *unbiased* Simulationsergebnisse zu zyklischen Peptiden an Lipidmembranen. Diese Ergebnisse geben Aufschluss über den Permeationsweg dieser Moleküle und eröffnen neue Designstrategien für permeable zyklische Peptide.

Kapitel 1 enthält eine Einführung in die Simulationsmethoden, die für diese Arbeit relevant sind.

Kapitel 2 enthält eine umfassende Übersicht über den aktuellen Stand und die Hürden für die Entwicklung makrozyklischer Wirkstoffe. Hierbei wird der Fokus darauf gelegt wie Computersimulationen und experimentelle Techniken sich ergänzen können, um neue Erkenntnisse im rationalen Design von permeablen Makrozyklen zu erlangen.

In Kapitel 3 untersuche ich das konformelle und kinetische Verhalten von zyklischen Dekapeptiden an einer polaren/apolaren Grenzschicht. Eine solche Grenzschicht imitiert die polare/apolare Grenzumgebung, die in der Lipidmembran zwischen den polaren Kopfgruppen und den apolaren Kohlenwasserstoffketten der Lipide herrscht. Diese Untersuchungen zeigen, dass diese besondere Umgebung zu einem Auftrennen in zwei Orientierungen führt und das Einnehmen der permeablen geschlossenen Struktur fördern kann.

Basierend auf den Ergebnissen von Kapitel 3 analysiert Kapitel 4 den Weg eines zyklischen Dekapeptids durch eine Lipidmem-

bran. Die Bildung des ersten Kontakts zwischen zyklischen Peptid und Membran sowie die konformellen Schliessungen und Öffnungen werden beschrieben. Basierend auf dem beschriebenen Permeabilitätsweg können neue Empfehlungen für das rationale Design von zyklischen Peptiden abgeleitet werden.

Sogenannte *enhanced sampling* Methoden erlauben es, langsame Dynamiken zu beschleunigen. Als Preis für diese Beschleunigung werden jedoch die Kinetiken des Systems verzerrt. Um die originalen Kinetiken zurückzuerlangen, wurden *dynamic reweighting* Methoden entwickelt. Kapitel 5 beschreibt zunächst bestehende Ansätze von *dynamic reweighting*. Die dabei gefundenen physikalischen Prinzipien und Verbindungen wurden anschliessend genutzt, um einen verbesserten *dynamic reweighting* Ansatz zu formulieren.

In Kapitel 6 re-analysiere ich die Konformationsdynamik von Cyclosporin A und zeigen die Anwendung von *dynamic reweighting*. Durch die Verbesserung und Korrektur früherer kinetischer Modelle für Cyclosporin A werden neue Einblicke in seine permeablen Zustände gewonnen sowie Herausforderungen und Fallstricke im Zusammenhang mit kinetischen Modellen hervorgehoben. Diese Ergebnisse zeigen, dass eine grosse Menge an Simulationen notwendig ist, um die seltene cis/trans-Isomerisierung in der Peptid-Hauptkette von Cyclosporin A zu untersuchen. Als Alternative zu diesem massiven Sampling zeige ich, wie *enhanced sampling* und *dynamic reweighting* auf dieses System angewendet werden können. Ich vergleiche die *reweighted* Gleichgewichts- und kinetisch-en Informationen sowie die Robustheit des Modells mit unserer Referenz. Darüber hinaus gebe ich Richtlinien zu Best Practices und heben Hindernisse für *enhanced sampling* und *dynamic reweighting* hervor.

Das finale Kapitel (Kapitel 7) enthält einen Ausblick auf die

Bedeutung der in dieser Dissertation vorgestellten Resultate für die Entwicklung makrozyklischer Wirkstoffe und die computergestützte Analyse seltener Ereignisse.

Publications

Articles in covered in this thesis:

1. S. M. Linker†, C. Schellhaas†, A. S. Kamenik, M. M. Veldhuizen, F. Waibl, H. Roth, M. Fouché, S. Rodde, S. Riniker, *Lessons for Oral Bioavailability: How Conformationally Flexible Cyclic Peptides Enter and Cross Lipid Membranes*, J. Med. Chem., 66 (2023) 2773–-2788.

2. S. M. Linker†, C. Schellhaas†, B. Ries, H. Roth, M. Fouché, S. Rodde, S. Riniker, *Polar/Apolar Interfaces Modulate the Conformational Behavior of Cyclic Peptides With Impact on Their Passive Membrane Permeability*, RSC Adv., 12 (2022) 5782–5796.

3. S. M. Linker, S. Wang, B. Ries, T. Stadelmann, S. Riniker, *Passing the Barrier – How Computer Simulations Can Help to Understand and Improve the Passive Membrane Permeability of Cyclic Peptides*, CHIMIA, 75 (2021) 518–521.

4. S. M. Linker, R. G. Weiß, S. Riniker, *Connecting Dynamic Reweighting Algorithms: Derivation of the Dynamic Reweighting Family Tree*, J. Chem. Phys., 153 (2020) 234106.

† These authors contributed equally.

Additional publications:

1. A. S. Kamenik, S. M. Linker, S. Riniker, *Matching Simulations and Experiments of Conformationally Flexible Cyclic Peptides: Steps Towards a Holistic Picture of Passive Membrane Permeability*, ACS Symp. Ser., 1417, 5 (2022) 137–154 .

2. A. S. Kamenik, S. M. Linker, S. Riniker, *Enhanced Sampling Without Borders: On Global Biasing Functions and How to Reweight Them*, Phys. Chem. Chem. Phys., 24 (2022) 1225–1236.

3. B. Ries, S. M. Linker, D. F. Hahn, G. König, S. Riniker, *Ensembler: A Simple Package for Fast Prototyping and Teaching Molecular Simulations*, J. Chem. Inf. Model., 61 (2021) 560–564.

4. J. Liu, V. Hürlimann, R. Emter, A. Natsch, C. Esposito, S. M. Linker, Y. Zou, L. Zhou, Q. Wang, S. Riniker, P. Kraft, *A New Family of Rigid Dienone Musks Challenges the Perceptive Range of the Human Olfactory Receptor OR5AN1*, Synlett, 10 (2020) 972–976.

5. K. Sekar[†], S. M. Linker[†], J. Nguyen, A. Grünhagen, R. Stocker, U. Sauer, *Bacterial Glycogen Provides Short-Term Benefits in Changing Environments*, Appl. Eviron. Microbiol., 86, (2020) e00049-20.

[†] These authors contributed equally.

1 Introduction

„Everything should be made as simple as possible, but not simpler“

Albert Einstein

This quote and chapter are dedicated to Carmen Esposito and her gift for well-structured and concise science

1.1 GENERAL OVERVIEW

Computer simulations are widely used as a "super microscope" to analyse the mode of action of biological systems.[1] Molecular dynamics (MD) simulations simulate molecular motion on the atomistic or coarse grained level using Newtonian physics. MD simulations are therefore especially valuable for studying the kinetics and energy states of biological and chemical systems. However, MD simulations are computationally demanding and currently limited to the microsecond time scale in routine applications, even on specialized computer clusters. In contrast, the mode of action of macromolecules like proteins is often dominated by slow, rare-event kinetics, like the dissociation of a ligand from the binding pocket of a protein.[2] Here, unbinding is rare and most simulation time is spent without simulating the desired process.[3] Therefore, the waiting time for rare events of biologically and chemically relevant systems is often not computationally feasible.

To overcome the limitations of slow kinetics, enhanced sampling algorithms were developed.[4] Enhanced sampling simulations add a biasing potential to the system, with the goal to accelerate crossing of energy barriers.[5] Biasing leads to faster sampling, at the cost of losing the information on the true kinetics. Only recently, approaches based on transition-path sampling, Markov models, or path ensemble averages were proposed to reweight such biased simulations in order to recover the correct kinetics.[6–8] The basic idea behind these methods is to find the reaction coordinate of the slowest processes, accelerate the transition by applying a bias, and afterwards reweight the simulation to estimate the unbiased kinetics. The reaction coordinate is a low-dimensional molecular observable and chosen to be an good

representation of the free-energy landscape and the corresponding highest barriers.[9,10] The highest energy barriers correspond to rare events and define the kinetics of the system. Knowledge of the reaction coordinate is crucial to reduce the energy barriers of the system in enhanced-sampling techniques. However, especially in complex biological systems, it is challenging to find a good representation of the reaction coordinate.

The combination of finding the reaction coordinate, enhanced sampling, and reweighting is a valuable tool to enable research on slow biological systems and further development of these methodologies is crucial for the field of computational biology. In this thesis, we apply unbiased simulations as well as enhanced sampling and reweighting to cyclic peptides in order to rationalize their membrane permeability process. Cyclic peptides display complex conformational dynamics and due to their peptidic origin mimic the behavior of proteins. Thus, they constitute an insightful test system to study the applicability and potential of enhanced sampling and reweighting techniques.

Cyclic peptides are a class of drug candidates, which can bind onto large and flat binding sites with high affinity. Therefore, they are possible new therapeutics for difficult targets such as protein-protein interfaces and class B G-protein coupled receptors (GPCRs) that are often considered to be undruggable.[11,12] Although the first cyclic peptide drugs – like the immunosuppressive cyclosporin A (CycA)[13–15] – are on the market, more insights into the mode of action and pharmacokinetics of cyclic peptides are needed. Especially the bioavailability is of great importance but difficult to achieve with this class of compounds. In order to reach intracellular targets, cyclic peptides need to cross the cell membrane barrier. Some cyclic peptides, like CycA, are able to passively diffuse through the membrane.[16] However, most cyclic

peptides are non-permeable. New insights into the dynamics of cyclic peptides will improve the efficiency of drug discovery and lead to the development of new therapeutics for currently undruggable targets.

MD simulations and Markov state models (MSMs) build the foundation of the studies and the reweighting techniques presented in this thesis. Thus, a short introduction in these topics are provided in the next sections.

1.2 MOLECULAR DYNAMICS SIMULATIONS

MD simulations describe molecules and their dynamics on an atomistic or coarse grained level. An empirical force field is used to evaluate the interactions of the simulated particles.[17–19] The type of force field determines the level physical accuracy. A force field that treats the particles on the level of classical mechanics can be used for simulations that reproduce conformational changes and dynamics, but is not applicable to study chemical reactions. The force acting on each particle can be calculated as the negative gradient of the potential energy defined by the force field.[20] To obtain a dynamic trajectory of the particle, Newton's equations of motion are integrated numerically with femtosecond time steps (Section 1.2.1). The time scales accessible to protein simulations have grown exponentially over the past decades. This advance was mainly achieved by efficient parallelization of MD software codes, better hardware, and sophisticated analysis of multiple independent trajectories.[21] However, the time scales of macrocyclic membrane permeation are often

still beyond the practical scope of MD simulations. Thus, enhanced sampling techniques, as described above, are important to achieve sufficient sampling of the relevant states. Another powerful method tackling the challenge of reaching relevant time scales that has evolved in the last decade is the combination of MD simulations and MSMs (Section 1.3).[22–24] MSMs can be used to achieve a speed-up by using embarrassingly parallel simulations and a statistical framework. This framework allows to construct a kinetic model from trajectories that are an order of magnitude shorter than the longest relaxation times of the system and thus well describes the systems' long-time dynamics.[25] Below we will give a short introduction into the integration schemes and ensembles used in MD simulations and the mathematical framework of MSMs.

1.2.1 Integration Schemes

In order to attain the dynamics of a system from its force field, one has to integrate the equations of motion. In practice, this is done using numerical integration and several integration methods with unique advantages and disadvantages were developed.

If the goal of the computation is to find the next local minima of the system on the potential-energy surface, optimization algorithms such as the steepest descent[26] or conjugated gradient[27] are the tool of choice. These algorithms are also widely used to minimize the system energy and relax the coordinates prior to a MD simulation.[20] If instead the goal is to sample the states of the system without the need for dynamic information, one can use stochastic approaches like the Monte-Carlo approach or the Metropolis-Hastings integrator.[28] Both approaches are based on the Metropolis-Monte Carlo criterion,[29] which assures that the

system is sampled according to its Boltzmann distribution. If both the state distribution and the dynamic information of the system are of interest, one should consider integration schemes based on Newton's equations of motion.[20,30] These integration schemes are usually subdivided based on their order of accuracy. MD simulations for biological systems often use second-order algorithms like the Verlet[31] and leap-frog algorithms.[32] Higher accuracy can for example be obtained using the Runge-Kutta algorithm,[33,34] which is often used to calculate spacecraft trajectories, but the multiple force evaluations per time step limit its application in molecular simulations. As an example, the leap-frog algorithm calculates the coordinates $\mathbf{r}_i$ of atom i at time $t + \Delta t$ and its velocities $\mathbf{v}_i$ at time point $t + \frac{1}{2}\Delta t$ using,

$$\begin{aligned} \mathbf{r}_i(t + \Delta t) &= \mathbf{r}_i(t) + \mathbf{v}_i(t + \frac{1}{2}\Delta t)\Delta t \\ \mathbf{v}_i(t + \frac{1}{2}\Delta t) &= \mathbf{v}_i(t - \frac{1}{2}\Delta t) + \mathbf{a}_i(t)\Delta t. \end{aligned} \tag{1.1}$$

Here, $\mathbf{a}_i(t)$ is the acceleration of on atom i calculated by the atomistic force $\mathbf{F}_i$, which in return is derived from the potential energy using,[35]

$$-\frac{\partial V(\mathbf{r}(t))}{\partial \mathbf{r}_i(t)} = \mathbf{F}_i(\mathbf{r}(t)) = m_i \mathbf{a}_i(t). \tag{1.2}$$

1.2.2 Simulation Ensembles

Thermodynamic ensembles describe how a large quantity of particles behaves under specific thermodynamic conditions in equilibrium. One generally discriminates microcanonical, canonical, grand canonical, and isobaric-isothermal ensembles. In the microcanonical (NVE) ensemble, the number of particles (N), the

system box volume (V), and the total energy of the system are fixed.[36] Thus, this ensemble describes a system that is completely isolated from its surroundings and can neither exchange particles nor thermal energy nor work. In contrast, the canonical (NVT) ensemble can exchange thermal energy with its surroundings. In this system, the number of particles (N), the system box volume (V), and the temperature (T) are fixed. The grand canonical (μVT) describes a system that can exchange particles with its surroundings but has a fixed chemical potential, volume, and temperature. The isobaric-isothermal (NPT) ensemble comes closest to the conditions applicable to a lab experiment. The number of particles, pressure, and temperature are kept constant.[36]

In order to keep the temperature or the pressure constant during a MD simulation, one has to apply so-called thermostats and barostats. Various algorithms are available for this purpose. First introduced by Berendsen[37] in 1984, the weak-coupling approach ranks among the oldest thermostats for MD simulations.[37] The weak-coupling thermostat scales the velocities by a factor $\lambda(T(t); \tau_T, \Delta t, T_0)$ that depends on the current temperature $T(t)$, the reference temperature T_0, a coupling constant τ_T, and the time step of the simulation Δt in order to maintain a constant temperature. The velocities are scaled according to,[37]

$$\lambda(T; \tau_T, \Delta t, T_0) = \sqrt{1 + \frac{\Delta t}{\tau_T}\left(\frac{T_0}{T(t)} - 1\right)}. \tag{1.3}$$

Other, more sophisticated approaches are the Nosé-Hoover[38–40] and Nosé-Hoover chain[41] thermostats. Typical algorithms to maintain a constant pressure are the weak-coupling barostat,[37] the Parinello-Rahman barostat,[42] Andersen barotstat,[43] or the Nosé-Hoover barostat.[44]

1.3 MARKOV STATE MODELS

In their original form, MSMs are based on unbiased MD simulations. Therefore, they are capable of describing realistic system kinetics.[23] However, as described above, using new reweighting techniques, MSMs can be combined with biasing and accelerating techniques,[6,45,46] e.g. replica exchange[47,48] or metadynamics/local elevation.[49,50] MSMs are able to describe the complete system dynamics based on independent short simulations that are incorporated in a statistical framework. Thus, a big advantage of MSMs is their embarrassingly parallel nature. In addition, the time scales covered by an MSM can be much longer that the time scales of the individual trajectories used to build the model.[22,51–53] Guided by knowledge – gained by experiments or previous simulations – one can also find an optimal distribution of simulation starting points that reduces sampling of uninteresting or already witnessed states. This process is called adaptive sampling.[54–56] Moreover, MSMs can simplify the simulation analysis by discarding very fast dynamics below the so-called lag time.[21] With choosing the lag time, it is possible to scale the model to multiple levels of complexity – from quantitatively accurate (high resolution) to human understandable (low resolution). The mathematical foundation of MSMs presented in the subsequent chapter is based on Refs. 22,53,57,58.

1.3.1 Mathematical Foundation of Markov State Models

MSMs can provide kinetic models on experimentally relevant time scales and statistical significance.[22] To explore the mathematical

foundation of MSMs, we consider a system in a state space Ω. The state of the system at time t is called $\mathbf{x}(\mathbf{t})$. In order to build a MSM, the system is supposed to have the following properties:

- Markovianity: The system has no memory. Thus, the state at time $\mathbf{x}(t + \Delta t)$ only depends on $\mathbf{x}(t)$ and the transition probability. States previous to t do not influence the outcome of $\mathbf{x}(t + \Delta t)$. The transition probability density $p(\mathbf{x}, \mathbf{y}; t)$ of going from point $\mathbf{x}$ to point $\mathbf{y}$ in the time τ is defined by

$$p_\tau(\mathbf{x}, \mathbf{y}; t) = P[\mathbf{x}(t + \tau) \in \mathbf{y} | \mathbf{x}(t) \in \mathbf{x}], \tag{1.4}$$

  with $\mathbf{x}, \mathbf{y} \in \Omega$.

- Ergodicity: The state space Ω is not subdivided in two or more dynamically disconnected sets. Therefore, for $t \to \infty$, the system will visit each state $\mathbf{x} \in \Omega$. In that limit, the time spent in state $\mathbf{x}$ is relative to its stationary distribution. Thus, the time average and the phase-space average of a function $f(\mathbf{x}(t))$ become equal, i.e.

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T f(\mathbf{x}(t)) dt = \int_\Omega f(\mathbf{x}) \mu(\mathbf{x}) d\mathbf{x}. \tag{1.5}$$

  Here, $\mu(\mathbf{x})$ is the stationary density distribution.

- Detailed balance: To fulfill detailed balance, the amount of transitions from state $\mathbf{x}$ to $\mathbf{y}$ need to equal the amount of transitions from state $\mathbf{y}$ to $\mathbf{x}$. Thus,

$$\mu(\mathbf{x}) p_\tau(\mathbf{x}, \mathbf{y}; t) = \mu(\mathbf{y}) p_\tau(\mathbf{y}, \mathbf{x}; t). \tag{1.6}$$

  Here, $\mu(\mathbf{x})$ is the stationary density distribution. On a phys-

ical level, detailed balance is justified by the following argument: If detailed balance is not fulfilled, there exists a state space loop, where more transitions happen in one direction then in the other. Using the preferred direction of this state space loop, work could be produced. However, a system in equilibrium is driven only by thermal energy. Thus, this would violate the second law of thermodynamics by converting pure thermal energy to work.

Suppose $\rho_t(\mathbf{x})$ is the density distribution at time t. Based on the transition probability density $p_\tau(\mathbf{x}, \mathbf{y}; t)$, this density distribution is changing over time. Assuming a continuous phase space, one could write

$$\rho_{t+\tau}(\mathbf{x}) = L(\tau) \circ \rho_t(\mathbf{x}), \tag{1.7}$$

with the continuous transport operator $L(\tau)$. The operator $L(\tau)$ describes the transport of densities and fulfills the Chapman-Kolmogorov equation:

$$\rho_{t+k*\tau}(\mathbf{x}) = [L(\tau)]^k \circ \rho_t(\mathbf{x}), \tag{1.8}$$

with k being a natural number. An important property of the stationary density distribution $\mu(\mathbf{x})$ is that applying $L(\tau)$ to the stationary density distribution will again yield the stationary density distribution, i.e.

$$\mu(\mathbf{x}) = L(\tau) \circ \mu(\mathbf{x}). \tag{1.9}$$

This is equivalent to an eigenvalue equation with an eigenvalue $\lambda_i = 1$ and an eigenvector $\mu(\mathbf{x})$. Due to ergodicity, there exists only one solution to the eigenvalue equation with the eigenvalue $\lambda_i = 1$ and the corresponding eigenvector is the stationary

density distribution. This can be validated using the following reasoning: If a second eigenvalue $\lambda_j = 1$ exists, then two dynamically disconnected subsets in Ω exist. This, by definition, contradicts the concept of ergodicity. By extracting the transport operator $L(\tau)$ from the MD simulation trajectory, it is thus possible to extract the corresponding equilibrium distribution by solving the eigenvalue equation. Due to detailed balance, all eigenvalues of $L(\tau)$ are real values and smaller or equal one.[59,60] The eigenvalue equation of $L(\tau)$ can exhibits additional solutions with eigenvalues $\lambda_j < 1$. These eigenvalues correspond to processes that decay over time. The smaller the eigenvalue, the faster is the decay. The time scale of the decay can be measured using the implied time scale t. The i^{th}-implied time scale is defined as

$$t_i = \frac{-\tau}{\ln \lambda_i}. \tag{1.10}$$

Figure 1.1 shows the generation of a MSM on a didactic one-dimensional energy landscape. The energy landscape is shown in Figure 1.1a. The energy landscape is characterized by four minima and a large energy barrier at the center. Panel 1.1 (b) displays the metastable eigenfunctions of $L(\tau)$. The eigenfunction of the first eigenvector corresponds to the stationary density distribution and only contains positive values. For the other eigenfunctions, interconverting states have opposite signs. Panel (c) displays the eigenvalue spectrum of the continuous transport operator $L(\tau)$. The eigenvalue spectrum can be used to distinguish the metastables processes (colored) and fast decaying processes. Those two classes are separated by a gap in the eigenvalue spectrum.

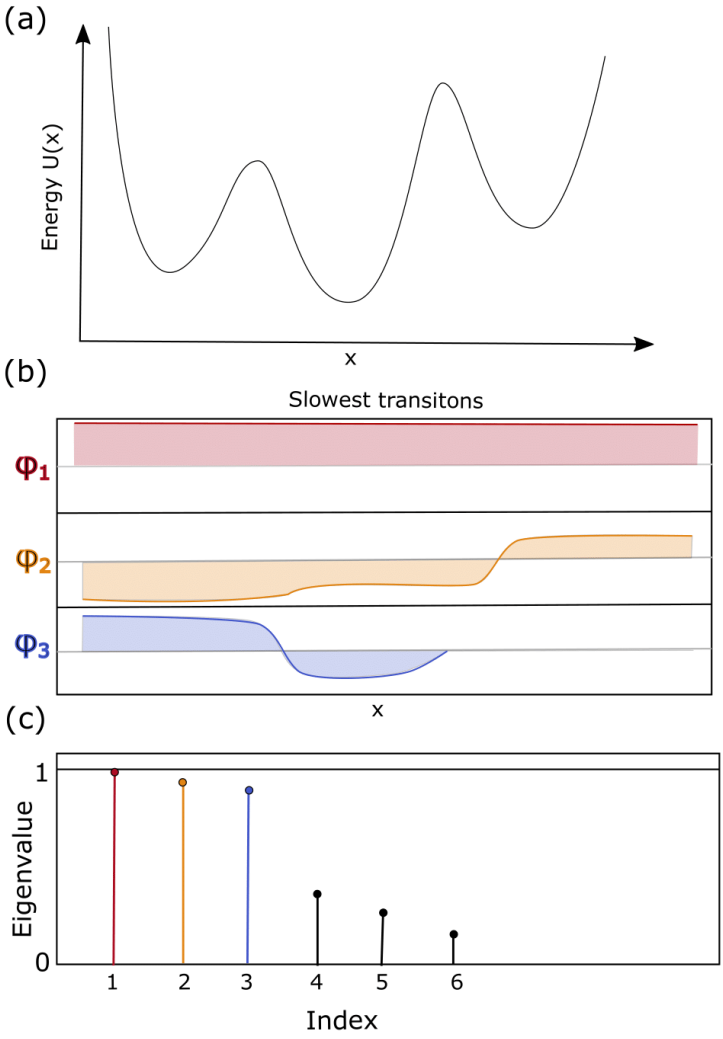


FIGURE 1.1: Schematic illustration of a one-dimensional Markov state model (MSM), adjusted from Ref. 23. (a): One-dimensional energy landscape with four distinct energy minima. (b): Eigenfunctions of the four slowest decaying processes. Interconverting states have opposite signs. (c): Sorted eigenvalue spectrum of $L(\tau)$.

1.3.2 Discrete Markov State Processes

In order to calculate the stationary density distribution of continuous processes, the continuous transport operator $L(\tau)$ is needed. This operator can be estimated from the MD simulations trajectories. However, in practice, the state space Ω has to be discretized to obtain a computational manageable description of $L(\tau)$. The discretization process divides the phase space into n micro-states S. Traditionally, these states are defined in such a way that they do not overlap, i.e.

$$\bigcup_{i=1}^{n} S_i = \Omega \qquad and \qquad S_i \cap S_j = \delta_{ij} S_i. \tag{1.11}$$

The discretization introduces an error.[58] Due to the loss of the exact locality of the state, the discrete jump process in no longer strictly Markovian. Thus, the dynamics of the corresponding discrete MSM are only approximated. A prototypical discretization process of a continuous 2D-trajectory is shown in Figure 1.2. Based on the discrete trajectory, the discrete transition matrix T can be calculated. Each entry T_{ij} of the transition matrix is defined as the probability for a transition from discrete state i to j. It can be calculated by dividing the probability of being in state i at time t and in state j at time $t+\tau$ by the probability of being in i at time t,[23]

$$T_{ij}(\tau) = P[\mathbf{x}(t+\tau) \in S_j | \mathbf{x}(t) \in S_i] = \frac{P[\mathbf{x}(t+\tau) \in S_j \cap \mathbf{x}(t) \in S_i]}{P[\mathbf{x}(t) \in S_i]}. \tag{1.12}$$

Using c_{ij}, the observed transition count in the discrete trajec-

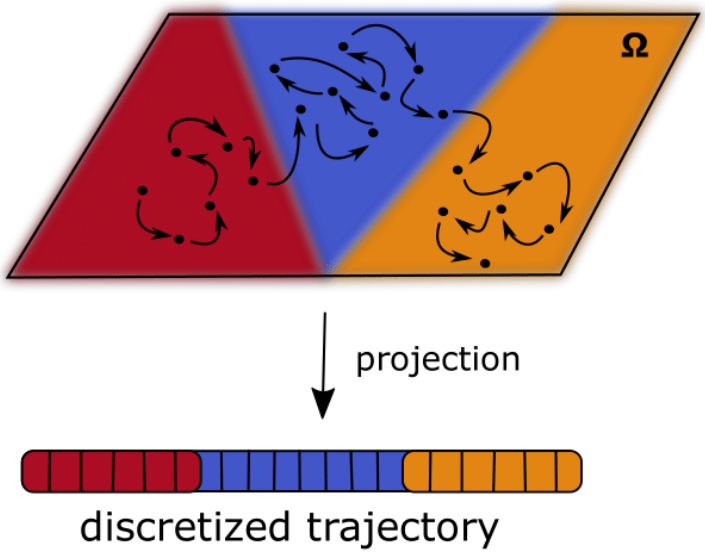
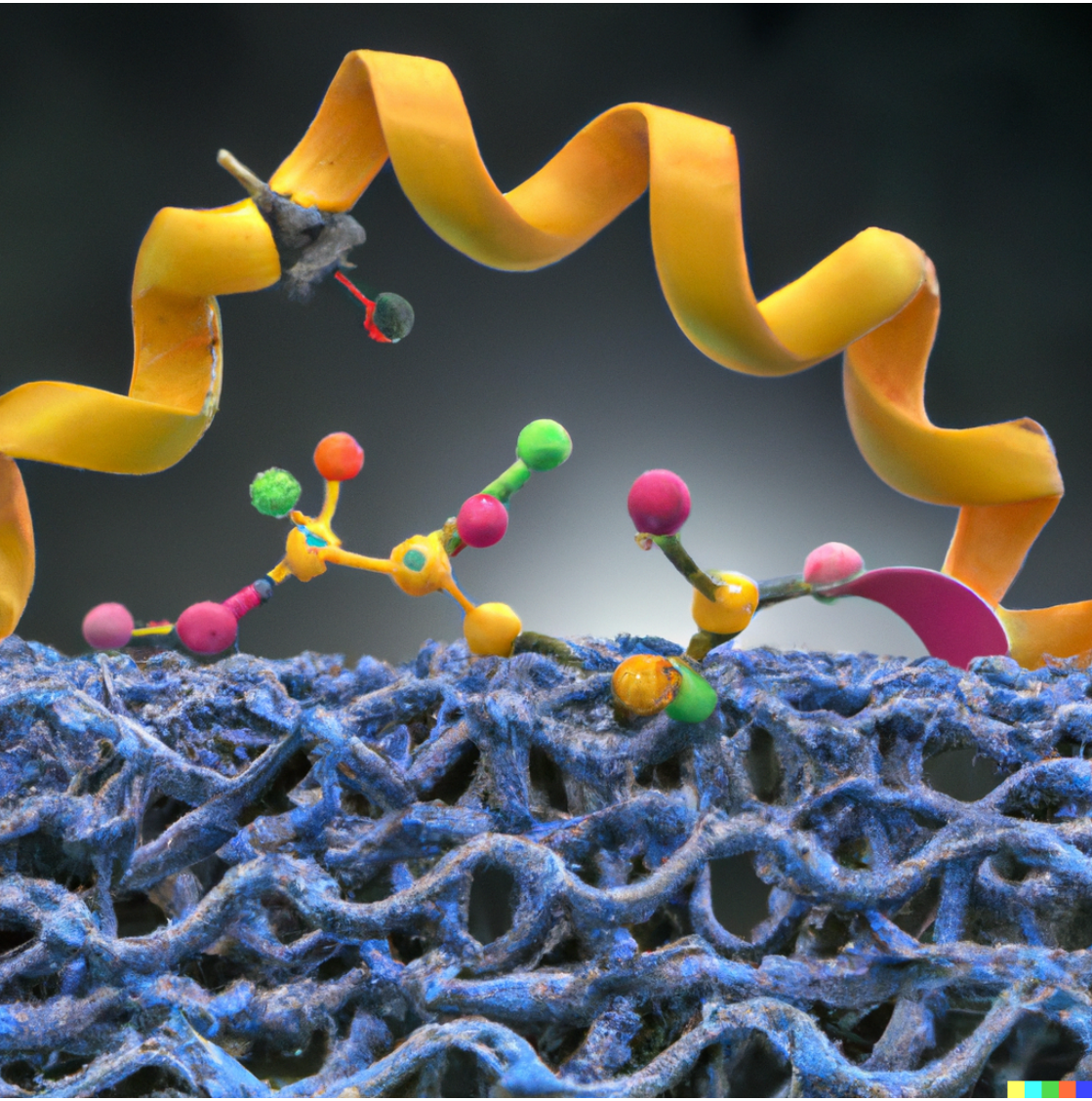


FIGURE 1.2: Discretization of the continuous phase space to a discrete trajectory. Due to the coarseness of the discretization, the information on the exact localization of the particle is lost.

tory, we can rewrite Eq. 1.12 as

$$T_{ij}(\tau) = \frac{c_{ij}(\tau)}{\sum_{j=1}^{n} c_{ij}(\tau)}. \tag{1.13}$$



2

Passing the Barrier - How Computer Simulations Can Help to Understand and Improve the Passive Membrane Permeability of Cyclic Peptides *

„After a certain high level of technical skill is achieved, science and art tend to coalesce in aesthetics, plasticity, and form. The greatest scientists are always artists as well.“

Albert Einstein

This quote and chapter are dedicated to Anna S. Kamenik and her fabulous and beautiful didactic figures

* This Chapter is reproduced in part from S. M. Linker, S. Wang, B. Ries, T. Stadelmann, S. Riniker, *CHIMIA*, 75 (2021) 518–521 and A. S. Kamenik, S. M. Linker, S. Riniker, ACS Symp. Ser. "Approaching the Next Inflection in Peptide Therapeutics: Attaining Cell Permeability and Oral Bioavailability" (2022), Ch. 5, 137–154, with the permission of ACS Publishing.

Proteins with large and flat binding sites as well as protein-protein interactions are considered "undruggable" with conventional small-molecule drugs. Cyclic peptides have been found to be capable of binding to such targets with high affinity, making this class of compounds an interesting source for possible therapeutics. However, the often poor passive membrane permeability of cyclic peptides still imposes restrictions on the applicability of cyclic peptide drugs. Here, we describe how computational methods in combination with experimental data can be used to improve our understanding of the structure-permeability relationship. Especially the conformational dynamic and chameleonic nature of cyclic peptides, which we investigate by a combination of MD simulations and kinetic modeling, is important for their ability to permeate passively through the membrane. The insights from such studies may enable to formulate design principles for the rational design of permeable cyclic peptides.

2.1 INTRODUCTION

Studies based on the sequence and structural similarity of the human genome estimate that traditional small-molecule drugs can only target 10-15% of human genes.[11] Small-molecule drugs primarily bind to deep and narrow binding pockets.[61] Many potential targets such as protein-protein interfaces or class B G-protein coupled receptors (GPCRs) have, however, large and flat binding sites and are thus "difficult-to-drug". To address this issue, increasing efforts in pharmaceutical research focus on the discovery of new compounds beyond small-molecule drugs. Cyclic peptides (and macrocycles in general) have been shown to be capable of binding to large and flat binding sites with high affinity,[62] making them possible therapeutics for difficult targets. In addition, cyclic peptides display low toxicity due to their amino acid makeup.[63] Although first cyclic peptide drugs – like the immunosuppressive cyclosporin A[13] or the anti-cancer drug romidepsin[64] – are already on the market, more insights into the mode of action and pharmacokinetics of cyclic peptides are needed.

Successful drug design requires not only strong and selective interactions between the drug and its target but also the ability of the drug to arrive at its site of action. This includes sufficient oral bioavailability (the drug fraction reaching systemic blood circulation).[65] Importantly, the key steps of oral bioavailability, which are absorption in the gut and passing the portal venous system of the liver to reach the blood stream, both require cell permeability.[65] Although there are many potential routes to enter a cell – including active transporters, endocytosis, or vesicle-based cargo systems – passive membrane permeability is the predominant mechanism for drugs.[66–68] Although cyclic peptides show

an improved membrane permeability in comparison to their linear counterparts, most cyclic peptides are still not permeable enough for oral administration.[69] In general, passive permeability decreases with increasing size (Figure 2.1).[70]

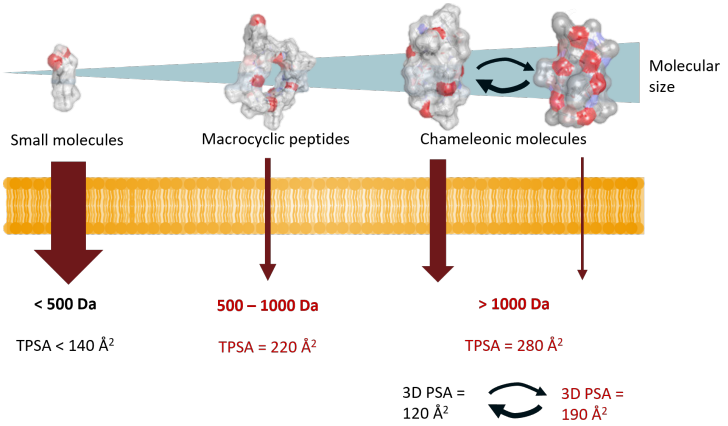


FIGURE 2.1: Molecular size impacts passive membrane permeability. High permeability is typically found for small molecules with a molecular size below 500 Da with a topological polar surface area (TPSA) below 140 $Å^2$. Chameleonic molecules like cyclosporin A can still achieve high permeability as they switch between conformations that are favorable in polar environments (high 3D PSA) and apolar environments (low 3D PSA). Adapted from P. Matsson and J. Kihlberg, *J. Med. Chem.* 2017, 60, 5, 1662–1664 with permission.

Despite violating several or all known guidelines for bioavailability,[71] the number of cell-permeable cyclic peptide drugs increases steadily.[72] The weights of typical cyclic peptides span a range of 500 to 2'000 Da (1 Da = 1 g/mol), which corresponds to a length of 5-14 amino acids and covers both relatively rigid and very flexible peptides.[73] Small cyclic peptides are conformationally restricted due to the strong ring tension imposed by the cyclization. Their passive permeability originates from their small

volume that allows them to hop in-between cavities that stochastically appear within the lipid membrane.[74,75] In contrast, with increasing size, cyclic peptides can have multiple rotatable bonds, resulting in a diverse conformational ensemble. The increased size, however, commonly leads to a reduced permeability as larger cavities are less likely to randomly emerge.[70] Some noteworthy exceptions exist. One of the first cases reported was the natural product cyclosporin A, which is orally available despite its relatively large ring size of 11 amino acids.[13] After its approval as an immunosuppressive drug in the 1980s, it quickly became established in transplantation medicine.[76,77] Most surprisingly at the time, this cyclic peptide was found to cross the cell-membrane via passive diffusion.[78] Hence, soon after its discovery, many researchers were inspired to understand why cyclosporin A has such a high membrane permeability despite its large size.[14,15,79–82] One eye-catching feature of cyclosporin A is its N-methylation pattern. While the introduction of N-methylations can improve cell-permeability due to removal of hydrogen-bond donors,[83–85] several comprehensive studies clearly have shown that the number of methylated peptide bonds alone cannot explain the cell-permeability of cyclic peptides.[86–89]

The predominant strategy to distinguish permeable and impermeable compounds is to synthesize a series of cyclic peptides and experimentally determine their membrane permeability. Typical approaches to improve the permeability in a given compound series are to investigate different backbone N-methylation patterns, side chain modifications, and stereochemical changes.[90] In some cases, the addition of a cell-penetrating peptide that triggers endocytosis was found to improve bioavailability.[91] However, these changes are often applied in a time-consuming trial-and-error process and improvements in permeability may come with

loss in binding affinity. Therefore, demands and efforts have increased to develop computational models to rationalize membrane permeability and support the development of bioavailable cyclic peptide drugs.

2.2 COMPUTATIONAL APPROACHES TO DESIGN CYCLIC PEPTIDES

Calculation of the membrane permeability of small molecules ranges back to Meyer and Overton.[92] The calculated lipophilicity measure AlogP[93] is often used as a heuristic for the passive permeability of drug-like molecules. However, no established descriptor that can be calculated from the molecular structure, such as the number hydrogen bond acceptors/donors, lipophilicity (as measured by logD or logP, i.e., the partition (or distribution) coefficient between polar and apolar solvents), radius of gyration, or the (3D) polar surface area is able to reliably predict the membrane permeability of macrocycles.[12,94,95] Over the past decades, evidence accumulated that the key to rationalize the structure-permeability relationship of cyclic peptides lies in their conformational flexibility and thus their conformational ensembles.[16,73,89,96–99] Indeed, the relationship between size and permeability differs distinctly between small molecules and macrocycles, an effect that cannot be explained with traditional permeation models.[70] As a result, the focus shifted to computational methods that can take into account the flexibility of molecules.

2.2.1 Molecular Dynamics Simulations of Cyclic Peptides

Molecular dynamics (MD) simulations use Newtonian physics to simulate the motion and interaction of molecules. They are a widely used tool to study kinetics, energetics, and conformations of biological and chemical systems (for a review see e.g. Ref. 19). Using MD, the dynamics of cyclic peptides can be analyzed in atomic detail, a resolution that cannot be achieved with experimental techniques. The hypotheses generated based on the simulations are then tested and validated in experiments. Therefore, it is common practice to combine MD simulations and experimental techniques to obtain a holistic picture.

Combinations of NMR and MD studies revealed that cyclic peptides display a chameleonic behavior.[16,89,96,100] Cyclic peptides can form intramolecular hydrogen bonds, leading to a so-called closed conformation, where most polar groups are shielded from the apolar environment. These molecules are conformationally flexible and can exhibit either lipophilic or lipophobic properties, depending on their environment (Figure 2.2).[12,100–103] Therefore, chameleonic molecules can combine high permeability with high solubility, both of which are key features for oral bioavailability.

Due to the shielding of polar groups in the closed conformation, the peptide has a lower desolvation energy barrier for moving into the membrane. Therefore, the closed conformation (also termed low-permittivity conformation[102]) is considered the main permeable "species".[102]) However, the existence of the closed conformation is a necessary but not sufficient condition for permeability.[104] Other properties like side chains, size, and the polar or hydrophobic surface area heavily influence the permeability

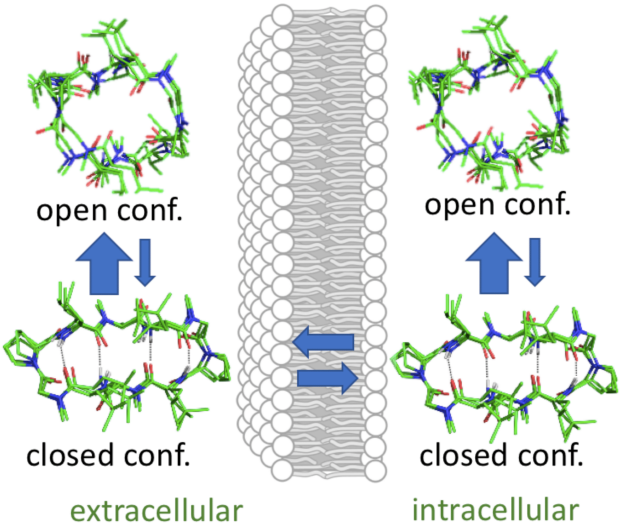


FIGURE 2.2: Model for the passive membrane permeability of cyclic peptides. In solution, cyclic peptides are predominantly in the open form that can form hydrogen bonds with the solvent. The closed form is characterized by internal hydrogen bonds. Those internal hydrogen bonds shield the polar groups of the cyclic peptides from an apolar environment making the cyclic peptide membrane permeable.

of cyclic peptides. A recent study combining MD simulations, NMR, and permeability assays showed that the size of the largest connected hydrophobic surface of the cyclic peptide is decisive for its permeability.[105] To increase the membrane permeability of cyclic peptides, the authors suggest to place N-methylations or hydrophobic residues in positions where they connect two separated hydrophobic surfaces. Recent work in our group revealed how a change in only a single stereocenter in two semi-peptidic macrocycles (Figure 2.3) leads to a "permeability cliff" in experimental permeability assays.[106] Using a combination of NMR and MD studies, we could show how the change affects the conformational ensemble of the macrocycles and through this the population of conformations with a maximal number of intramolecular hydrogen bonds in the apolar (chloroform) environment.

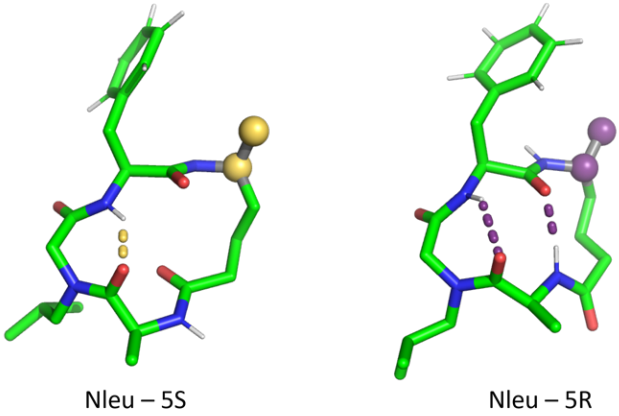


FIGURE 2.3: The two four amino acid cyclic peptides, Nleu-5S (golden) and Nleu-5R (purple) differ only in the fifth C-atom stereocenter in the alkyl linker (spheres). The shown structures are snapshots from our extensive MD simulations representing the differences in both molecules' hydrogen-bond patterns in apolar environments.

The findings described in these studies would not have been possible without the atomistic insights on the cyclic peptides provided by MD simulations. Furthermore, computer simulations may not only help to improve the permeability, but also to identify peptide candidates for specific targets. For example, MD simulations have been used to pre-select cyclic peptides drug candidates for targeting protein-protein interfaces.[107]

2.2.2 Kinetic Models of Cyclic Peptides

As discussed above, conformational flexibility of cyclic peptides is a crucial determinant for their membrane permeability. Kinetic information is needed to fully account for the dynamics of cyclic peptides. In contrast to many other computational techniques, MD simulations can provide such kinetic information. The MD procedure we have developed for permeability analysis is shown in Figure 2.4.[89,108] First, a diverse conformer set is generated for a cyclic peptide of interest. When an experimental structure (X-ray or NMR) is available, this can be done by taking snapshots from an enhanced sampling simulation starting from the experimental structure. Alternatively, *in silico* conformer generators can be used. Due to their size, cyclic peptides span a far larger conformational space than can be exhaustively sampled. It is therefore instrumental to bias sampling towards the more relevant phase space. We recently improved the RDKit conformer generator ETKDG specifically for macrocyclic species by incorporating experimental torsion information as well as shape and interaction heuristics.[109] Once selected, each conformer is solvated in water (mimicking the polar extracellular condition) and chloroform (mimicking the apolar dielectric environment within the lipid core). Next, extensive simulations are performed in parallel

in the two solvents, starting from the conformer set. The individual simulations are typically on the order of hundreds of nanoseconds and total sampling time on the order of tens of microseconds per solvent. The parallel trajectories are used to build Markov state models (MSMs).[22,23,110,111] MSMs extract metastable conformational states and the interconversion rate between them, and enable the comparison of the conformational behavior of structurally similar compounds. They are therefore a valuable tool to rationalize membrane permeability. The kinetic models obtained can also be used to compare internal flexibility between slowly exchanging core conformations with experimentally determined NMR T2 relaxation times.[112] When constructing MSMs, we use density-based clustering algorithms.[89] Recently, we could show that a hierarchical density-based clustering is able to better capture metastable states residing at free-energy wells of varied depth.[113]

Our kinetic models based on MD simulations of the permeable cyclosporin A revealed the existence of so-called congruent conformational states,[16] which are populated in both polar and apolar environments. Such congruent conformational states can facilitate the transition between different environments and therefore increase permeability (Figure 2.2). The closed conformation of cyclosporin A is an example of a congruent conformation. In addition, MSMs suggested that the main difference between cyclosporin A and the structurally closely related but impermeable cyclosporin E are the interconversion rates between their metastable states.[96] When applying our workflow to rationalize the membrane permeability differences of six cyclic decapeptides sharing the same backbone N-methylation pattern but differing in their side chain composition,[89] the kinetic models revealed that the population of the closed conformational state in water cor-

relates with the passive membrane permeability. Based on this observation, we expanded our analysis onto a larger set of 24 decapeptides, systematically covering side chain compositions. We find that the closed state population in water remains a distinguishing metric of passive permeability and the MD simulations provided mechanistic insights into the formation of the closed conformation.[108]

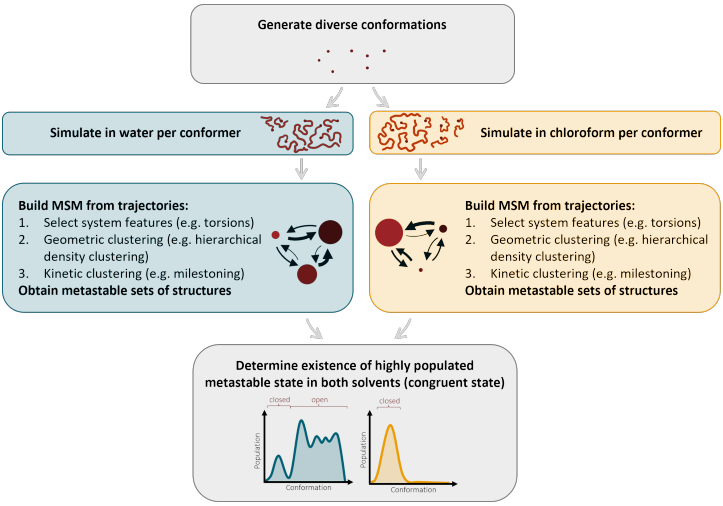


FIGURE 2.4: Workflow illustration to identify sufficiently populated congruent conformations from extensive MD simulations in water and chloroform.

2.2.3 Simulations of Cyclic Peptides at Membranes and Interfaces

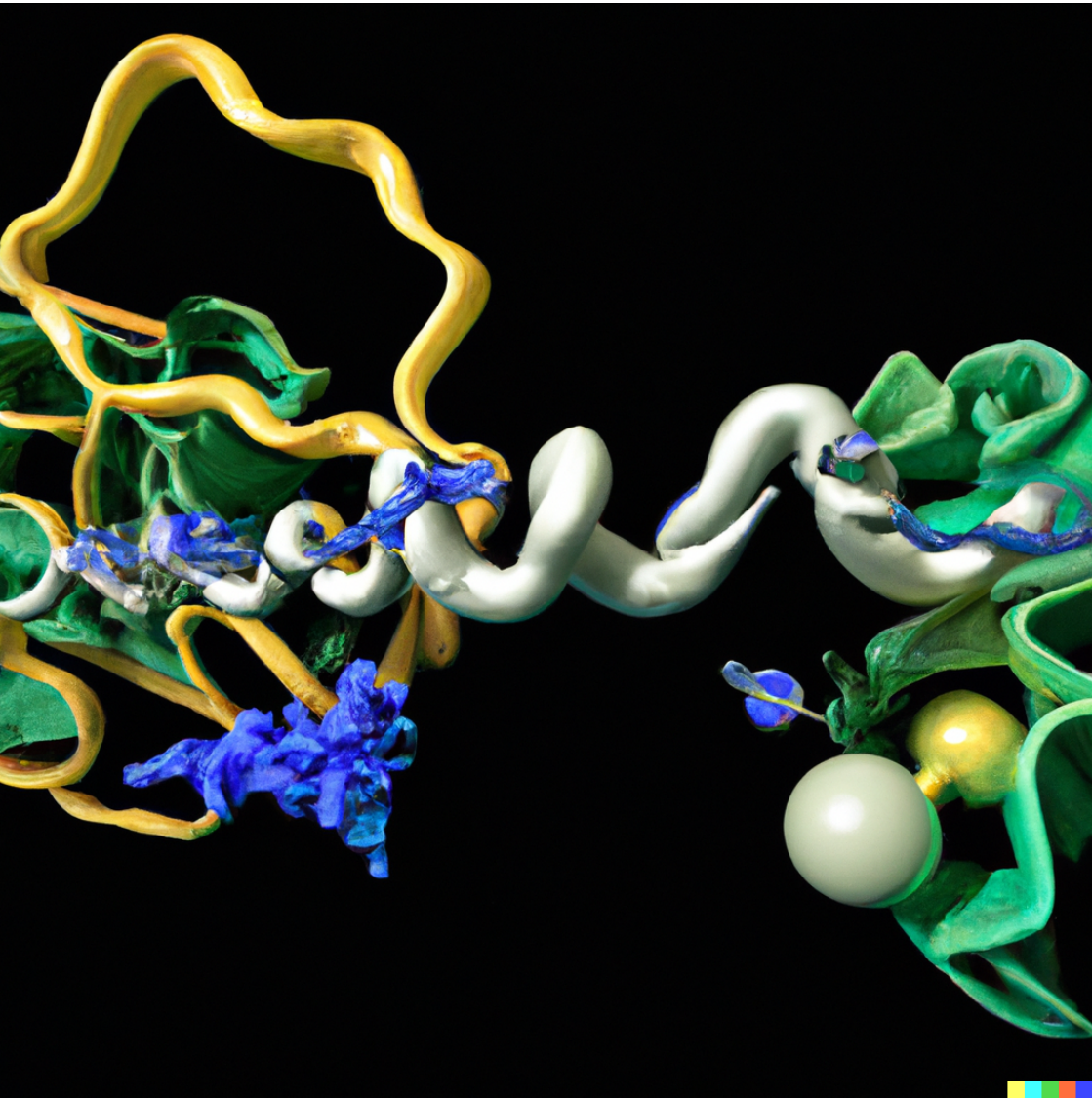
To better understand the passive membrane permeability of cyclic peptides, it would be desirable to directly simulate their interaction with lipid membranes. However, these large and complex

systems require tremendous computational resources. Studies investigating these interactions for passively permeating peptides are therefore still rare, and the few available works involved computational tricks to speed up the simulations.[80,114] For example, Wang *et al.*[80] simulated cyclosporin A at a POPC lipid bilayer by increasing the flexibility of the peptide backbone by 40% and increasing the temperature to over 400 K. At such elevated temperatures, the membrane in the simulation becomes porous. Nevertheless, their observations supported the view that conformational flexibility is important for membrane permeability and the authors hypothesized that a higher flexibility of the peptide leads to a higher permeability. An alternative is to mimic the water-membrane interface with a less resource-intense model system. We found that cyclic peptides have the same preferred orientation as at a biological membrane while residing at a chloroform/water interface.[115] Therefore, this simplified system can be useful to model properties that depend on the local polar/apolar environment of the peptides.[115]

2.3 CONCLUSION

MD simulations yield atomistic insights into the conformational dynamics of cyclic peptides. Together with experimental efforts, the findings have already helped to improve our understanding of the passive membrane permeability of cyclic peptides. Future research will focus on the role of the membrane in the permeation process and on further unveiling the complex relationship between structure and conformational behavior of cyclic peptides. This knowledge will help to improve the rational design of peptidic

drug candidates and the development of novel therapeutics for currently undruggable targets.



3

Polar/Apolar Interfaces Modulate the Conformational Behavior of Cyclic Peptides with Impact on Their Passive Membrane Permeability *

Ist es ein lebendig Wesen,
Das sich in sich selbst getrennt?
Sind es zwei, die sich erlesen,
Dass man sie als eines kennt?

Johann Wolfgang von Goethe

This quote and chapter are dedicated to Christian Schellhaas, who with his dedication and sharp mind greatly advanced our understanding of permeability

* This Chapter is reproduced in part from S. M. Linker†, C. Schellhaas†, B. J. Ries, H.-J. Roth, M. Fouché, S. Rodde, S. Riniker *RSC Adv.*, 12 (2022), 5782–5796 with permission from the RSC. †These authors contributed equally.

Cyclic peptides have the potential to vastly extend the scope of druggable proteins and lead to new therapeutics for currently untreatable diseases. However, cyclic peptides often suffer from poor bioavailability. To uncover design principles for permeable cyclic peptides, a promising strategy is to analyze the conformational dynamics of the peptides using molecular dynamics (MD) and Markov state models (MSMs). Previous MD studies have focused on the conformational dynamics in pure aqueous or apolar environments to rationalize membrane permeability. However, during the key steps of the permeation through the membrane, cyclic peptides are exposed to interfaces between polar and apolar regions. Recent studies revealed that the free energy minima of the permeation process (i.e. the most favorable position of the peptide) was at these interfaces. Thus, a deeper understanding of the behavior of cyclic peptides at polar/apolar interfaces is desired. Here, we investigate the conformational and kinetic behavior of cyclic decapeptides at a water/chloroform interface using unbiased MD simulations and MSMs. The distinct environments at the interface alter the conformational equilibrium as well as the interconversion kinetics of cyclic peptide conformations. For peptides with low population of the permeable conformation in aqueous solution, the polar/apolar interface facilitates the interconversion to the closed conformation, which is required for membrane permeation. Comparison to unbiased MD simulations with a POPC bilayer reveals that not only the conformations but also the orienta-

tions are relevant in a membrane system. These findings allow us to propose a permeability model that includes both ‘prefolding’ and ‘non-prefolding’ cyclic peptides – an extension that can lead to new design considerations for permeable cyclic peptides.

3.1 INTRODUCTION

Existing pharmaceutical drugs cover only a small fraction of the over 20'000 proteins encoded in the human genome.[11,116,117] Most of the druggable proteins share distinct structural features, compared to the undruggable fraction: They have restricted, pocket-shaped binding sites that favor interactions with small organic molecules.[61,118] In addition, their amount of well-defined rigid domains is significant. However, bioinformatic studies estimate that 85 % to 90 % of proteins do not contain suitable pocket-shaped binding sites and/or well-defined rigid domains, and are therefore difficult to target by small-molecule drugs.[11,116] In contrast, macrocyclic drugs like cyclic peptides can bind to larger binding sites with flat profiles or protein-protein interfaces.[62,119–124] In addition, cyclization prevents rapid metabolic clearance.[63] Therefore, cyclic peptides have the potential to vastly extend the scope of druggable proteins and lead to therapeutics for currently untreatable diseases.[125]

The therapeutic applicability of cyclic peptides is, however, limited by their often low cell permeability and oral bioavailability.[69,94] To address this issue, many experimental and computational studies have focused on the molecular mechanism of cell permeation of cyclic peptides with the aim to define strategies for the rational design of permeable cyclic peptides.[16,89,96,105,108,126–129] N-methylation of the peptide backbone, change of stereocenters, conformational flexibility, and side-chain modifications can all influence the permeability. Unfortunately, their effect is non-linear and highly site-dependent.[98,106,108,130–132] The cell permeability of cyclic peptides often drops with increasing peptide size.[70] Nevertheless, some larger cyclic peptides can display internal con-

formational changes, which are crucial for membrane permeablility.[101,102,104] The different conformations can be distinguished by the amount of intramolecular hydrogen bonds. In the so called 'closed' conformation, the internal hydrogen bonds shield the polar groups from the environment leading to a low polar surface area. In contrast, polar groups are exposed to the environment in the 'open' conformation.[16,100,102] Therefore, cyclic peptides can adapt to polar and apolar environments by changing from one conformation to another, an ability called the 'chameleonic' behavior of cyclic peptides.[12,102]

The shielding of polar groups in the 'closed' conformation lowers the desolvation energy barrier for moving into the apolar membrane interior. Therefore, the 'closed' conformation is assumed to be the main permeable species.[72,89,95,102] However, the possibility to adopt a 'closed' conformation in an apolar environment alone does not necessarily imply membrane permeability.[89,104] Molecular dynamics (MD) and nuclear magnetic resonance (NMR) studies have revealed that congruent conformational states (i.e. conformations occurring in both polar and apolar environments, of which the 'closed' conformation can be one) facilitate the transition between different environments and therefore increase membrane permeability.[16,89,108] The peptide composition, size, and hydrophobic surface area heavily influence the conformational behaviour of cyclic peptides and thus also the permeability.[70,89,98,105,130] The interplay between all these factors is not trivial to decipher and therefore, it is difficult to establish structure-permeability relationships.

During their path through the membrane, cyclic peptides pass different environments. They start in a polar aqueous environment outside the cell, cross the polar and often charged lipid head-group region, move through the apolar lipid-tail region, and

again pass the head-group region in order to reach the interior of the cell. Thus, cyclic peptides not only face different environments, but also multiple interfaces between polar and apolar regions. The conformational and dynamic behavior at these interfaces is hardly understood since previous simulation approaches have focused mainly on homogeneous environments.[89,108] Only few studies have been reported that targeted cyclic peptides in non-homogeneous environments.[80,114] In both cases, biased enhanced sampling approaches were employed to achieve sufficient sampling in the available simulation time. Wang *et al.*[80] simulated cyclosporin A at a water/chloroform interface as well as in a 1-palmitoyl-2-oleoylphosphatidylcholine (POPC) bilayer system. To enhance sampling, the authors increased the temperature of the system to 490 K and reduced specific force constants acting on the ω-dihedral angles. Sugita *et al.*[114] used steered MD[133] and umbrella sampling[134] to pull more than 100 different cyclic peptides into a POPC bilayer and calculate their free-energy profile. In addition, they focused on small cyclic peptides with little conformational flexibility in order to obtain converged results. These studies led to important new insights into the permeability of cyclic peptides. The free-energy minimum along the permeation pathway (i.e. the most favorable position of the peptide) was neither in the aqueous phase nor in the apolar lipid-tail region, but directly beneath the lipid head-group region at the polar/apolar interface.[80,114] This behavior was conserved over a wide range of peptides with different hydrophilicity. The free-energy minimum was more distinct for more lipophilic peptides, but even very hydrophilic peptides showed this minimum.[114] As cyclic peptides are likely to spend a large proportion of their permeation process located at this minimum, a detailed and non-biased understanding of how the interface influences the conformational behavior

and dynamics of cyclic peptides is desired.

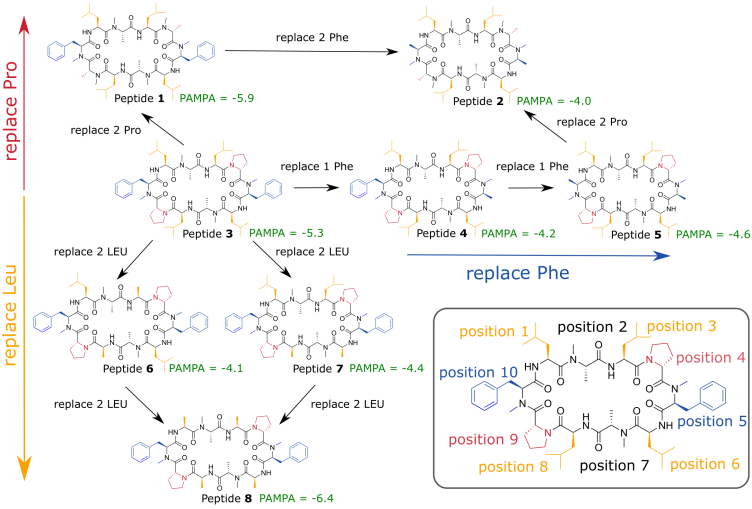


FIGURE 3.1: Cyclic decapeptide (CDP) series used in this chapter. The backbone scaffold was reported by Fouché *et al.*[135,136] and is kept constant. The side-chain residues are systematically varied between leucine/alanine at position 1, 3, 6, and 8 (yellow), proline/alanine at position 4 and 9 (red), and D-phenylalanine/D-alanine at position 5 and 10 (blue). In position 2 and 8 are alanines in all peptides. In position 5 and 10 are D-amino acids to enable the correct β-turn conformation. The parallel artificial membrane permeation assay (PAMPA) data were taken from Ref. 108. Note that CDP **1** and **3** in this chapter correspond to CDP **6** and **4**, respectively, in Refs. 89,108.

In this chapter, we use extensive unbiased MD simulations to investigate the behavior of a series of eight cyclic decapeptides (CDPs) at a water/chloroform interface, and compare the results to unbiased simulations at a POPC bilayer. In contrast to the work of Sugita *et al.*,[114] the peptides are chosen to show complex internal conformational dynamics. The simulations are performed without biases and at room temperature to avoid ar-

tifacts like the distortion of the interface or POPC bilayer or the formation of pores. In addition, this ensures that the observed conformations and kinetics reflect the natural behavior of the cyclic peptides. The backbone scaffold of the CDPs and their N-methylation pattern was introduced by Fouché *et al.*[135,136] and is kept throughout our series while the side chains are varied. Figure 3.1 illustrates the variations that are performed in three dimensions: (i) a switch from leucine to alanine at position 1, 3, 6, and 8, (ii) proline to alanine at position 4 and 9, and (iii) D-phenylalanine to D-alanine at position 5 and 10. In the closed conformation, as observed by NMR and in crystal structures, the peptides form two β-strands (amino acids at position 1, 2, 3 and 6, 7, 8) and two β-turns (amino acids at position 4, 5 and 9, 10). In this conformation, all unmethylated amide nitrogen atoms face towards the peptide's interior and build the typical four H-bonds pattern.[89] We investigate how the interplay between polar/apolar interfaces and the peptide orientation modulate the conformational and kinetic behavior of the CDPs, and showcase the important role of interfaces in the passive permeation process of cyclic peptides.

3.2 METHODS

Experimental data shown in this chapter was generated by our collaborators at Novartis Institutes for BioMedical Research. Methodological details are described in Ref. 115.

3.2.1 MD Simulations

Initial Conformations

All MD simulations were initiated with CDPs in their 'closed' or major 'open' conformation. The 'closed' conformers were obtained from measurements in chloroform by NMR spectroscopy.[135] The generation of the major 'open' conformers has been described in detail in Refs. 89,108. Briefly, a set of 100 seed conformers from enhanced sampling runs facilitating an opening of the 'closed' conformers was selected to perform repeated parallel MD simulations of 100 ns length until the subsequent core-set Markov model (CSMM) converged. The 'open' conformers of CDPs **2** and **4**-**8** were adapted by *in silico* mutagenesis from CDP **1** and **3**, respectively, and used as an ensemble set of seeds for the corresponding MD simulations.

Water/Chloroform Interface

All MD simulations at the water/chloroform interface were performed using the Groningen Molecular Simulation (GROMOS) software package[137] and the GROMOS 54A8 force field.[138] The simulations were carried out under NPT conditions with rectangular periodic boundary conditions. The leapfrog scheme[32] was used to integrate Newton's equations of motion with a time step of 2 fs. The simple-point-charge (SPC) water solvent model[139] was used. Weak coupling[37] to three separate temperature baths for the peptide, the chloroform phase and the water solvent was applied with a reference temperature of 300 K and a relaxation time of 0.1 ps. The pressure was maintained around 1.013 bar (1 atm) by weak coupling to a pressure bath with a relaxation time of 0.5 ps and a isothermal compressibility of $0.4575\,\mathrm{nm}^2/\mathrm{N}$

under isotropic scaling of the simulation box. For the nonbonded interaction, a twin-range cutoff scheme was used with 0.8 nm and 1.4 nm cutoffs. A reaction-field force[140] with a relative dielectric permittivity of 61.0[141] was used for the treatment of electrostatic interactions beyond the long-range cutoff. The SHAKE algorithm[142] was used to constrain bond lengths with a relative tolerance of 10^{-4}. Coordinate and energy trajectories were written out every 5 ps for data analysis.

The biphasic simulation system consisted of 400 chloroform molecules and an equal volume of water molecules. The simulation box measured 7.52 nm x 3.76 nm x 3.76 nm. A cyclic peptide in either the 'closed' conformation or the major 'open' state was placed in the simulation box in various orientations either at the interface of the water and chloroform phases, or at the center of the aqueous phase, with approximately 1.88 nm distance between the CDP's center of mass and the closest chloroform atom at the interface. Each simulation was preceded by 20 ps NVT thermalisation and equilibration under positional restraining of the solute atoms. Initial velocities were generated using a random number generator seed at an initial temperature of 300 K. If not mentioned otherwise, 50 MD simulations with different starting orientations were performed for 200 ns per peptide and starting conformation, resulting in 20 µs of sampling of each CDP. The first 2 ns of each simulation were discarded from the analysis for equilibration.

POPC Bilayer

The MD simulations at the POPC lipid bilayer (a mimic for a cellular membrane) were performed using the Groningen Machine for Chemical Simulations (GROMACS) 2020.5 software package[143] and the GROMOS 54A8 force field.[138] Lipid parame-

ters were adopted from the POPC model of Marzuoli *et al.*[144] to improve solvation properties of the head-group region. The simulations were carried out under NPT conditions with rectangular periodic boundary conditions. Again, the SPC[139] water model was used as solvent. The leapfrog scheme[32] was used to integrate Newton's equations of motion with a time step of 2 fs. Weak coupling[37] to three separate temperature baths at 303 K for the peptide, the lipids and the solvent was applied with a relaxation time of 0.1 ps. The pressure was coupled semi-isotropically to a Parrinello-Rahman barostat[145] at 1.0 bar with a coupling constant of 2.0 ps and an isothermal compressibility of $0.45\,\mathrm{nm}^2/\mathrm{N}$. For both the short-range electrostatic and van der Waals interactions, a single cutoff of 1.2 nm was used. The long-range electrostatics were calculated by the particle mesh Ewald (PME) algorithm.[146] The linear constraint solver (LINCS) algorithm[147] was used to impose constraints on the bond lengths with fourth order expansion. Center-of-mass (COM) motion removal was applied in every simulation step to remove the motion of the bilayer relative to the solvent. Coordinate trajectories were written out every 100 ps for data analysis.

The topology of the simulation box containing 512 POPC lipids in a bilayer (256 per leaflet) was adopted from Marzuoli *et al.*[144] The CDPs were placed either in their 'closed' conformation or the major 'open' state at the center of the aqueous phase in the simulation box, with approximately 3 nm distance between the CDP's center of mass and the closest head-group atom of the POPC lipids. Each simulation was preceded by 100 ps NVT thermalisation and 1 ns NPT equilibration. In total, 50 runs with 100 ns length were started in the 'open' and 'closed' conformation of CDP **1** and CDP **3** each. After manual assessment, those runs that showed an initial contact with the membrane were selected.

For CDPs **1** and **3**, two and three of the 100 runs showed an initial contact with the membrane, respectively. The last frame of these simulations was used as the seed for five continuation runs with 100 ns each.

Construction of Markov Models

MSMs provide a statistical framework to describe the complete system dynamics.[22] Using MSMs, one can compute stationary quantities and long-time kinetics from ensembles of short simulations. Importantly, these short simulations need to be in 'local equilibrium' within the MSM states, but are not required to be in 'global equilibrium'. Thus, we can combine multiple short simulations of 200 ns to obtain the system's kinetics in the µs regime.

For all trajectories, the sine and cosine of the backbone torsion angles ϕ and ψ were extracted using the PyEMMA Python library,[148] representing the first 40 feature dimensions. Additionally, the z-position in the simulation box and orientation of the CDPs with respect to the z-vector of the simulation box were extracted from the trajectories using the MDTraj[149] Python library, resulting in a total of 42 feature dimensions. In the system description, no side-chain information was explicitly included. The 42 feature dimensions were reduced to 6–11 collective coordinates (the exact number depends on the peptide) by time-lagged independent component analysis (TICA).[150]

The hierarchical density-based Sittel-Stock clustering algorithm[151] was used for the spatial clustering. The Sittel-Stock algorithm requires a cutoff for the minimal number of members for a cluster (k). Here, a cutoff k=100 was used. To obtain effective transition probabilities between the conformational states, the core-set Markov model technique (CSMM)[152–155] was used. For the Markov model, a lagtime τ of 1.5 ns was chosen to ensure

Markovianity. Furthermore, robust Perron cluster-cluster analysis (PCCA+)[156] was performed to group the microstates into metastable conformational states. Depending on the peptide, three or four macrostates were chosen. Visual inspection of these showed that this procedure lead to the separation of a 'closed', an 'open orientation A', and an 'open orientation B' state for each peptide.

The mean first passage times (MFPTs) $\mathbf{E}_x[T_y]$ of the interconversion processes describe the expected hitting times of one target state y in Y when starting in state x in X. MFPTs were calculated from the transition matrix T with the following equation,

$$\mathbf{E}_x[T_y] = \begin{cases} 0 & x = y \\ 1 + \sum_z T_{x,z}\mathbb{E}_z[T_y] & x \neq y \end{cases} \tag{3.1}$$

Bootstrapping was performed to obtain the average and standard deviation for the steady-state populations of the Markov models. A total of 50 bootstrapping iterations were performed. In each iteration, n trajectories were picked from the total set of trajectories with replacement, where n equals the total number of trajectories.

Data Analysis

If not stated otherwise, all trajectories were analyzed using the MDTraj[149] Python library.

Position. The position of the CDP inside the biphasic simulation box was calculated by dividing the number of atom–atom contacts within a threshold of 0.5 nm between the CDP and the chloroform molecules with the number of atom–atom contacts between the CDP and molecules of the water and chloroform phase

as follows,

$$\text{Position} = \frac{\text{Atom contacts }_{\text{Peptide}}(CHCl_3)}{\text{Atom contacts }_{\text{Peptide}}\ (CHCl_3 + H_2O)} \tag{3.2}$$

Orientation with respect to the simulation box. To calculate the orientation of the peptides with respect to the simulation box, the cosine of α, the scalar product between the normal vector to the plane set up by the CDP ($\vec{a}$) and a vector parallel to the length of the simulation box ($\vec{b}$) was determined,

$$\text{Orientation} = \cos(\alpha) = (\vec{a} \cdot \vec{b}) \tag{3.3}$$

Because of the periodic boundary conditions, the vector parallel to the length of the biphasic simulation box ($\vec{b}$) was calculated as the vector pointing from the COM of the aqueous phase to the COM of the chloroform phase. In the simulation with the POPC bilayer, on the other hand, the z-axis of the system was used as $\vec{b}$. In each case, these vectors ($\vec{b}$) constitute the normal vector to the polar/apolar interface.

Orientation with respect to the water/chloroform interface by side-chain interactions. To calculate the orientation of the peptides with respect to the water/chloroform interface by side-chain interactions, the number of atom–atom contacts within a threshold of 0.5 nm between the non-hydrogen atoms of the leucine side chains (alanine if substituted) and chloroform molecules was divided by the number of atom–atom contacts between the non-hydrogen atoms of the leucine side chains (alanine if substituted) and molecules of the water and chloroform phase as follows,

$$\text{Orientation} = \frac{\text{Atom contacts }_{\text{Leu/Ala}}\ (CHCl_3)}{\text{Total Solvent Atom contacts }_{\text{Leu/Ala}}} \tag{3.4}$$

RMSD. The atom-positional backbone root-mean square deviation (RMSD) was calculated with respect to the NMR solution structure of the 'closed' state.

State assignment. Re-assignment of simulation frames to the macrostates identified by Markov state modelling was used to analyze the distribution of total energies as well as the positions and orientations of the CDPs within these states. If not stated otherwise, only the frames with a probability $> 90\,\%$ of belonging to a certain metastable set were included.

Data and Software Availability

The structure and topology files of the CDPs in the closed and major open conformation are available on GitHub (`https://github.com/rinikerlab/cyclic_peptide_at_interfaces`). The GitHub repository also contains a sample Jupyter notebooks for the MSM analysis of CDP **1**.

Further information, custom scripts, or production trajectories are available from the corresponding author (S.R.) upon request. The freely available software can be obtained via the following links: GROMOS (`http://www.gromos.net/`), GROMACS (`https://www.gromacs.org/`), and PyMol (`https://github.com/schrodinger/pymol-open-source`).

3.3 RESULTS AND DISCUSSION

3.3.1 3D Hydrophobicity Profile

The 3D conformations of the cyclic decapeptides (CDPs) show a directed hydrophobicity profile, i.e. one side of the peptides is

more hydrophobic than the other. This directionality is mainly caused by leucine residues, a common feature of the CDP series.[136] Figure 3.2 displays the conformation of CDP **1** in the 'closed' and the highest populated 'open' state obtained by a MSM of the peptide in water. The solvent-accessible surface of the hydrophobic leucine residues is highlighted in orange. In both the 'closed' and 'open' conformation, all leucine residues point outwards of the plane defined by the backbone macrocycle. We will call this direction the peptide's 'normal'. The leucine side-chain orientations are relative stable throughout the simulations. In contrast, the orientation of the phenylalanine side chains is flexible and can span a 180° angle. Thus, the CDPs show a higher occurrence of hydrophobic side chains on one side, which leads to the directed hydrophobicity pattern. In the 'closed' conformation, the leucine side chains build a continuous, large hydrophobic patch, while this patch is split into smaller ones in the 'open' conformations. The directed hydrophobicity and conformational dependency was observed for all peptides of the series except CDP **8**, which contains no leucines (see Figure 3.3).

When the peptides are simulated in an isotropic environment like a water box, the observed directed hydrophobicity has no effect on the conformational dynamics of the peptides. However, this may be different at interfaces between polar and apolar environments. Sugita *et al.*[114] have shown that the most favorable position of cyclic peptides at the membrane is directly underneath the lipid head-group region at the polar/apolar interface. Such an interface has its own directionality that can interact with the directed hydrophobicity of the CDPs and effect the conformational behavior. As previous work has mainly focused on isotropic environments, the role of the directionality could not be captured. The observed differences in the hydrophobic patch between the

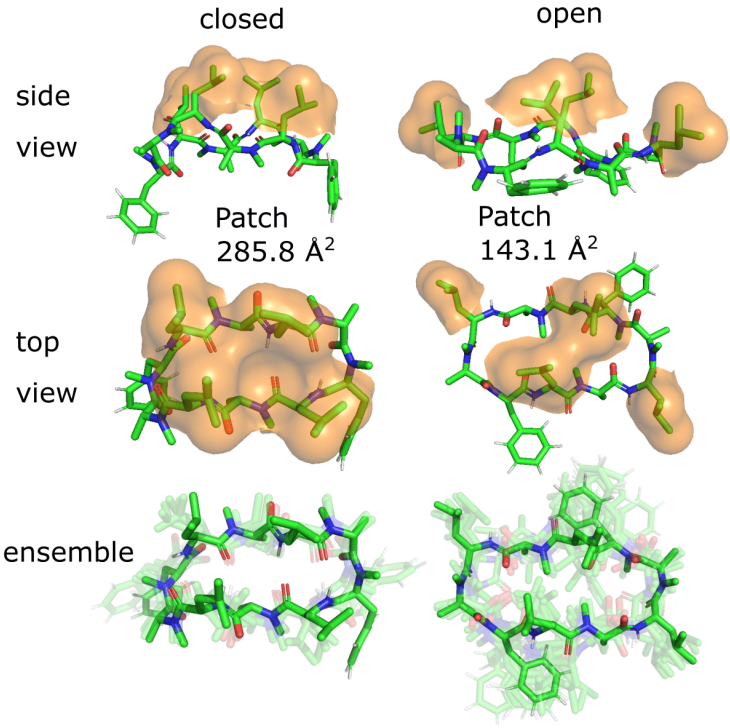


FIGURE 3.2: Metastable conformations of CDP **1** in water. The 'closed' as well as the highest populated 'open' state are shown. In the two top panels, the solvent-accessible surface of the hydrophobic leucine residues is indicated by orange spheres. The conformational ensemble of the metastable states is shown at the bottom.

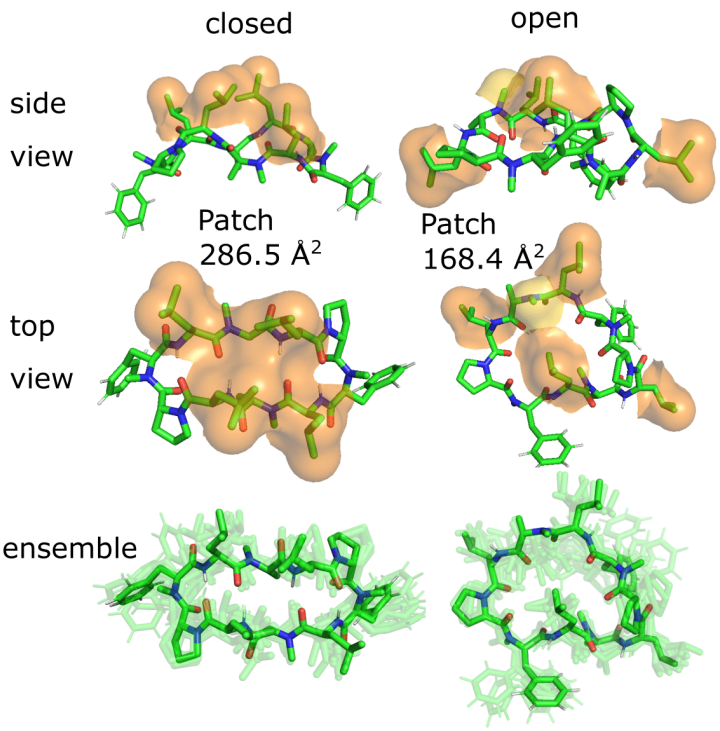


FIGURE 3.3: Metastable conformations of proline-containing CDP **3** in water. The 'closed' as well as the highest populated 'open' state are shown. In the two top panels, the solvent-accessible surface of the hydrophobic leucine residues is indicated by orange spheres. For the 'open' conformation, the solvent-accessible surface of a backbone N-methylation, which bridges the hydrophobic leucine patches, is indicated by a yellow sphere. The conformational ensemble of the metastable states is shown at the bottom.

'open' and 'closed' state are thereby of special interest, as Hoang and co-workers[105] found that the size of the largest continuous hydrophobic surface patch correlates with the membrane permeability in a series of cyclic hexapeptides and heptapeptides. This raises the question whether the different patch sizes of 'open' and 'closed' conformers also lead to permeability effects.

3.3.2 Peptide Orientations at Interfaces

A water/chloroform interface was chosen as a model system. This combination allows straightforward comparison to previous work that was performed in pure water or pure chloroform.[16,89,105,108] We first tested how the CDPs orient towards the interface. Therefore, CDP **1** was placed at the interface in the 'closed' and in the major 'open' conformation with different orientations and simulated without constraints. Interestingly, only two stable orientations, called A and B in the following, were observed in the open conformation. In the closed conformation, only one stable orientation, orientation A, was seen. All other orientations quickly rotated towards one of the stable ones. In both orientations A and B, the plane defined by the peptide backbone laid flat on the interface. In orientation A, the leucine side chains face towards the chloroform phase, whereas in orientation B they face towards the water phase.

To confirm that the stable orientations occur naturally and are not an artifact of placing the peptide at the interface, 60 short simulations of 20 ns length with the peptides starting in the aqueous phase in random orientations were performed. Indeed, within few nanoseconds of simulation time, the peptides diffused towards the interface and adopted the same stable orientations A and B. Interestingly, 75 % of the 'open' peptides initially adopted

orientation A at the interface, whereas 25 % initially interacted with the interface in orientation B. This shows that while establishing the first contact with the interface, there is an initial preference for orientation A. This preference is amplified after a few nanoseconds equilibration time at the interface, resulting in 97 % in orientation A versus 3 % in orientation B (Table 3.1). Based on these findings, the simulations of CDPs **2**-**8** were directly started from the water phase with random orientations. Again, only the two stable orientations A and B were observed (Table 3.1). Note that Table 3.1 displays the simulation time spent in the orientations over 50 simulations of 200 ns length. As we show later in this chapter, the relaxation timescales for this system are in the order of µs. Thus, at this stage, the reported values do not reflect equilibrium distributions but rather mimic an initial distribution after drug administration. For equilibrium populations, we refer the reader to Section 3.3.5.

TABLE 3.1: Fraction of simulation time spent in orientation A and B after equilibration for CDPs **1**-**8** in the 'open' and 'closed' conformation

Peptide ID	'open' orient. A	'open' orient. B	'closed' orient. A	'closed' orient. B
1	97%	3%	100%	0%
2	>99%	< 1%	100%	0%
3	76%	24%	100%	0%
4	88%	12%	100%	0%
5	96%	4%	100%	0%
6	54%	46%	100%	0%
7	57%	43%	100%	0%
8	60%	40%	87%	13%

Figure 3.4 shows representative simulations of CDP **1** in the stable orientations at the interface. The simulation in the top left panel was started from the 'closed' conformation in orientation A. This orientation was stable throughout the full simulation time

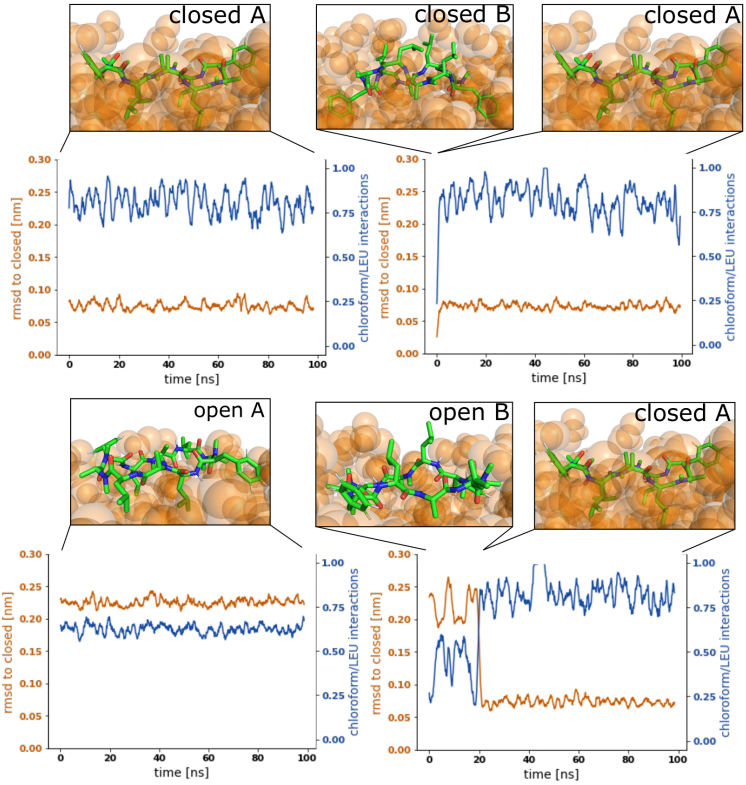


FIGURE 3.4: Representative simulations of CDP **1** at the water/chloroform interface. Blue lines show the fraction of leucine/chloroform interaction with respect to all leucine/solvent interactions as a measure for the peptide's orientation. The root mean square deviation (RMSD) of the peptide with respect to its 'closed' conformation is shown in orange. Snapshots of the simulations are provided to illustrate the location and orientation of the peptide towards the interface. Chloroform molecules are shown as orange spheres. Water molecules are not shown for visual clarity. Simulations were started from the following combinations of conformation and orientation. (Top left): 'closed'/A. (Top right): 'closed'/B. Bottom left: 'open'/A. Bottom right: 'open'/B.

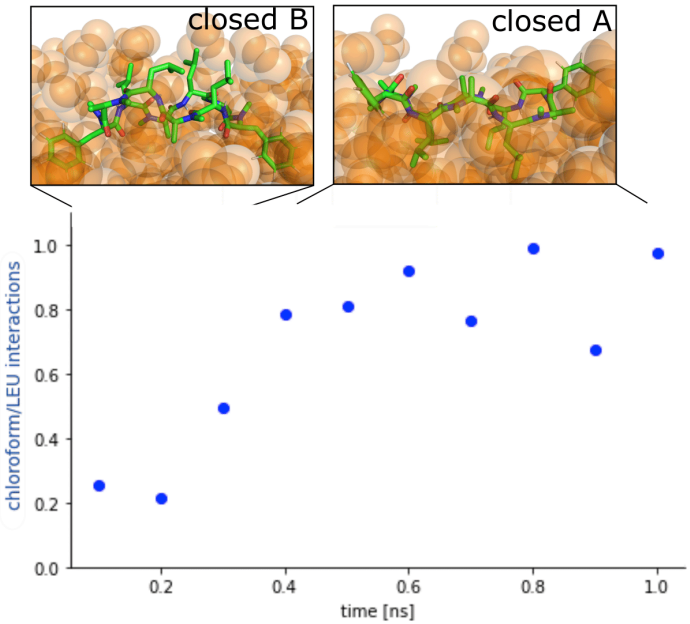


FIGURE 3.5: Zoom in of the first nanosecond of the rotation from orientation B to A for CDP **1** in the 'closed' conformation. Blue points show the orientation of the peptide with respect to the membrane measured by the side-chain contacts. In the simulation snapshots, the chloroform molecules are shown as orange spheres. Water molecules are not shown for clarity.

of 100 ns. In contrast, the simulation in the top right panel was started from the 'closed' conformation in orientation B. Within 1 ns of simulation time, the peptide rotates from orientation B to orientation A, indicating that orientation B is unstable in the 'closed' conformation. A higher resolution of the rotation process is shown in Figure 3.5.

Simulations started from the 'open' conformation are shown in the bottom panels of Figure 3.4. When the peptide was started from orientation A (left), the orientation was stable throughout the full simulation as in the 'closed' case. Interestingly, orientation B appears to be metastable in the 'open' conformation (right panel). After starting in orientation B, the peptide remained in this orientation for a few nanoseconds, before rotating as well towards orientation A. During the rotation, the peptide stayed in the 'open' conformation in some simulations while it closed in others (as in the example in Figure 3.4, a rotation from orientation B to A without closing is depicted in Figure 3.6). The rotation of the peptide occurs along its long axis and thus, the contacts with the chloroform phase are temporarily increased until orientation A is reached.

3.3.3 Composition Effects on the Orientation Preference at the Interface

From Table 3.1, it is evident that all CDPs favor orientation A at the interface. In the 'closed' conformation, all peptides except CDP **8** are even exclusively observed in orientation A. The stronger preference for orientation A in the 'closed' state in comparison to the 'open' state aligns well with the hydrophobic patches shown in Figures 3.2 and 3.3. In the 'closed' states, the leucine side chains form a large continuous hydrophobic patch,

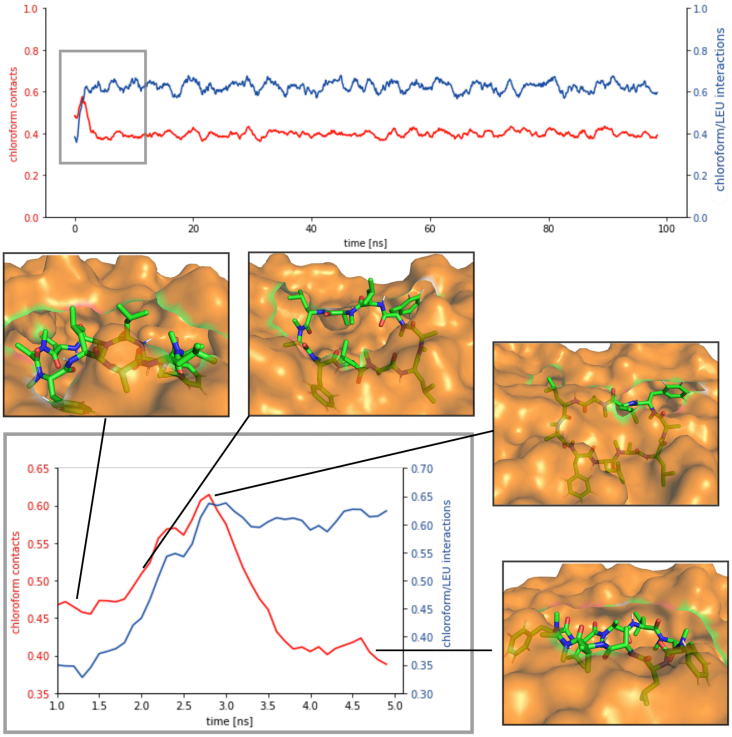


FIGURE 3.6: Representative simulation and zoom into the first 5 ns of the rotation from orientation B to A for CDP **1** in the 'open' conformation. Blue lines show the orientation of the peptide with respect to the membrane measured by the side-chain contacts. Red lines show the fraction of chloroform contacts with respect to all solvent contacts of the peptide. Simulation snapshots are shown at specific points of the simulation indicated by the black lines. Chloroform molecules are colored orange and shown in surface representation. Water molecules are not shown for clarity.

while the patch is broken up in the 'open' state. In orientation A, the hydrophobic patch faces the hydrophobic chloroform phase resulting in favorable interactions. These are stronger with the large continuous patch of the 'closed' state, leading to a stronger preference for orientation A. Hence, not only the overall amount of hydrophobic surface area is important but also its distribution in different conformers.

The effect of the hydrophobic patch can be confirmed by the orientation preferences of CDPs **6**-**8**, which lack half (two) or all (four) leucines side chains in comparison to the other peptides. In the 'open' conformation, CDPs **6** and **7** (lacking two leucines) have nearly equal preference for orientation A and B (Table 3.1). CDP **8** (lacking all four leucines) is the only peptide for which orientation B in the 'closed' conformation is observed. It is also the peptide with the lowest passive membrane permeability. Other amino acids also seem to influence the orientation preference. Phenylalanine even shows a titratable effect. Upon removal of both phenylalanine side chains from CDP **1**, the fraction of 'open' frames in orientation B drops from 3 % to <1 %. When removing one phenylalanine from CDP **3**, the fraction drops from 24 % to 12 %, while removing an additional phenylalanine leads to a further decrease to 4 %. The presence of proline, on the other hand, seems to reduce the preference for orientation A. CDPs **1** and **3** as well as **2** and **5** only differ in the presence of proline residues. In the first pair, the fraction of orientation B in the 'open' conformation increases from 3 % to 24 %, and in the second pair it increases from <1 % to 4 % upon introduction of the proline side chain. CDPs that contain proline residues have a more round shape in the 'open' states (compare Figure 3.2 and Figure 3.3). Therefore, the hydrophobic patch of those peptides is more fragmented in the 'open' state than for peptides without

proline. A more separated hydrophobic patch thus leads to a lower preference for orientation A.

The CDPs can only adopt distinct orientations at polar/apolar interfaces. The preference between the orientations is modulated by the amino-acid composition as well as the conformation of the peptide. This particular feature of cyclic peptides could only be observed by advancing from simulations in simple isotropic environments to anisotropic environments such as a polar/apolar interface. However, the question remains if the observations from the water/chloroform interface are transferable to a real lipid membrane system.

3.3.4 Comparison with Lipid Membranes

Lipid membranes and simple polar/apolar interfaces share important features like the directed hydrophobicity, but also differ in many points. Lipid membranes consist of a polar and often charged head-group region and an apolar lipid-tail region. Each layer of the membrane exhibits a dipole moment that points from the membrane's middle plane toward the head-group region. The head-group region is usually hydrated by water molecules. In comparison to chloroform molecules, the lipids are much larger, more bulky, and therefore not that easily displaced. The orientations A and B of the CDPs maximize the interface surface area occupied by the peptide (i.e., the peptide lies parallel to the interface), and are thus expected to distort the membrane surface more than a peptide in a tilted orientation. This raises the question how transferable the results from the water/chloroform system are towards a lipid membrane.

As an approximation for a biological membrane, we used a large patch containing 512 POPC lipids (256 in each layer). This

system was chosen to minimize finite size and buckling effects. Due to the increased size and complexity of the system, the simulation program was changed to GROMACS with the same force field (see Method section). To ensure that this change does not cause significant differences in the results, we compared the simulations of CDP **1** in water between the two simulation programs (the MSMs are displayed in Figure 3.7). The kinetic timescales as well as the fraction of 'closed' and 'open' conformational states were within a ratio of 1.5 of each other. In addition, the stationary distributions of the peptide overlapped in the TICA subspace with the exception of two minor conformational states (marked with red boxes). These small deviations can be explained by statistical fluctuations of the simulations or the difference between the twin-range versus single-range cutoff scheme used in the GROMOS and GROMACS software, respectively.

Simulations of the POPC patch with CDPs **1** and **3** were performed at room temperature without the addition of any bias. The two peptides were selected as representatives of the peptide classes with and without proline residues. The equilibrium population of 'closed' states in water is 45 % for CDP **3**, but only <1 % for CDP **1**.[89] Therefore, these peptides are good candidates to test the role of conformation preference for entering the membrane interface. We use no biases in the simulations in order to remain as realistic as possible and avoid membrane distortion or pore formation, unwanted effects that would alter the results. The simulations were initialized with peptides placed in the aqueous phase. In the chloroform/water simulations, the peptides moved to the interface within a few nanoseconds. In contrast, it took much longer in the lipid membrane system for the peptide to reach the membrane interface. In most simulations, the peptide remained in the aqueous phase entirely. Due

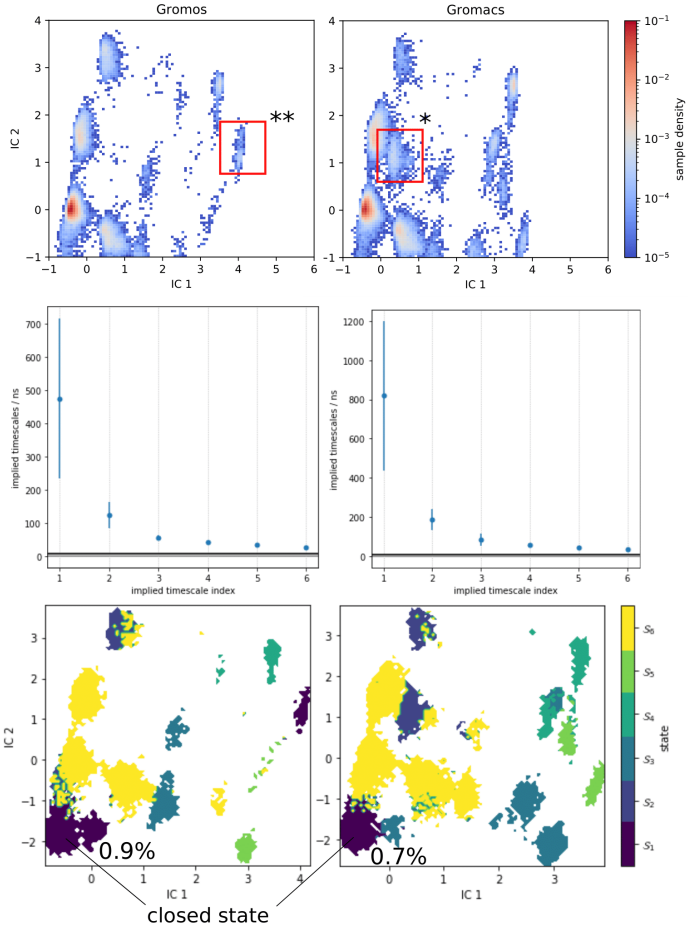


FIGURE 3.7: Comparison of Markov state model (MSM) results of CDP **1** in water with the GROMOS and GROMACS simulation software packages. (Top): Stationary distributions according to the MSM in the TICA subspace. The distribution overlap, except for two minor conformational states marked with red boxes. (Middle): Implied timescales of the MSMs. (Bottom): State assignment of the PCCA+ algorithm. The 'closed' state and its population is highlighted.

to their conformational flexibility, the peptides adopted conformations according to their equilibrium distribution during the elongated stay in the aqueous phase. CDP **3** adopted both 'open' and 'closed' conformations in the aqueous phase. Interestingly, the peptide was able to move to the membrane interface in both 'open' and 'closed' conformations. CDP **1** was nearly exclusively found in 'open' conformations in the water phase. Accordingly, only open conformations were seen inserting into the membrane interface.

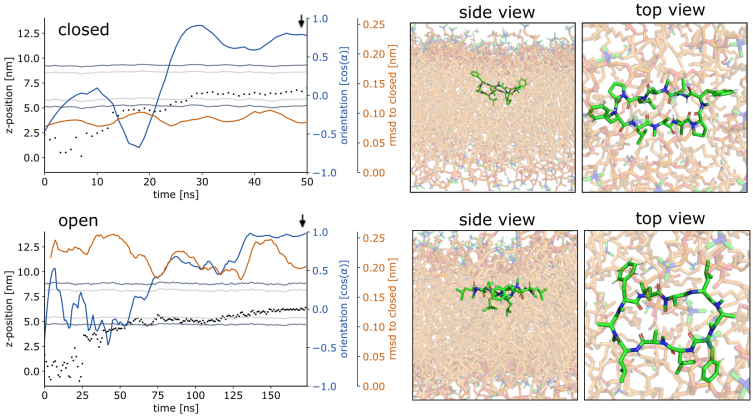


FIGURE 3.8: Representative simulations of CDPs **3** and **1** inserting into the membrane interface. (Top): CDP **3** in the 'closed' conformation. (Bottom): CDP **1** in the 'open' conformation. Black dots show the COM position of the peptide. The position of the membrane is indicated by dark grey (head groups) and light grey (tails) lines. The orientation of the peptide with respect to the membrane is shown in blue, α indicates the angle between the peptide normal and the membrane normal. The orange lines indicate the RMSD of the peptide with respect to its 'closed' NMR solution structure. The black arrow indicates the time point corresponding to the snapshots shown on the right. In the snaphots, the peptide is shown in green and the membrane in orange.

Figure 3.8 shows representative simulations of CDPs **1** and **3**

inserting into the membrane interface in the 'closed' and 'open' state, respectively. In the trajectory of CDP **3**, the peptide remained in the 'closed' conformation throughout the simulation. After an initial phase of diffusion in the water phase (0-15 ns), the peptide stayed in the proximity of the membrane for some time (15-25 ns), before entering the membrane in a tilted orientation (22-28 ns), and eventually moving deeper into the membrane and adopting orientation A (30 ns onward). The membrane thickness was unperturbed by the embedding of the CDP. The area per lipid, which was $0.637\,\mathrm{nm}^2$ without the CDP, dropped slightly ($0.02\,\mathrm{nm}^2$) upon CDP insertion, but relaxed back into the equilibrium value within a few ns. Orientation A was relatively stable over the rest of the simulation. Looking at the snapshots, one can confirm that the peptide is indeed in orientation A with the leucine residues pointing towards the apolar tail region. The lipids were thereby pushed aside to make room for the peptide. In the top view, one can observe that the peptide is not covered by the lipids but still has contacts with the aqueous phase through a water funnel. This water funnel together with the polar headgroups in the proximity of the peptide and the apolar tail region creates a local polar/apolar interface for the peptide. The environment at this interface is comparable with that observed in the water/chloroform system. In some frames in the 'closed' conformation (but never in 'open' conformations), lipids fully cover the peptide, leading to the disappearance of the water funnel and a slightly deeper penetration of the peptide in the membrane.

CDP **1**, which inserts into the membrane in the 'open' conformation, shows a similar trajectory. It exhibited more conformational dynamics but never closed. It also entered the membrane in a tilted orientation that is stable between 60-125 ns. Eventually, it adopts orientation A with the leucines pointing towards

the tail region (125 ns onward). In contrast to orientation A of the 'closed' conformation, the phenylalanine side chains point towards the membrane plane and not towards the head-group region. Nevertheless, the environment is very similar to that in water/chloroform and also shows the characteristic polar/apolar interface formed by the water funnel and the head groups.

Before fully entering the membrane, the peptides position themselves at the membrane/water interface. Our findings highlight that conformationally flexible peptides can do this in both 'closed' and 'open' conformations. In all simulations with peptides in the 'open' conformation and in a majority of the simulations with peptides in the 'closed' conformation, the peptide remains in contact with the aqueous phase through a water funnel and also in proximity of the head-group region. Thus, the peptide creates its own local interface environment. Especially peptides that preferentially adopt the 'open' conformation in water seem to insert into the membrane in the 'open' conformation. For peptides that predominantly enter the membrane in the 'open' conformation, the closing dynamics at the interface is potentially the decisive factor for their membrane permeability. Within our simulation time, the peptides appear to be trapped in one leaflet of the membrane, indicating that crossing the interior of the membrane is connected with a substantial energy barrier for the large and flexible CDPs. Unfortunately, the insertion events into the membrane are very rare. A total of 10 µs simulation time per peptide was necessary to sample two and three membrane insertion events for CDPs **1** and **3**, respectively. Additionally, simulating a big POPC patch is very computer resource intensive. Therefore, this analysis was only performed for CDPs **1** and **3** and no statistical evaluation of the distribution between orientation A and B was possible.

In summary, striking similarities in terms of positioning, conformation, and orientation preferences of the CDPs were observed between the water/chloroform interface and the membrane interface, indicating that the former model system can be used as an approximation for the conformational dynamics at lipid membranes. Due to its large size, the computational costs to simulate the membrane system are very high. Additionally, the dynamics of the peptides at the membrane are slower due to friction effects with the bulky lipids. Collecting sufficient data to construct an MSM would take considerably more simulation time than available for the membrane system in this chapter and is unfeasible for a larger number of peptides. Therefore, we continued to study the conformational dynamics of the CDPs using the simpler and faster equilibrating water/chloroform system, where it is possible to obtain sufficient data for the construction of MSMs.

3.3.5 Interface Catalyzes the Closing of Cyclic Peptides

Next, we investigated how the interface influences the conformational dynamics of the CDPs. Figure 3.11 shows a trajectory of CDP **1** interconverting from the 'open' to the 'closed' conformation at the interface. Interestingly, the backbone torsional angle ϕ of leucines 3 and 8 is highly correlated with the change in RMSD with respect to the 'closed' state, but not the backbone torsions of leucine 1 and 6 or any other residue. In the trajectory, we can see first half closing of the peptide (at approx. 10 ns) followed by a full closing event (at approx. 28 ns). For the half closing, the characteristic intramolecular hydrogen bonds are only formed on one side of the peptide (see Figure 3.11C and E). For the full closing, a change in the backbone torsional angle ϕ of leucines 3 and 8

of around 90°is necessary. In comparison, no closing events were seen in our previous simulations of CDP **1** in pure water (CDP **6** in Ref. 89,108). Similar closing trajectories were observed for all eight peptides (Figures 3.9 and 3.10 show a selection of these closing and opening events).

To better understand the kinetics and metastable conformational states of the CDPs, we built MSMs for each of the peptides. Previously, we used the backbone torsional angles as input features for the MSMs.[89,108] To take into account the role of the interface, we included two additional features: the position and orientation of the peptide relative to the interface. The position of the peptide with respect to the interface was measured by the ratio between the number of peptide-chloroform contacts and the total number of peptide-solvent contacts. The orientation was described with the angle between the peptide and the membrane normal. Four or three metastable conformational states were identified for CDPs **1**-**8**. The corresponding implied timescales are shown in Figure 3.20.

Figure 3.12B shows the metastable state assignment mapped on the orientation feature and the RMSD with respect to the 'closed' conformation. The model clearly distinguishes between a 'closed' state (blue) and two different 'open' states (red and orange). The two 'open' states are separated by their orientation with respect to the interface. The 'closed' state can adopt a larger variety of orientations. This first seems to be in contrast with the observed stability of only one orientation in the 'closed' state (Table 3.1). However, as discussed below, the 'closed' state can diffuse into the apolar phase. There, it experiences an isotropic environment, which leads to more orientational variety. Therefore, the 'closed' state has a highly preferable orientation while positioned at the interface but looses this preference once it diffuses

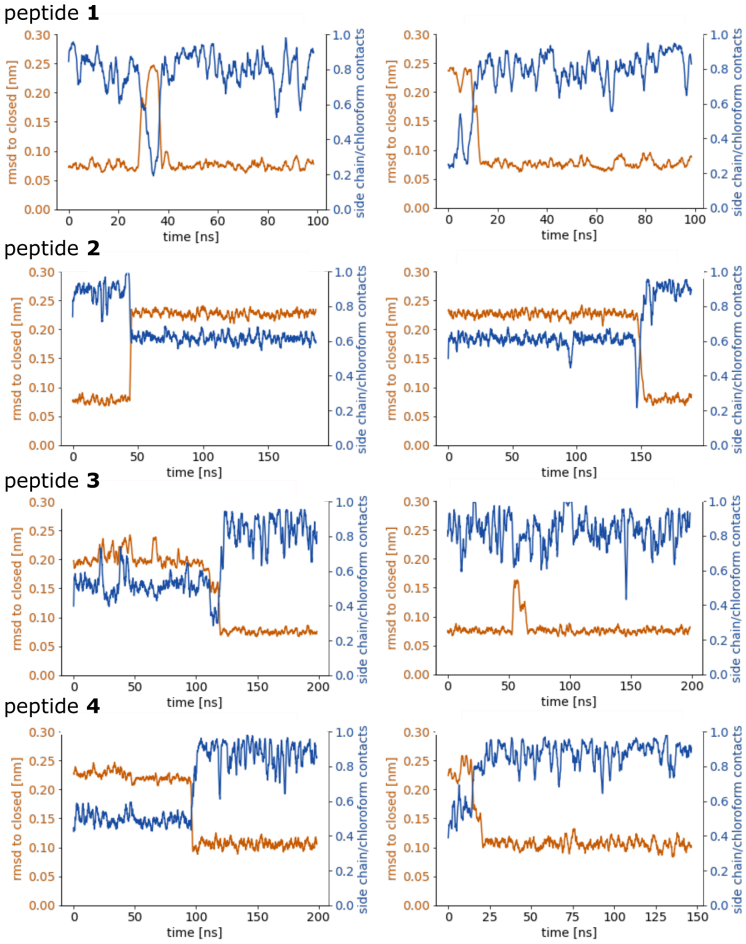


FIGURE 3.9: Trajectories with closing events for CDPs **1-4** at the interface. The orange line represents the RMSD with respect to the 'closed' NMR solution structure, and the blue line gives the orientation of the peptide measured by the side-chain contacts.

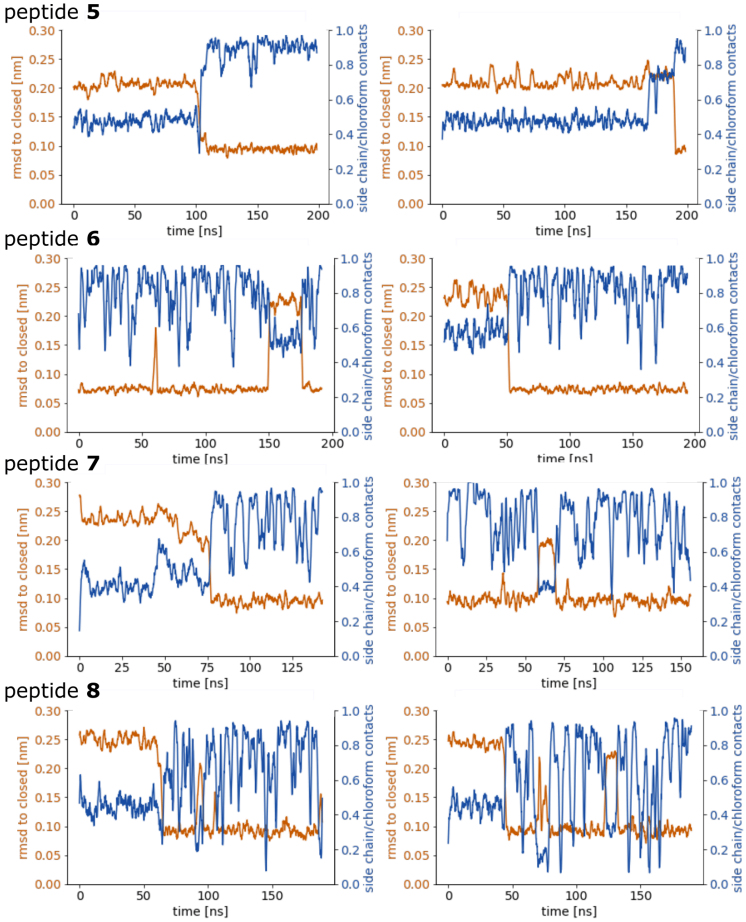


FIGURE 3.10: Trajectories with closing events for CDPs **5-8** at the interface. The orange line represents the RMSD with respect to the 'closed' NMR solution structure, and the blue line gives the orientation of the peptide measured by the side-chain contacts.

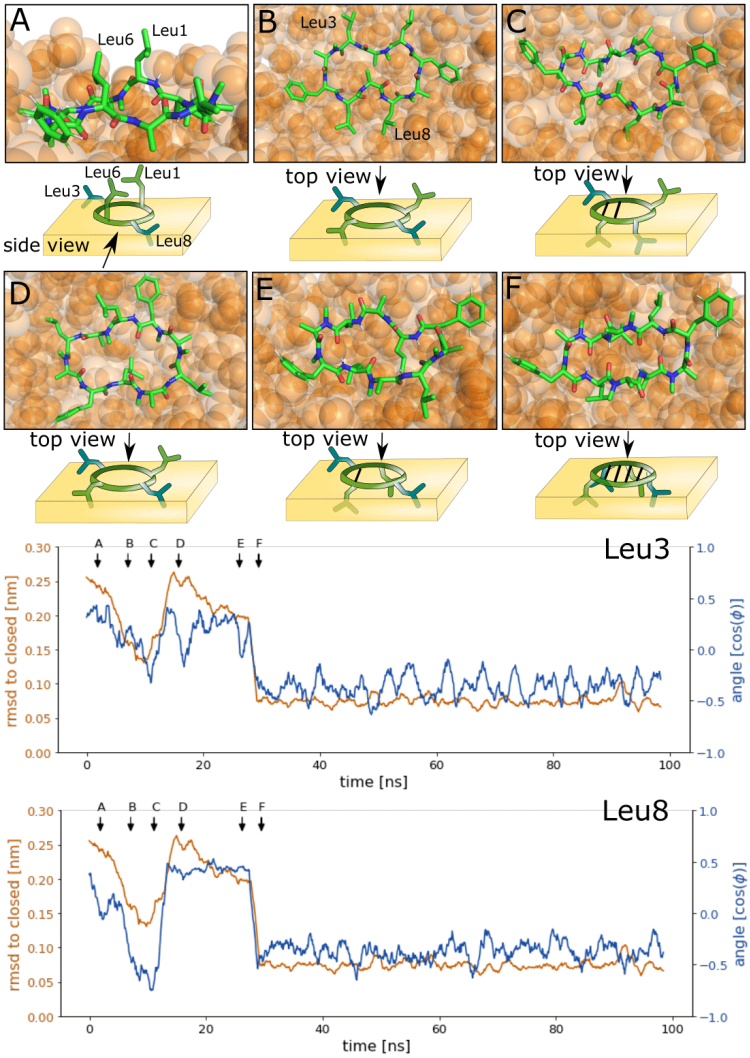


FIGURE 3.11: Interconversion of CDP **1** from 'open' to 'closed' conformation at the interface. The orange line represents the RMSD with respect to the 'closed' NMR solution structure, and the blue line the cosine of the backbone ϕ angle. Arrows indicate the time points of the snapshots shown on top.

into the apolar phase. Panel C shows the MSM states mapped on the position feature and the RMSD with respect to the 'closed' state. From this projection, it can be seen that only the 'closed' state (but not the 'open' one) can fully diffuse into the apolar phase (position = 1).

The potential energy contributions in the MSM states of CDP **1** are compared in Figure 3.12D and E. As expected from the equilibrium populations, the 'open' orientation B (state 1) has a higher energy (less favorable) than the 'open' orientation A (state 3). The higher energy of state 1 might facilitate the closing process to state 2. Panel E splits the energy into water, chloroform, and intramolecular contributions. In the case of CDP **1**, the peptide has more favorable interactions with both water and chloroform in orientation A (state 1) than in orientation B. For the peptides containing proline, the preference for orientation A can primarily be attributed to more favorable interactions with water molecules.

Figure 3.12 shows the MSM of CDP **1** based on the interface simulations. The results of all other CDPs are provided in Figures 3.13-3.19 and Figure 3.21.

Importantly, the kinetic model distinguishes for all peptides metastable states in orientation A and B for 'open' but not 'closed' conformations. The model was not biased to make this distinction. The orientation feature was treated as any other of the 42 MSM input features. Its significance for the kinetic model underlines the previous observation that it is important to consider both orientations for 'open' conformations. The equilibrium populations as well as the MFPTs between the metastable states are shown in panel A. Interestingly, the 'closed' state of CDP **1** is highly populated at the interface (43 %). In contrast, its 'closed' population in water is <1 %.[89] Thus, CDP **1** mainly adopts 'open'

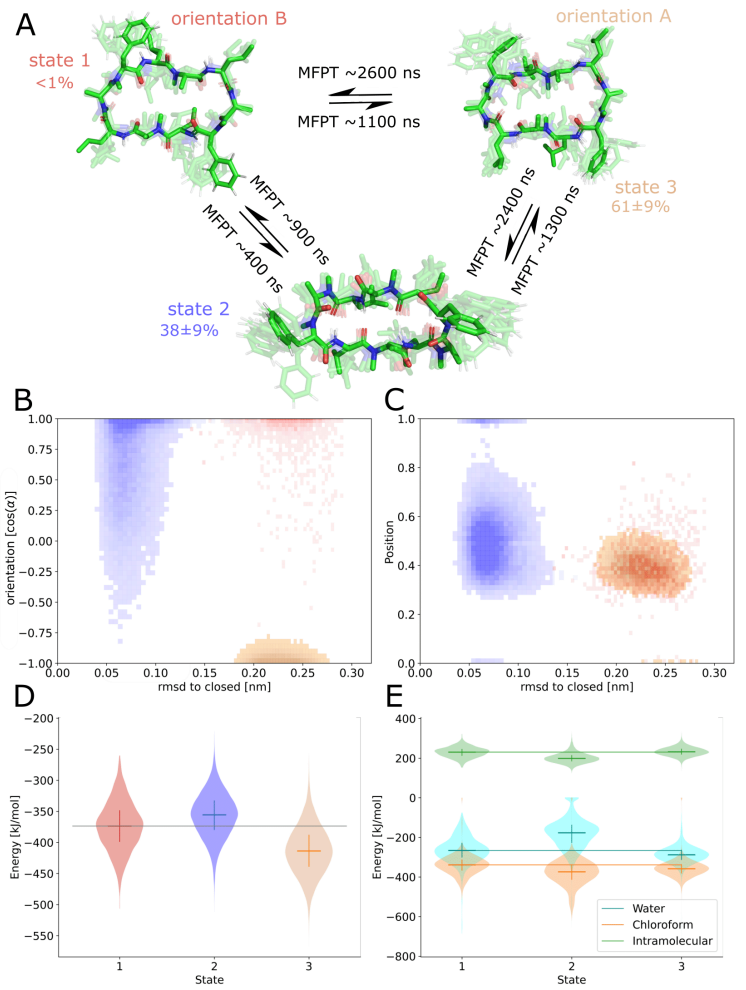


FIGURE 3.12: Markov state model (MSM) of CDP **1**. **A**: Representative members of the metastable states, their equilibrium populations as well as the mean first passage times (MFPTs) between them. **B**: Metastable state assignment mapped to the orientation feature and the RMSD with respect to the 'closed' state. **C**: Metastable state assignment mapped to the position feature and the RMSD with respect to the 'closed' state. **D**: Violin plot of the interaction energies of the metastable state members. **E**: Same as **D**, but split into water, chloroform, and intramolecular contributions. Horizontal lines are added to help guide the eye.

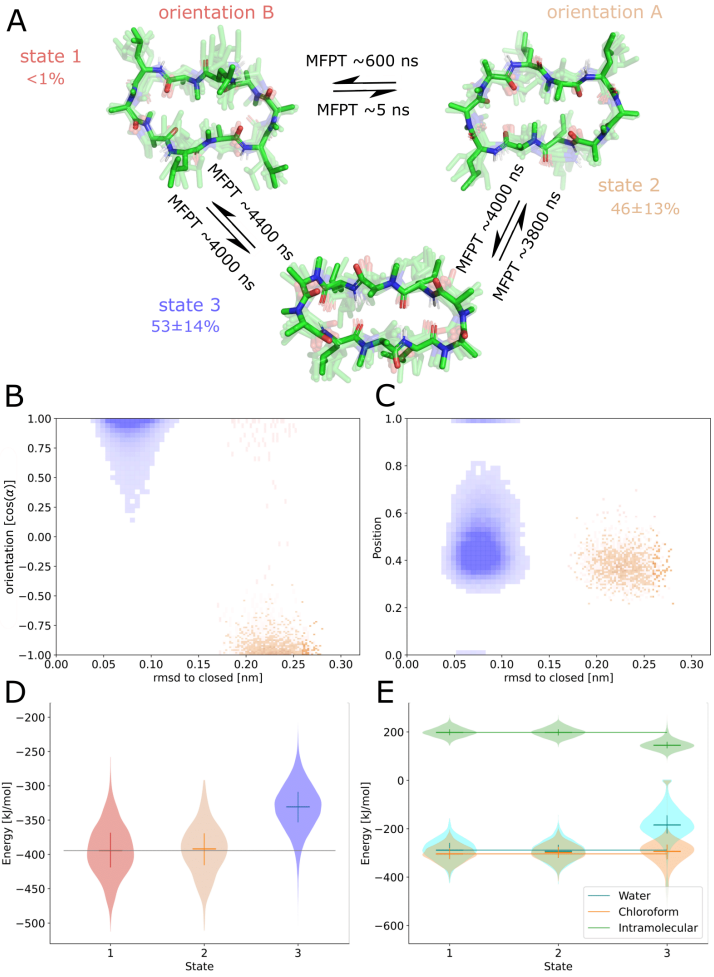


FIGURE 3.13: MSM of CDP **2** in the water/chloroform system. **A**: Representative members of the metastable states, their equilibrium populations and MFPTs between them. **B**: Metastable state assignment mapped to the orientation feature and the RMSD with respect to the 'closed' state. **C**: Metastable state assignment mapped to the position feature and the RMSD with respect to the 'closed' state. **D**: Violin plot with the interaction energies of the metastable state members. **E**: Same as **D**, but split into water, chloroform, and intramolecular contributions. Horizontal lines are added to guide the eye.

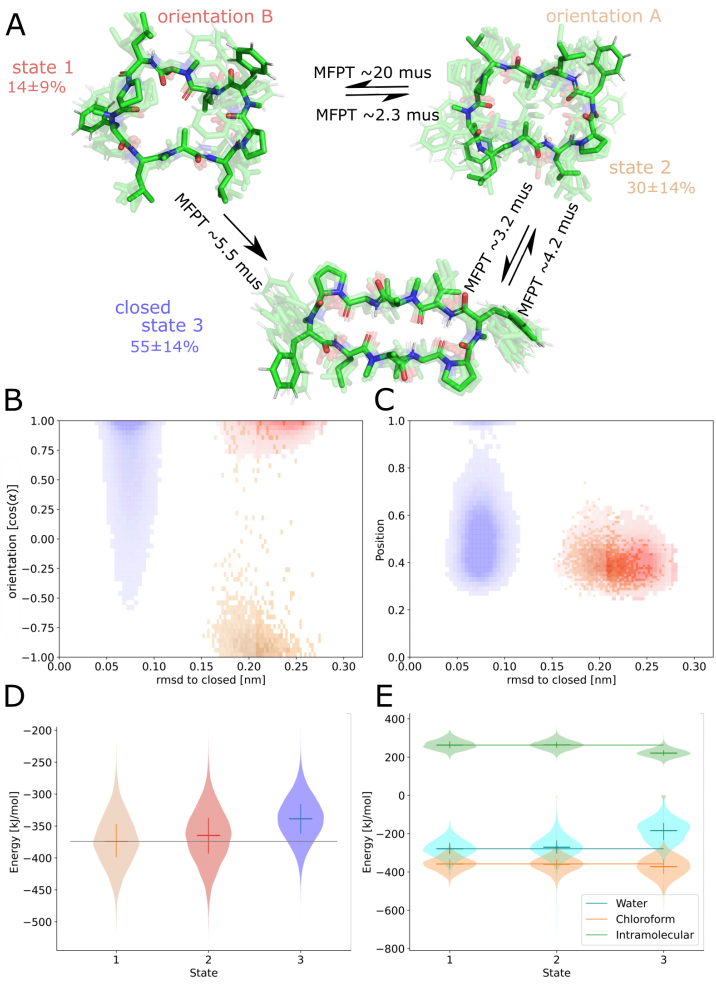


FIGURE 3.14: MSM of CDP **3** in the water/chloroform system. **A**: Representative members of the metastable states, their equilibrium populations and MFPTs between them. **B**: Metastable state assignment mapped to the orientation feature and the RMSD with respect to the 'closed' state. **C**: Metastable state assignment mapped to the position feature and the RMSD with respect to the 'closed' state. **D**: Violin plot with the interaction energies of the metastable state members. **E**: Same as **D**, but split into water, chloroform, and intramolecular contributions. Horizontal lines are added to guide the eye.

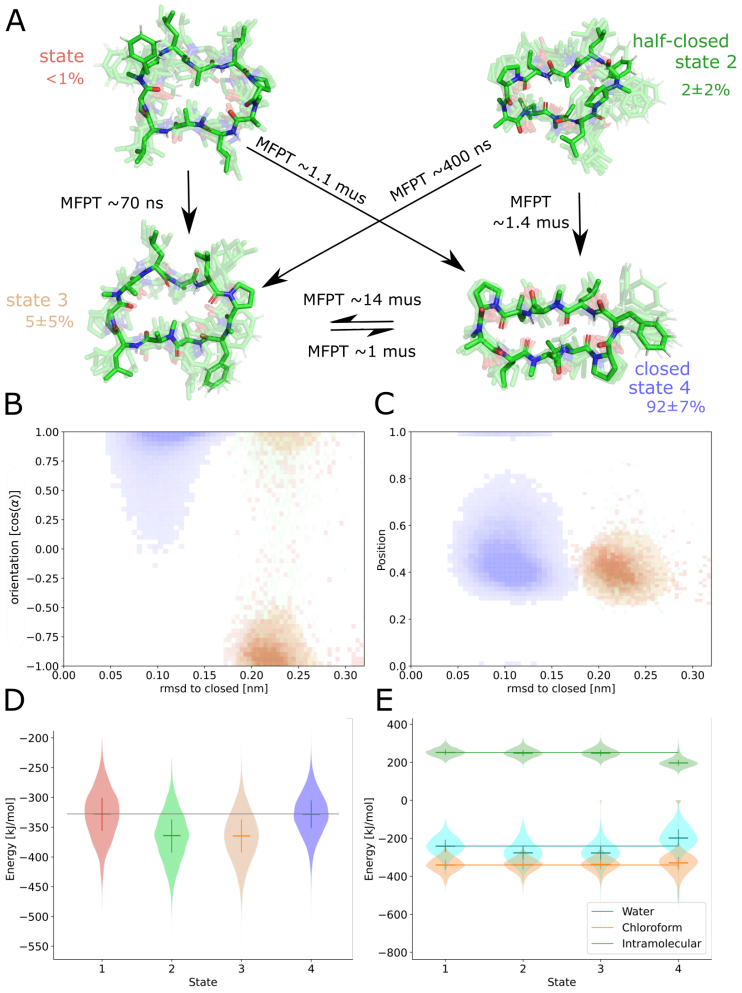


FIGURE 3.15: MSM of CDP **4** in the water/chloroform system. **A**: Representative members of the metastable states, their equilibrium populations and MFPTs between them. **B**: Metastable state assignment mapped to the orientation feature and the RMSD with respect to the 'closed' state. **C**: Metastable state assignment mapped to the position feature and the RMSD with respect to the 'closed' state. **D**: Violin plot with the interaction energies of the metastable state members. **E**: Same as **D**, but split into water, chloroform, and intramolecular contributions. Horizontal lines are added to guide the eye.

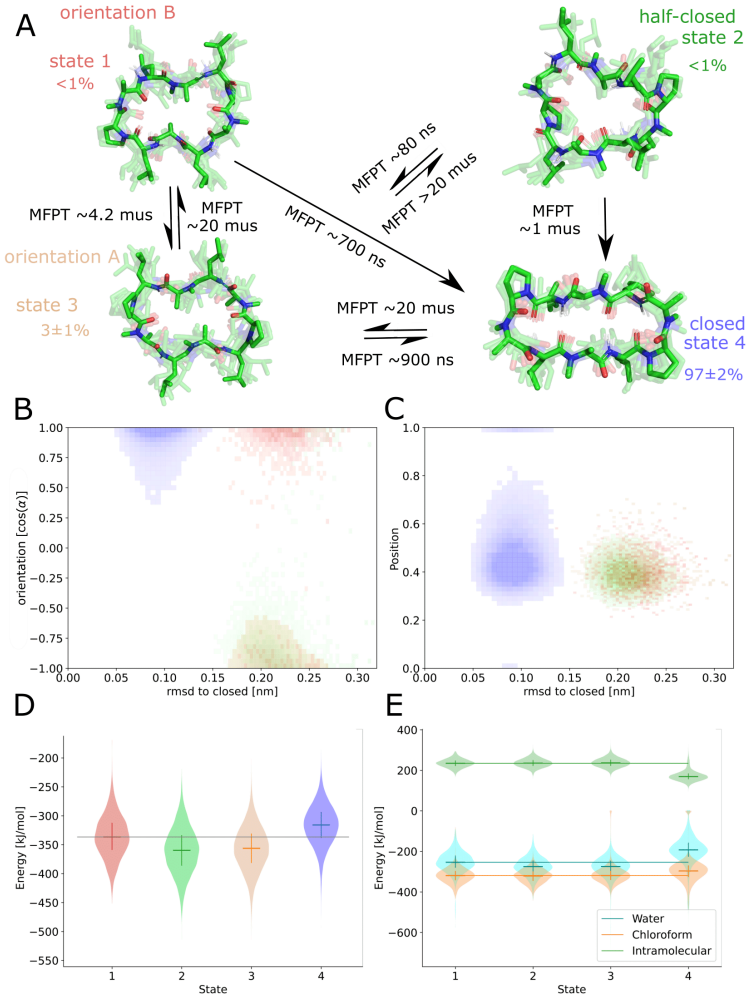


FIGURE 3.16: MSM of CDP **5** in the water/chloroform system. **A**: Representative members of the metastable states, their equilibrium populations and MFPTs between them. **B**: Metastable state assignment mapped to the orientation feature and the RMSD with respect to the 'closed' state. **C**: Metastable state assignment mapped to the position feature and the RMSD with respect to the 'closed' state. **D**: Violin plot with the interaction energies of the metastable state members. **E**: Same as **D**, but split into water, chloroform, and intramolecular contributions. Horizontal lines are added to guide the eye.

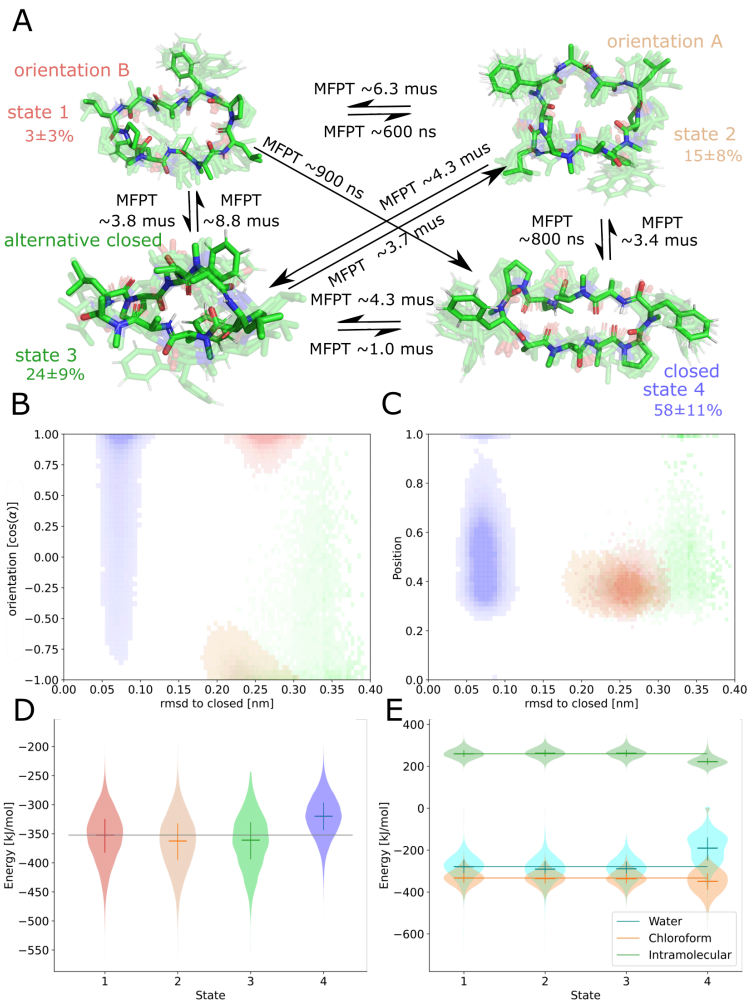


FIGURE 3.17: MSM of CDP **6** in the water/chloroform system. **A**: Representative members of the metastable states, their equilibrium populations and MFPTs between them. **B**: Metastable state assignment mapped to the orientation feature and the RMSD with respect to the 'closed' state. **C**: Metastable state assignment mapped to the position feature and the RMSD with respect to the 'closed' state. **D**: Violin plot with the interaction energies of the metastable state members. **E**: Same as **D**, but split into water, chloroform, and intramolecular contributions. Horizontal lines are added to guide the eye.

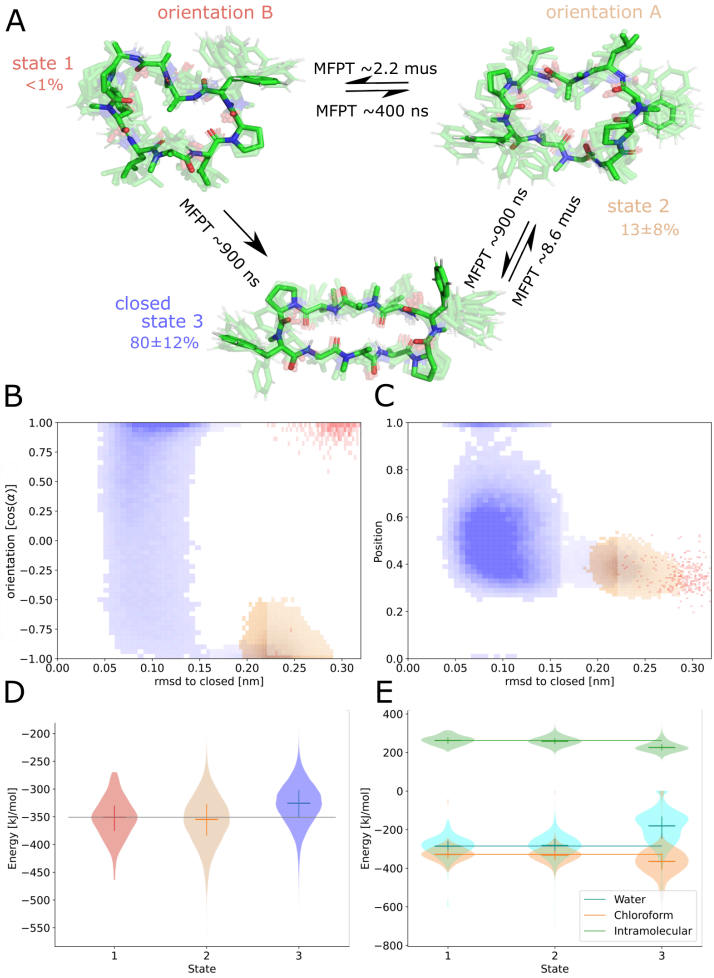


FIGURE 3.18: MSM of CDP **7** in the water/chloroform system. **A**: Representative members of the metastable states, their equilibrium populations and MFPTs between them. **B**: Metastable state assignment mapped to the orientation feature and the RMSD with respect to the 'closed' state. **C**: Metastable state assignment mapped to the position feature and the RMSD with respect to the 'closed' state. **D**: Violin plot with the interaction energies of the metastable state members. **E**: Same as **D**, but split into water, chloroform, and intramolecular contributions. Horizontal lines are added to guide the eye.

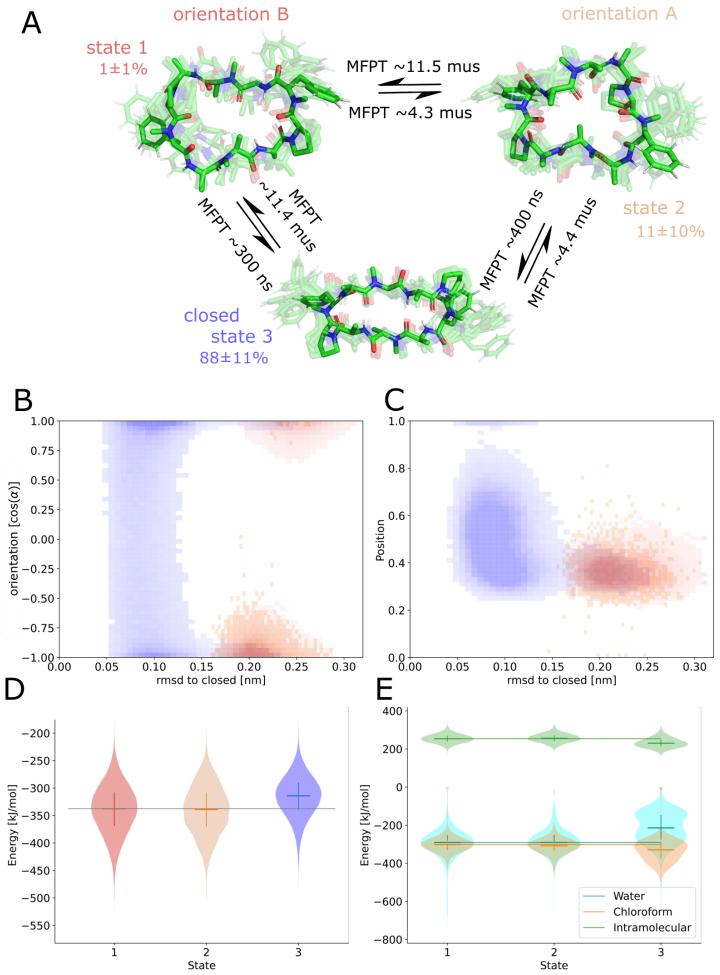


FIGURE 3.19: MSM of CDP **8** in the water/chloroform system. **A**: Representative members of the metastable states, their equilibrium populations and MFPTs between them. **B**: Metastable state assignment mapped to the orientation feature and the RMSD with respect to the 'closed' state. **C**: Metastable state assignment mapped to the position feature and the RMSD with respect to the 'closed' state. **D**: Violin plot with the interaction energies of the metastable state members. **E**: Same as **D**, but split into water, chloroform, and intramolecular contributions. Horizontal lines are added to guide the eye.

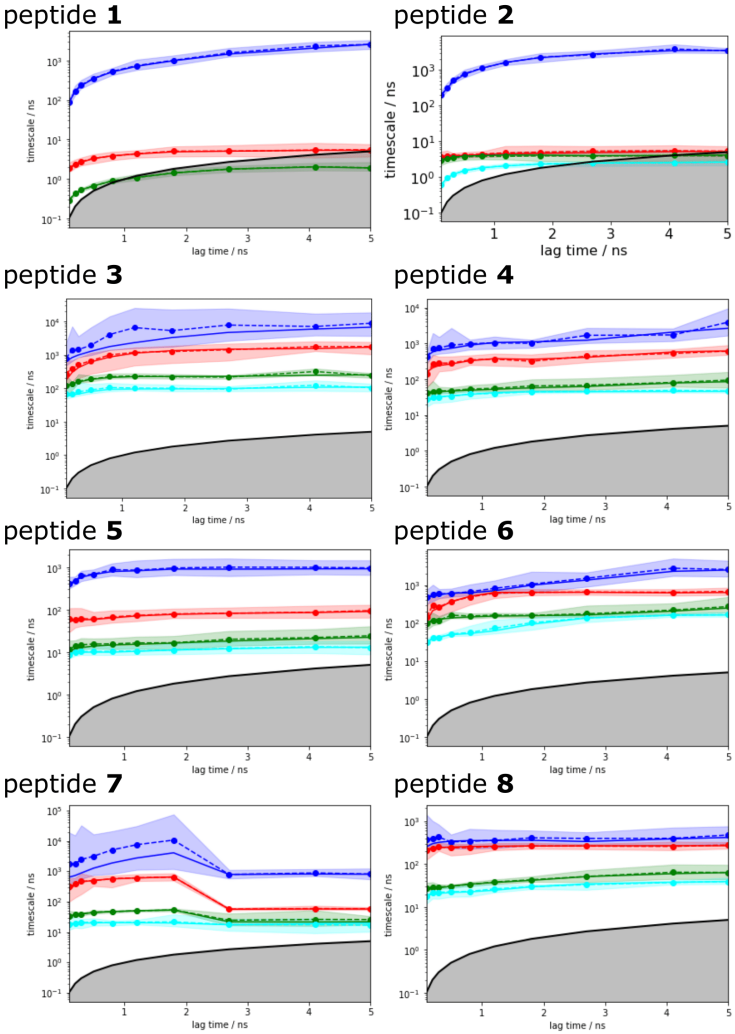


FIGURE 3.20: Implied timescales of the MSMs of CDPs **1**-**8**.

conformations in water and also likely inserts into the interface in an 'open' conformation (compare to Figure 3.8). Once at the interface, the equilibrium and kinetics of CDP **1** are shifted compared to the aqueous phase. The closing process becomes much faster. Thus, the interface can act as a catalyst for the interconversion between 'open' and 'closed' conformations. Additionally, the MFPT for the closing process is much shorter for orientation B than for orientation A. In our equilibrium model, orientation B is hardly populated. However, we have found that in 25 % of cases CDP **1** initially interacted with the interface in orientation B (Section 'Peptide Orientations at Interfaces'). Therefore, in a real world non-equilibrium scenario where the cyclic peptide is administered as a drug, the importance of orientation B may be higher than anticipated by the equilibrium model.

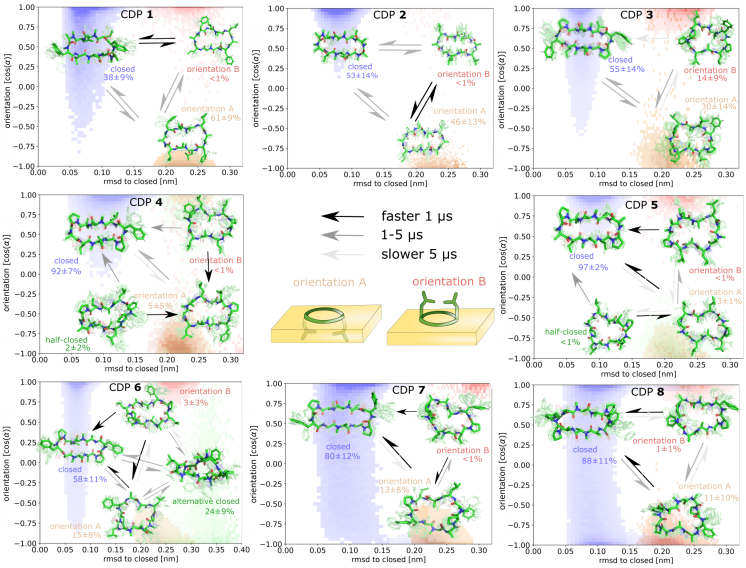


FIGURE 3.21: Markov state models (MSMs) of CDP **1**-**8**. Representative members of the metastable states, their equilibrium populations as well as the mean first passage times (MFPTs) between them are shown on top of the simulation data points mapped to the orientation feature and the RMSD with respect to the 'closed' state. Note that the simulation data points represent non-equilibrium data while for the MSMs the equilibrium populations are shown.

Figure 3.21 displays an overview of the MSMs of all eight peptides. The kinetics and equilibrium populations of the metastable states are shown on top of the simulation data points mapped to the features RMSD to the 'closed' state and the orientation. The simulation data points were colored according to their metastable state assignment. All CDPs showed a distinct separation between at least three metastable states: the 'closed' conformations and the 'open' conformation in orientation A and B. CDPs **4**-**6** populate an additional metastable state that could be assigned to either the 'half-closed' or an 'alternative closed' conformation. For all eight CDPs, the 'closed' state was highly populated in equilibrium with the interface present (>38 % for peptides without proline and >55 % for peptides with proline). Additionally, orientation A was always higher populated, and thus energetically more favorable, than orientation B.

We were able to show that the unique environment of the interface alters the conformational equilibrium and favors the 'closed' state. This effect is especially strong for peptides that rarely close in aqueous solution. For these peptides, the interface effect can increase the fraction of 'closed' conformations by a factor > 50. In addition, peptides starting from an 'open' conformation in orientation A and B have different closing dynamics with faster closing in orientation B. Next, we will have a closer look at how the 'closed' peptides behave at the interface.

3.3.6 Permeable Conformations

Proceeding from the peptide in the 'closed' conformation at the interface, a total of ∼500 events of peptides diffusing into the chloroform phase were counted. These events will be referred to as 'membrane diffusions' in the following. All diffusion events start

from the 'closed' orientation A with the leucine side chains interacting with the apolar phase. In most cases, the peptide rotates along its long axis until the leucine side chains point towards the aqueous phase. Then, it diffuses into the apolar phase until it is fully surrounded by chloroform molecules. An example trajectory of this process is provided for CDP **1** in Figure 3.22. Not a single event of 'membrane diffusion' in the 'open' conformation was observed. This is in line with the findings reported in Ref. 80 for cyclosporine A.

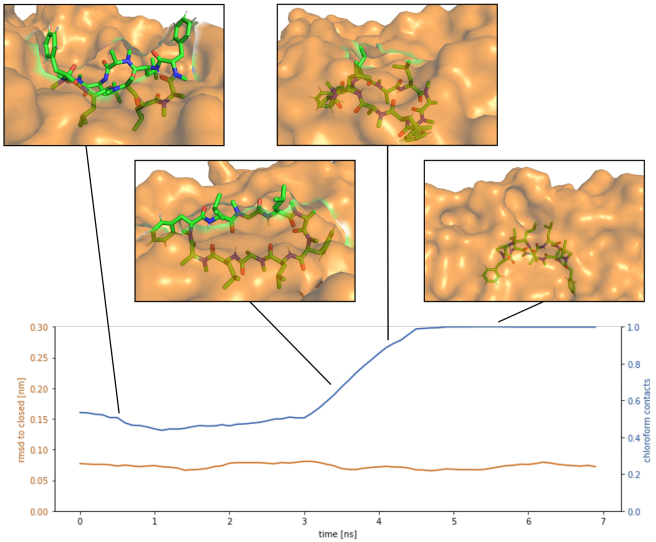


FIGURE 3.22: Representative apolar diffusion trajectory of CDP **1**. The orange line shows that the peptide remains in the 'closed' conformation throughout the diffusion process. The blue line indicates the fraction of peptide-chloroform contacts of the total peptide-solvent contacts. The diffusion event starts in orientation A. Next, the peptide rotates along its long axis until the leucine side chains point towards the aqueous phase. Then, it diffuses into the apolar phase until it is fully surrounded by chloroform molecules.

Interestingly, an 'alternative closed' state was identified for CDP **6** (Figure 3.23). In this 'alternative closed' state, the phenylalanine and proline side chains build a cage structure and the backbone resembles a twisted 'eight'. Four intramolecular hydrogen bonds stabilize the structure with the center carbonyl oxygen contributing to two hydrogen bonds. The MSM revealed that the 'alternative closed' state has a significant equilibrium population of around 24 % (Figure 3.17). Therefore, it might contribute notably to permeability. The existence of this 'alternative closed' state showcases the versatility of cyclic peptides and highlights the need to have exhaustive knowledge of their conformational behavior.

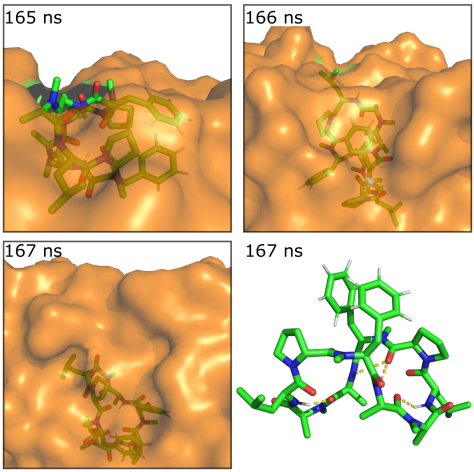


FIGURE 3.23: Apolar diffusion process of the 'alternative closed' state of CDP **6**. In this 'alternative closed' state, the phenylalanine and proline side chains build a cage-like structure. The backbone conformation resembles a twisted 'eight'. Four intramolecular hydrogen bonds stabilize the structure with the center carbonyl oxygen contributing to two hydrogen bonds. The peptide diffuses into the apolar phase in a side orientation with one leucine residue pointing upwards.

The finding that 'membrane diffusion' events only occur in the 'closed' or 'alternative closed' conformations strengthens the long-standing hypothesis that large and flexible cyclic peptides can permeate only when they can adopt conformations in which the polar groups are shielded by intramolecular hydrogen bonds. Although cyclic peptides in the 'open' state are mainly located at the interface, they can not fully immerse into the apolar phase.

3.4 CONCLUSION

In this chapter, we investigated the conformational and kinetic behavior of CDPs at a polar/apolar interface. All examined peptides preferred to be located at the interface over a position in pure polar or apolar solvent. Due to their directed hydrophobicity (Figure 3.2), the CDPs adopted two distinct orientations with respect to the interface: the hydrophobic patch formed by the leucine residues pointed either towards the apolar (orientation A) or polar phase (orientation B). The existence of these orientations was evident from both visible inspection of the simulations (Figure 3.4) and the MSMs (Figure 3.12). The same conformations and orientations were also observed in the POPC bilayer system (Figure 3.8). Interestingly, the local environment of the peptides in the membrane matched the environment of the water/chloroform interface. On one side the peptide was in contact with the apolar lipid tails, while a water funnel and the head-group region of the lipids formed a polar surrounding on the other side. The equilibrium populations and interaction energies of orientation A and B showed that orientation A is clearly preferred. Interestingly, this preference was modulated by both

the conformation and the amino-acid composition of the CDPs (Table 3.1). These modulations can be linked to changes in the hydrophobic patch on the peptide surface. A larger continuous patch led to a stronger preference for orientation A. Peptides in orientation B showed a faster closing dynamics, probably due to the higher (less favorable) energy of this orientation (Figure 3.12). Although the equilibrium population of orientation B was rather low in the MSMs, a significant fraction of CDPs that started in the water phase initially docked to the interface in orientation B. Thus, in a non-equilibrium scenario, the importance of orientation B might be higher than anticipated by the equilibrium kinetic models.

The MSMs furthermore revealed that the unique environment of the interface not only led to distinct orientations but also influenced the conformational equilibrium and kinetics of the CDPs. Importantly, the presence of the interface facilitated the closing process for all examined peptides. Especially for CDP **1**, which rarely adopts the 'closed' conformation in aqueous solution, the equilibrium population of the 'closed' state was increased by a factor of approximately 50 (i.e. from <1 % to 38 %). Thus, the interface might function as a catalyst for the closing process. This is especially relevant for peptides with low 'prefolding' in water. Furthermore, we were able to explicitly show that the peptides can only diffuse into the apolar phase in the 'closed' conformation, marking it as the 'permeable species'.

We showed that cyclic peptides can insert into the interface between the head-group and tail regions of the membrane in both the 'open' and 'closed' conformations (Figure 3.8). This, in combination with the observed catalytic ability of the interface, leads to the proposition of a refined hypothesis for membrane permeation (Figure 3.24). An equilibrium exists between 'open' and

'closed' states in water. Peptides with a significant equilibrium population of the 'closed' conformation in water (i.e. 'prefolding') can insert into the interface directly in the 'closed' state, and subsequently diffuse into the apolar phase (blue + black route). Peptides with no or low population of the 'closed' conformation in water may still be able to permeate, although along a different route. They insert into the interface in the 'open' state. The interface modulates the equilibrium between 'open' and 'closed' conformations, facilitating the closing process. Once in the 'closed' conformation, the peptide is able to diffuse into the apolar phase (red + black route). One can further speculate that different amino acids contribute differently to the closing process in water and in a lipid bilayer. Bulky residues may hinder the dynamics due to steric clashes inside the membrane but not in water. Future work is necessary to test whether different design principles apply for 'prefolding' and 'non-prefolding' cyclic peptides. In particular, the influence of different amino-acid compositions on the closing dynamics has to be explicitly tested in the presence of a membrane, as the water/chloroform interface is not able to mimic the steric hindrance caused by the lipids. Thus, future work will show the predictive power of the proposed model.

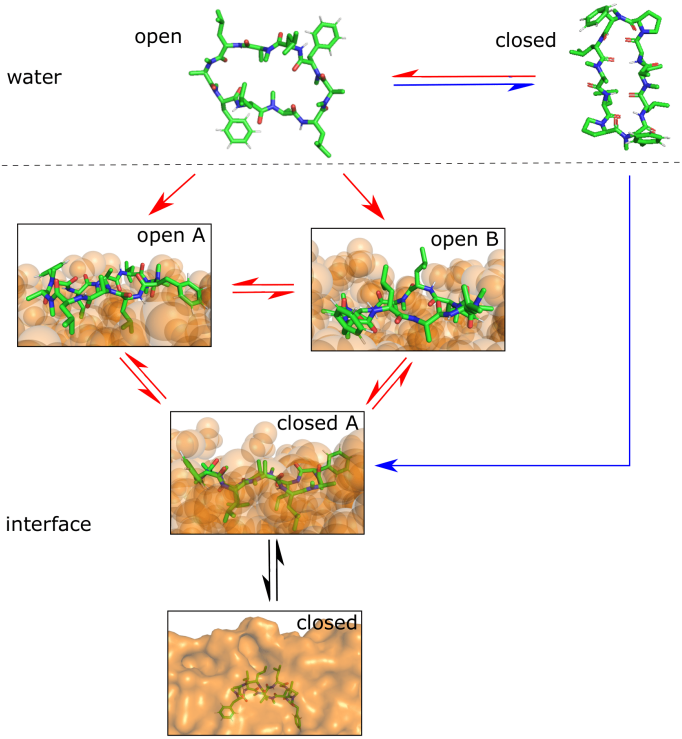
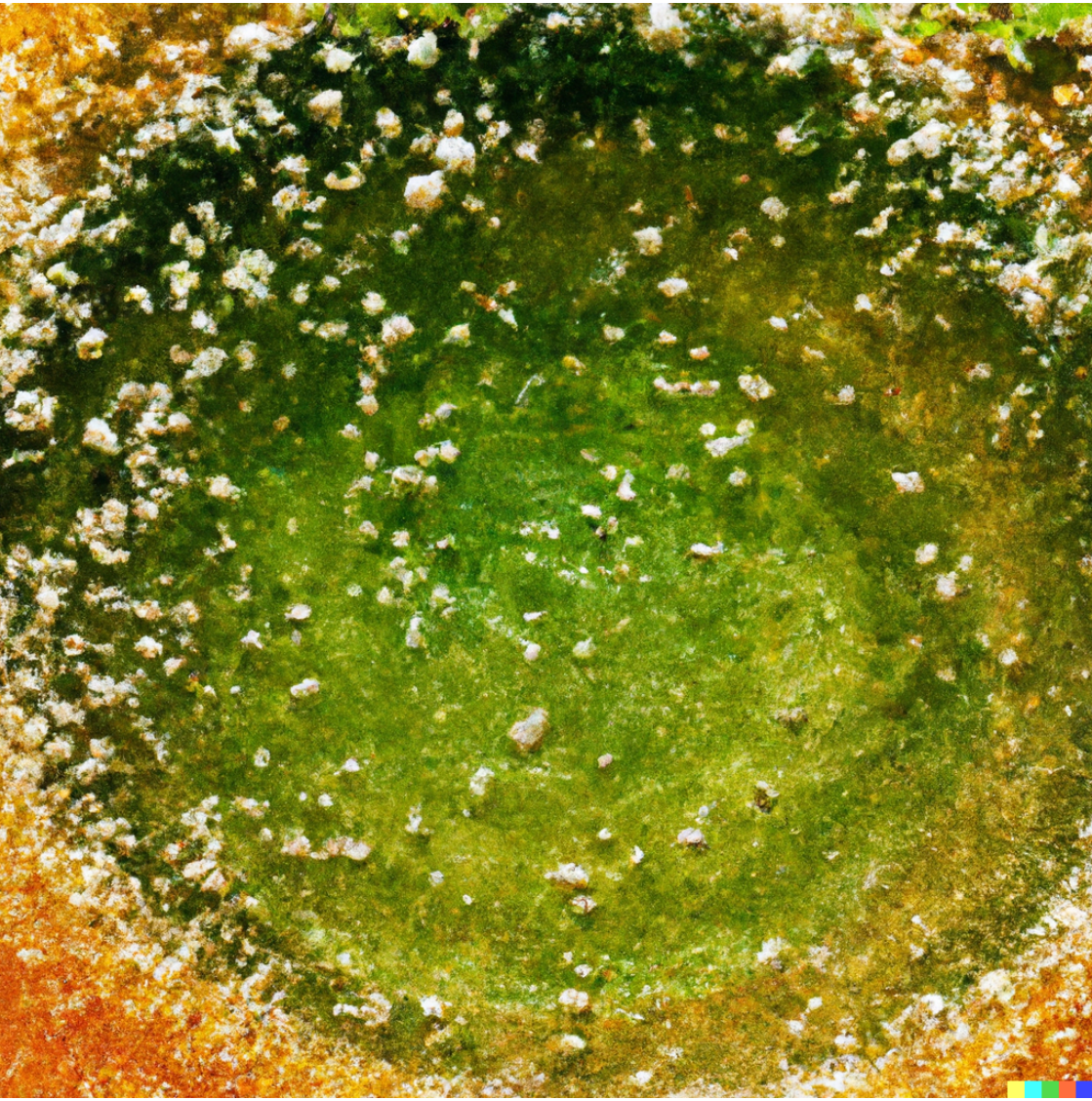


FIGURE 3.24: Hypothesis for the permeation process of 'prefolding' (closing in water) and 'non-prefolding' (closing in the membrane) cyclic peptides. Peptides with a significant population of the 'closed' conformation in water can insert into the interface or membrane directly in the 'closed' state (blue route). Peptides with a no or a low population of the 'closed' conformation in water insert into the interface in the 'open' state. There, the conformational dynamics are modulated by the presence of the interface, resulting in a new equilibrium between 'open A', 'open B' and 'closed' states. The interface environment acts as a catalyst, which facilitates the closing process (red route). Once in the 'closed' conformation at the interface, the peptide can diffuse into the apolar phase (black arrows).



4

Lessons for Oral Bioavailability: How Conformationally Flexible Cyclic Peptides Enter and Cross Lipid Membranes *

"We build too many walls and not enough bridges" and *"Niemand hat die Absicht, eine Mauer zu errichten"*

Isaac Newton and Walter Ulbricht

This quote and chapter are dedicated to Benjamin Ries and his ability to build bridges where others left walls

* This Chapter is reproduced in part from S. M. Linker†, C. Schellhaas†, A. S. Kamenik, M. M. Velhuizen, F. Waibl, H.-J. Roth, M. Fouché, S. Rodde, S. Riniker, *J. Med. Chem.*,66, (2023), 2773–2788 with permission of ACS Publications. †These authors contributed equally.

Cyclic peptides extend the druggable target space due to their size, flexibility, and hydrogen-bonding capacity. However, these properties impact also their passive membrane permeability. As the "journey" through membranes cannot be monitored experimentally, little is known about the underlying process, which hinders rational design. Here, we use molecular simulations to uncover how cyclic peptides permeate a membrane. We show that side-chains can act as 'molecular anchors', establishing the first contact with the membrane and enabling insertion. Once inside, the peptides are positioned between head groups and lipid tails – a unique polar/apolar interface. Only one of two distinct orientations at this interface allows for the formation of the permeable 'closed' conformation. In the closed conformation, the peptide crosses to the lower leaflet via another 'anchoring' and flipping mechanism. Our findings provide atomistic insights into the permeation process of flexible cyclic peptides and reveal design considerations for each step of the process.

4.1 INTRODUCTION

Macrocyclic compounds like cyclic peptides represent promising candidates to address difficult drug targets.[121,125,157–161] Traditional small-molecule drugs typically show sufficient binding affinity only for relatively deep and narrow protein binding sites.[61,118] Consequently, this limits their target space to an estimated 10-15% of the human proteome.[11,116,117] In comparison to small-molecules, macrocyclic compounds hold the potential to vastly extend the scope of druggable proteins. Due to their potential to target larger binding sites with flat profiles,[62,119–122,124] like protein-protein interaction (PPI) interfaces,[123,162,163] they can lead to new therapeutics for currently untreatable diseases. However, high binding affinity is not enough for a molecule to be pharmaceutically relevant. One crucial aspect in drug design is drug delivery, where passive membrane permeability is important for oral bioavailability of drugs. Key steps of drug delivery like the gastrointestinal absorption and passing the portal venous system,[65,164] predominantly rely on trans-cellular diffusion.[165] Moreover, a majority of drugs with intracellular targets also pass the final cell-membrane barrier via passive diffusion.[71] Macrocycles often possess high molecular weights that are associated with low passive permeability,[70] hampering the therapeutic applications of macrocyclic compounds.[69,70,94,97]

Cyclic peptides are macrocycles that are composed of amino acids.[63] In comparison to their linear counterparts, they are associated with better passive cell membrane permeability and metabolic stability.[63,69,166] Interestingly, there are examples of cyclic peptides that can be administered orally,[13,64] despite violating conventional drug-likeliness rules based on their increased molecular

weight and high number of hydrogen-bonding atoms.[70,71,97,167,168] Nonetheless, designing orally bioavailable cyclic peptides with simultaneous high binding affinity to the target has been difficult so far, and is often only achieved via a tedious trial-and-error process.[169–172] Addressing this issue, previous studies have substantially advanced our understanding of the structure-permeability relationship of cyclic peptides, and thus provide first guidance for their design as therapeutics with oral bioavailability.[70,114,126,127,129,166,173–176] N-methylation of the peptide backbone,[128,177,178] changes of stereocenters,[179,180] tuning the amphiphilicity,[181] and side-chain modifications[108,182] were identified as membrane permeability factors. The effects of these modifications are unfortunately non-linear and highly site-dependent.[98,106,108,130–132] Even small structural modifications can lead to global conformational rearrangements and thus change the physico-chemical properties and membrane permeability of a compound.[161,183,184] Coherently, it has been shown that the conformational behavior of a cyclic peptide in different environments is particularly impactful for passive permeability.[16,89,95,96,101,102,104,115,185]

Cyclic peptides with so-called 'chameleonic' behavior are able to adapt to their environment and adopt different conformational states that exhibit varying lipophilic properties.[12,16,100–104,186] These conformational states typically can be distinguished by their number of intramolecular hydrogen bonds: In so-called 'open' conformations, hydrogen-bonding atoms are exposed, allowing for formation of favorable contacts with polar solvents (e.g., in the blood stream or the cytosol). In the so-called 'closed' conformation, intramolecular hydrogen bonds are formed, which leads to a less polar surface area and a lower desolvation energy for entering apolar environments like the cell membrane interior. Intuitively, this behavior of chameleonic cyclic peptides is beneficial for oral

bioavailability, as good permeability is combined with good solubility.[187]

It is generally assumed that the 'closed' conformation is the main permeable species.[16,72,89,95,102,108] However, little is known about the structural origin of the chameleonic properties, let alone the mechanistic details of the path of a chameleonic cyclic peptide through the cell membrane with respect to its conformational behavior. While the composition (amino-acid sequence), size, and hydrophobic surface have been identified as important determinants of chameleonic conformational behavior,[70,89,98,105,130] their mechanistic interplay is complex and not yet understood well enough to elucidate structure–permeability relationships. Addressing this challenge, molecular dynamics (MD) simulations can serve as a computational microscope to track the pathway and conformational dynamics of cyclic peptides in a time and spatial resolution that is not (yet) feasible with current experimental techniques. In combination with experimental data (e.g., PAMPA membrane permeability coefficients[188]), this enables the development of a holistic view of the passive permeability of cyclic peptides.[185]

In this chapter, we perform extensive all-atom MD simulations of a series of eight cyclic decapeptides that show complex internal conformational dynamics in a 1-palmitoyl-2-oleoylphosphatidylcholine (POPC) bilayer system to decipher their pathway through the membrane and rationalize the relationship between structure and passive membrane permeability. Previous approaches have either focused on simulations in homogeneous solvents,[89,108] at very high temperatures,[80] or under application of steered pulling forces.[114,189] Here, we report an alternative sampling strategy where insights from unbiased MD simulations are used to seed biased simulations that allow for a step-wise enrichment of key

events across the permeation pathway. Note that this strategy enabled us to describe the membrane crossing pathway in an unbiased manner at room temperature, avoiding artifacts like pore formation or other distortions of the POPC bilayer.

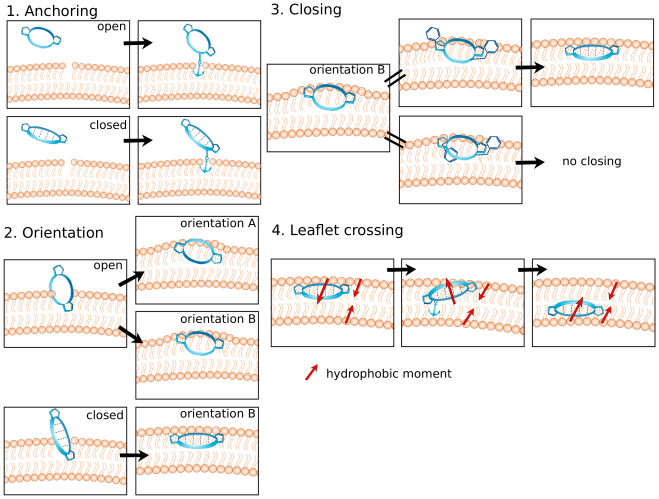


FIGURE 4.1: Summary of the four steps for the passive membrane permeation of conformationally flexible cyclic peptides. The peptide is shown in blue and the membrane in orange. Peptides in the 'closed' conformation are indicated by their four hydrogen bonds (dashed lines). Side-chain residues that anchor the peptide in the membrane are depicted with an anchor symbol. The hydrophobic moment is pointing from the polar to the apolar part of a molecule.

Our key findings are summarized in Figure 4.1, highlighting the four steps, which we identified, for the passive membrane permeation of conformationally flexible cyclic peptides. First, specific side-chain residues can act as 'molecular anchors', which establish the contact between a cyclic peptide and a membrane before insertion. Second, the peptide positions itself directly at the interface between the polar head groups and the apolar tail

region. There, the cyclic peptides show a preference for one of two distinct orientations. Third, we observed conformational interconversion into the permeable 'closed' state from only one of these orientations. Last, only this 'closed' conformation can detach from the polar/apolar interface and diffuse across the lipid membrane leaflets, which again requires a unique anchoring and flipping mechanism. Based on these steps, we identified multiple design considerations and opportunities that can improve the development of cyclic peptides with oral bioavailability.

4.2 METHODS

4.2.1 MD Simulations

The simulations were carried out with the Groningen Machine for Chemical Simulations (GROMACS) 2020.5 software package[143] in combination with the GROMOS 54A8 force field[138] and the POPC model of Marzuoli *et al.*[144] Simulations were performed under periodic boundary conditions with a leapfrog integration scheme[32] and a time step of 2 fs. Peptides, lipids, and solvent were coupled to three separate thermostats at 303 K using a weak coupling scheme[37] and a relaxation time of 0.1 ps. A semi-anisotropic Parrinello-Rahman barostat[145] at 1.0 bar with a coupling constant of 2.0 ps and isothermal compressibility of 0.45 nm^2/N was used. Long-range electrostatics were treated using the particle mesh Ewald algorithm.[146] Bond lengths were constrained using the linear constraint solver (LINCS) algorithm.[147] A center-of-mass (COM) motion removal was applied every step to eliminate the movement of the bilayer relative to the solvent.

In order to avoid finite size effects, a relatively large lipid patch containing 512 POPC molecules was used. The CDP starting structures in the 'closed' conformation were obtained by NMR spectroscopy.[135] The protocol to obtain the starting structures in the major 'open' conformation has been described in detail in Refs. 89,108 and Chapter 3. The CDPs were initially positioned in the aqueous phase approximately 3 nm away from the head-group region of the POPC bilayer. The system was solvated with the SPC[139] water model, and equilibrated using a 100 ps NVT thermalisation and 1 ns NPT equilibration. Unbiased production runs (NPT) had a length of 100 ns.

4.2.2 Biased simulations

Pulling simulations were performed using the GROMACS internal pull code. Peptides started in the aqueous phase approximately 3 nm away from the head-group region of the POPC bilayer. A constant pulling force of $50\,\mathrm{kJ\,nm^{-1}\,mol^{-1}}$ was applied on the distance between the COM of the selected anchor amino acid and the COM of the POPC lipids. The pulling force was chosen small enough so that, if applied on the COM of the peptide, it did not perturb the membrane or lead to a membrane insertion event (see Figure 4.2). To increase anchoring events, five pulling simulations with a length of 5 ns each were performed for every leucine, phenylalanine, and proline residue of the respective CDP. For the pulling simulations of leaflet crossing, twenty simulations with a length of 10 ns each were performed for every leucine, phenylalanine, and proline residue of the respective CDP.

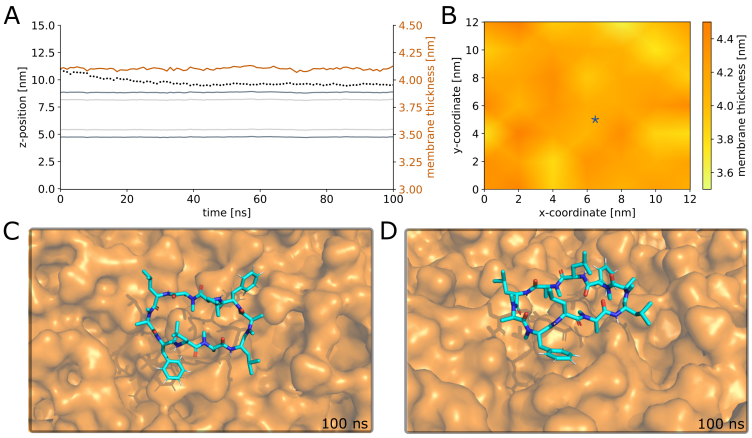


FIGURE 4.2: Pulling on the center-of-mass (COM) does not lead to membrane insertion or membrane crossing. A COM pulling force identical to the anchor pulling force was applied to the peptide. (**A**): The peptide is pulled to close proximity of the membrane but does not enter. The average membrane thickness over time is shown in orange. (**B**): Heatmap of the membrane thickness for simulation frame 100 ns. The position of the peptide is indicated with a blue asterisk. No substantial fluctuations in membrane thickness were observed around the peptide. (**C**): Top view of the peptide with COM pulling towards the membrane at frame 100 ns. The membrane surface is shown in orange. (**D**): Same as **C** but in side view.

4.2.3 Data Analysis

If not stated otherwise, all trajectories were analyzed using the Python library MDTraj.[149]

Position. The z-coordinate of the COM was used to indicate the peptide position. As a reference, the head-group position and the beginning of the tail group of the POPC bilayer are shown in the figures. The average z-coordinate of the head groups was calculated using the nitrogen position of the choline group. The starting of the tail group was calculated using the carbon atom succeeding the ester group.

Orientation with respect to the membrane. The peptide normal $\vec{a}$ was determined using the cross product of the major and minor axis of the peptide backbone. As the membrane position remains approximately fixed throughout the simulation, the z-axis $\vec{b}$ was used as an approximation for the normal axis of the membrane. The angle between these normal vectors was calculated as,

$$\text{Orientation} = \alpha = \arccos\left(\frac{\vec{a}\cdot\vec{b}}{|\vec{a}||\vec{b}|}\right). \tag{4.1}$$

RMSD. The atom-positional backbone root-mean square deviation (RMSD) with respect to the NMR solution structure of the 'closed' conformation was calculated using the Python library MDTraj.[149]

Treating periodic boundary conditions. Due to the periodic boundary conditions, peptides can access the membrane from above and below. For consistency and didactic reasons, all trajectories were transformed such that the peptide entered the membrane at the upper leaflet. Thus, both the position and the orientation were shifted by a 180° rotation of the simulation box

around the x-axis if the peptide entered from below.

4.2.4 Biasing Simulation for the Phenylalanine Lock

Pulling simulations were performed using the GROMACS internal pull code for dihedral angles. An umbrella pulling force with a pulling rate of 0.1 nm and a force constant of 500 kJ nm^{-1} mol^{-1} was applied on the phenylalanine's backbone ψ angle. In total 40 pulling simulations with a length of 20 ns were performed. This resulted in two successful lock removal events.

4.2.5 Distribution of Head-Group Gap Size

The hydrophobic residues anchor to the membrane via gaps between the lipid head groups. To better understand this process, we simulated a pure POPC membrane as well as a membrane containing a molar fraction of 30 % cholesterol. Consistently with the rest of this work, each membrane contained 512 POPC molecules. The second system contained additional 216 cholesterol molecules, which resembles the 30% cholesterol fraction observed in mammalian cells. This resulted in a 163.5 nm^2 sized membrane patch for pure POPC and a 169.5 nm^2 sized membrane patch for POPC+cholesterol (both after equilibration.) After an equilibration phase of 50 ns, we simulated each membrane for 150 ns, saving frames every 100 ps, and computed the occurrence of head-group gaps. As the average lifespan of a gap is below 20 ps,[190] we expect only weak correlations between consecutive frames.

We used Packmem[190] to analyze the occurrence of gaps. Packmem creates a 2D grid of 0.1 nm resolution along the x/y-coordinates

of a leaflet of the membrane. At each chosen x/y-coordinate, the highest z-coordinate is determined, which overlaps with the van der Waals radius of a membrane atom. If this atom belongs to the hydrophobic part of POPC or chlosterol, the respective grid point is defined to be in a head-group gap. The hydrophobic part of POPC was defined as all atoms below the carbon bound to the phosphate group, and that of cholesterol as everything but the C-OH group.

The area of a head-group gap was computed by searching continuous patches of tiles that are in a gap. The occurrence of head-group gaps of each size was normalized by the number of frames times two, because there are two leaflets. This gives us the probability of finding a head-group gap of a given size in a single MD snapshot.

The minimum area needed to accommodate a side chain anchor was computed by noting that the broadest part of the sidechain are the two methyl groups in front, and by modelling their area as the obround shape defined by two circles of 0.2 nm radius and a distance of 0.25 nm. We performed an exponential fit of the data to determine the decay constant for both membrane setups. As recommended in Ref.,[190] we only considered head-group gaps $> 0.15\,\mathrm{nm}^2$ for the fit.

4.2.6 Construction of Markov State Models

Markov state models (MSMs) allow for the calculation of equilibrium quantities and long-time kinetics from ensembles of short simulations.[22] MSM require 'local equilibrium' within the MSM states, but not a 'global equilibrium' between all MSM states.

The PyEMMA Python library[148] was used to construct the MSMs. The sine and cosine of theϕ and ψ backbone torsional

angles as well as the position with respect to the membrane center and the orientation of CDP **4** with respect to the membrane normal were extracted from the trajectories. This resulted in 42 input features, which were reduced to 12 collective coordinates by time-lagged independent component analysis (TICA).[150] The trajectory was clustered using k-means clustering with 30 states. A lagtime τ of 10 ns was chosen to ensure Markovianity. To group the microstates into metastable conformational states, robust Perron cluster-cluster analysis (PCCA+)[156] was performed.

4.2.7 Data and Software Availability

An example Jupyter notebook to analyse the CDP trajectories is available on the rinikerlab GitHub (`https://github.com/rinikerlab/decapeptides-membrane`). This repository also contains the topology and structure files of the POPC system as well as topology files of the CDPs and the 'open' and 'closed' starting structures used in this chapter. The freely available software can be obtained via the following links: GROMACS (`https://www.gromacs.org/`) and PyMol (`https://github.com/schrodinger/pymol-open-source`).

4.3 RESULTS AND DISCUSSION

To investigate the permeability pathway of flexible cyclic peptides, we performed extensive all-atom MD simulations of a series of cyclic decapeptides (CDPs) at room temperature. The backbone scaffold and N-methylation pattern of the CDPs was introduced by Fouché *et al.*[135,136] and kept constant. The variable amino acids in this series are highlighted in color in Figure

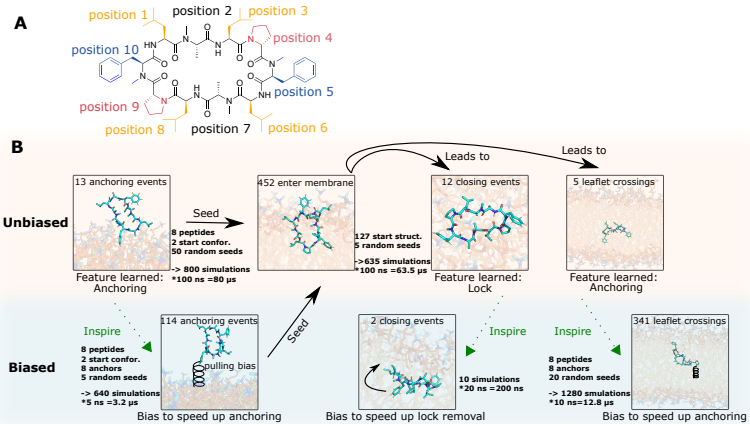


FIGURE 4.3: (**A**): Backbone scaffold and amino acid composition of the cyclic decapeptide (CDP) series used in this chapter. The colored residues were systematically replaced according to Table 4.1. The backbone scaffold was reported by Fouché *et al.*[135,136] and is kept constant. (**B**): Schematic workflow showing the different conditions, total simulation time, and number of observed events. Unbiased MD simulations were used to elucidate the membrane permeation pathway of CDPs. Biasing along this pathway was used to enrich sampling and to obtain starting structures for new unbiased simulations.

4.3A and listed in Table 4.1. The peptides are characterized by two β-strands (residues 1-3 and 6-8) and two β-turns (residues at positions 4, 5 and 9, 10). In the 'closed' conformation, four intramolecular hydrogen bonds are formed that shield the polar groups of the CDP from the environment.[89] The NMR solution structures and passive permeability data for the CDPs were taken from the literature.[89,108,115,135]

The permeation process of the CDPs was tracked through cycles of unbiased and biased simulations. Unbiased simulations were used to learn the features important for permeability. This knowledge was utilized to define the collective variable for biasing. From the end states of the biased simulations, new unbiased simulations were started (see Figure 4.3 B). Unless clearly marked otherwise, all results and conclusions in this paper were drawn from the analysis of the unbiased simulations.

TABLE 4.1: Amino-acid composition of the CDPs used in this chapter. D-amino acids are marked with the letter 'D', methylated amino acids are marked with the letter 'M'. The amino acids at position 2 and 7 (MAla) were kept constant. The parallel artificial membrane permeation assay (PAMPA) coefficients were taken from Ref. 115. CDP **1** and **8** with $\log P_e$ <-5.5 can be considered non-permeable.

CDP	1	3	4	5	6	8	9	10	PAMPA
Pos.									
1	Leu	Leu	DAla	MPhe	Leu	Leu	DAla	MPhe	-5.9
2	Leu	Leu	DAla	MAla	Leu	Leu	DAla	MAla	-4.0
3	Leu	Leu	DPro	MPhe	Leu	Leu	DPro	MPhe	-5.3
4	Leu	Leu	DPro	MPhe	Leu	Leu	DPro	MAla	-4.2
5	Leu	Leu	DPro	MAla	Leu	Leu	DPro	MAla	-4.6
6	Ala	Leu	DPro	MPhe	Ala	Leu	DPro	MPhe	-4.1
7	Ala	Ala	DPro	MPhe	Leu	Leu	DPro	MPhe	-4.4
8	Ala	Ala	DPro	MPhe	Ala	Ala	DPro	M-he	-6.4

4.3.1 Cyclic Peptides Enter Lipid Membranes Using 'Anchor' Residues

The simulations of the CDPs started in the aqueous phase at a distance of around 3 nm from the membrane and were allowed to freely diffuse through the aqueous phase. Within the sampled 100 ns per simulation, only a small fraction of the simulations resulted in peptide-membrane contacts. Interestingly, all these interactions followed the same unique mechanism: First, transient and small gaps between the head groups of the membrane formed due to thermal fluctuations. During this transient time span, the apolar lipid tails are exposed to the aqueous phase. Figure 4.4A visualizes the opening of these gaps, where the solvent-accessible surface of the membrane is color coded as head groups (blue) or tail region (orange). If, by chance, the CDP is in close proximity to such a gap, its apolar residues can interact with the exposed lipid tails and the peptide is stabilized at the water-membrane interface. Hence, we find that an apolar side chain acting as a molecular anchor is a key feature for this contact initiating step (see Figure 4.4C). However, as the head-group gaps are short-lived and small, contact formation happens on average only once per 6 µs simulation time. he area per lipid is not significantly perturbed upon entry of the peptide (see Figure 4.5).

Figure 4.4B depicts a simulation of a CDP anchoring and entering the lipid bilayer in more detail. To better track the position of the peptide, the dark grey points indicate the position of the upper head-group region and the light grey points indicate the beginning of the apolar tails as defined in Section 4.2.3. For the first 16 ns, the peptide diffuses freely in the aqueous phase. Then, the first membrane contact is established. The peptide stays anchored to the membrane for 26 ns. In this anchored position, the

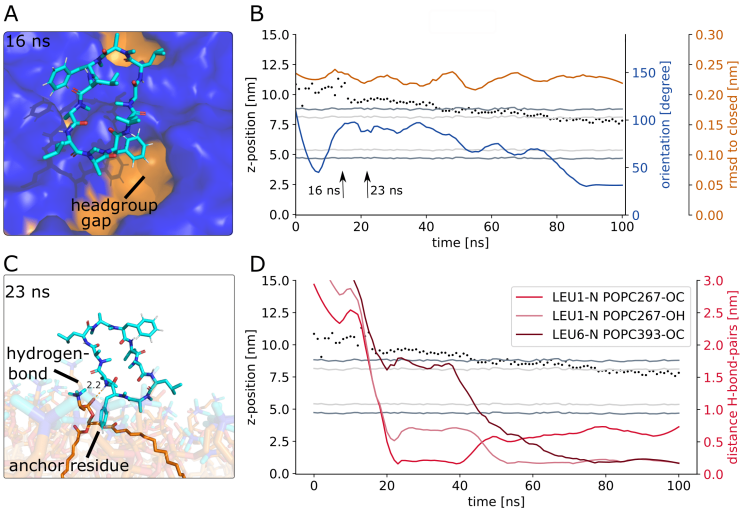


FIGURE 4.4: (**A**): Snapshot of the CDP **1** directly before anchoring to the membrane. The atoms of the POPC head groups are colored in blue, the atoms of the tails are colored in orange. Thermal fluctuations of the lipids can lead to temporary head-group gaps and the apolar tails underneath become exposed (in this snapshot the gaps manifest as orange patches). If the CDP is close to such a transient gap, its apolar side chains can 'anchor' to the lipid membrane. (**B**): Trajectory of CDP **1** entering a membrane. The z-position of the CDP is indicated with black dots. The position of the head-group and tail-group region are indicated with dark grey and light grey lines, respectively. The angle between the normal vectors of the peptide and the membrane is shown in blue. The RMSD with respect to the 'closed' conformation of the CDP is shown in orange. (**C**): Snapshot of the CDP while anchoring to the membrane. A hydrogen bond between the CDP backbone and the polar head-group atoms can stabilize the anchoring. (**D**): During the anchoring process, three consecutive hydrogen bonds are formed. The distance between the hydrogen-bond pairs are shown in red. The z-position of the CDP and the membrane are shown as in panel B.

angle between the normal vectors of the peptide and the membrane is around 90° (see blue line). Thus, the peptide is oriented perpendicular to the membrane. At around 42 ns, the peptide slowly penetrates deeper into the membrane. This causes no significant perturbation in the area pear lipid (see Figure 4.5). The peptide first moves through the head-group region in its initial nearly perpendicular orientation. Then, the peptide gradually rotates to be parallel with the membrane plane as it continues the permeation process. Towards the end of the simulation (100 ns), the peptide is located directly at the interface between the polar head groups and apolar tails (see also Figure 4.6). As indicated by the RMSD plotted in orange in Figure 4.4B, in this particular simulation CDP **1** inserts into the membrane in the 'open' conformation. While we also observed anchoring events in the 'closed' conformation for other CDPs (see Figure 4.6), there was a strong imbalance toward anchoring in the 'open' conformation (11 out of 13 unbiased anchoring events, see Table 4.2). This was surprising because the 'closed' and 'open' conformations were equally represented in our starting structures and most of the CDPs have a significant equilibrium population of the 'closed' conformation in water.[108,115] More so, in four simulations we observed CDPs that started in the 'closed' conformation and opened prior to entering the membrane.

Figure 4.4C shows a snapshot of the peptide (here CDP **1**) anchored at the membrane. In this particular simulation, a phenylalanine side chain acts as the 'anchor'. In addition to phenylalanine, we observed that also the leucine and proline side chains can act as membrane anchors. Here, the phenylalanine 'pulls' the peptide towards the membrane through a favorable contact with the apolar tail region. In addition, this anchored structure is stabilized by a hydrogen bond between a POPC head group

TABLE 4.2: Summary of the observed unbiased permeation steps across CDPs **1**-**8**.

CDP	1	2	3	4	5	6	7	8
Anchoring 'open'	2	0	1	2	4	0	2	0
Anchoring 'closed'	0	0	1	0	1	0	0	0
Opening	1	0	0	5	0	1	2	2
Closing	0	0	1	2	1	1	1	7
Half-closing	0	1	0	2	2	0	3	5
Leaflet crossing permanent	0	2	1	0	0	0	1	1
Leaflet crossing transient	0	2	1	0	0	0	0	1

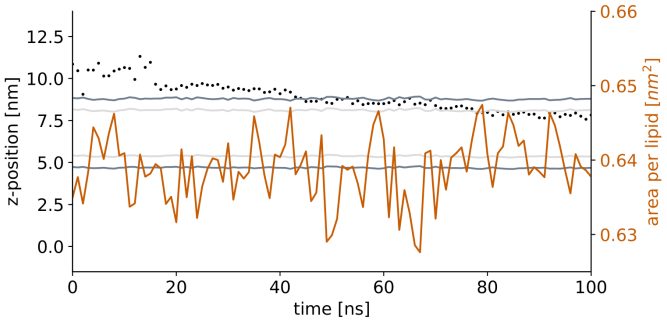


FIGURE 4.5: Change in the area per lipid (brown, right axis) over time. The z-position of the CDP and the membrane are indicated with black and grey dots, respectively. No significant changes in the area per lipid were observed upon the peptide entering the membrane (starting at 23 ns).

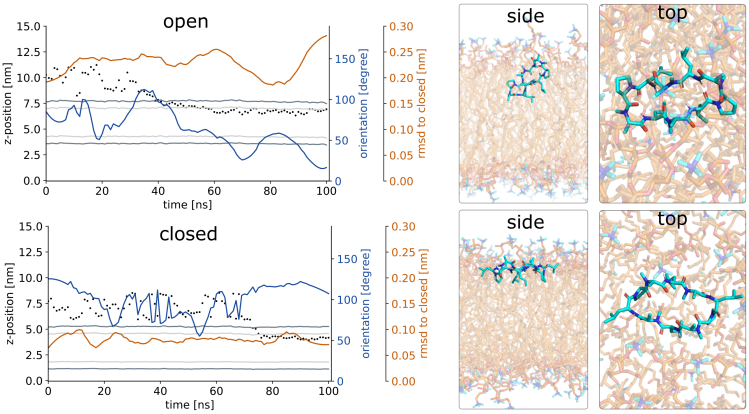


FIGURE 4.6: (Left): Trajectory of CDP **5** entering the membrane in the 'closed' and in the 'open' conformation. The z-position of the CDP and the membrane are indicated with black and grey dots, respectively. The angle between the normal vectors of the peptide and the membrane is shown in blue. The RMSD with respect to the 'closed' conformation of the CDP is shown in orange. (Right): Snapshots of the CDP at the end of the simulations. In both cases, the peptide backbone is nearly parallel to the membrane plane.

and the backbone of leucine at position 1. As indicated in Figure 4.4D, three hydrogen bonds are consecutively formed between the peptide and the membrane in the course of the anchoring and entering process. The first hydrogen bond, as described above, is formed between the phosphate group and leucine 1. As the peptide penetrates deeper into the membrane (at approximately 50 ns), the initial hydrogen bond is replaced by a new hydrogen bond between the same peptide residue and the ester oxygen of the lipid tail. Later (at approximately 75 ns), an additional hydrogen bond between the backbone nitrogen of leucine 6 and the phosphate group of another POPC molecule is formed. Stabilizing hydrogen bonds were observed for all anchoring events in the 'open' conformation. In contrast to the 'open' conformation, the 'closed' conformation is characterized by four intramoleuclar hydrogen bonds. Hence, such stabilizing interactions require a transitional breaking of an intramolecular hydrogen bond in favor of membrane peptide interactions. We observed such transient hydrogen bond breaking and forming in both 'closed' anchoring events. The associated energy barrier might be one reason that membrane contacts in the 'open' conformation are more prevalent in our simulations compared to contacts in the 'closed' conformation.

The unbiased simulations described so far are an adequate MD simulation approach to answer the question how cyclic peptides enter lipid membranes with only limited bias through the simulation setup. However, the high computational cost associated with these simulations prevented us from obtaining sufficient statistics to derive hypotheses for the rational design of permeable cyclic peptides. Nevertheless, the information gathered in the unbiased simulation provided an ideal starting point for enhanced sampling, in our case pulling simulations. The basic idea

in pulling simulations is to introduce a biasing potential along a chosen 'reaction coordinate'. With this technique, we could significantly enhance the occurrence of anchoring events within feasible computational time. Based on the mechanism observed in the unbiased simulations, we applied a weak pulling force on each of the potential anchor residues and pulled towards the membrane center (for details see Section 4.2.2 and Figure 4.3B). Additionally, we performed simulations where we applied the pulling force on the center-of-mass (COM) of the peptide (see Figure 4.2). However, only when the bias is applied to the potential anchor residues we observed that the peptides do anchor and subsequently enter the membrane. This emphasizes the importance of selecting an appropriate reaction coordinate for biased simulations. Hence, in our case insights on the permeation mechanism from the unbiased simulations were vital to be able to bias effectively but cautiously, with minimal distortion of the system. By applying the bias, anchoring events occurred on average every 30 ns, which corresponds to a striking 250 fold speed-up.

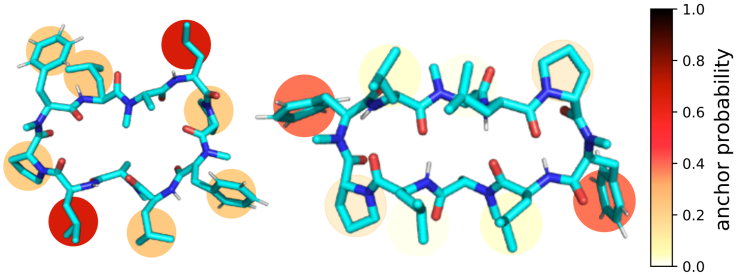


FIGURE 4.7: 'Anchor quality' of the different side chains for the CDPs in the 'open' (left) and 'closed' (right) conformations. The anchor probability was determined as the fraction of successfully anchoring events after pulling the CDPs on that respective side chain towards the membrane. The probabilities are averaged over the eight CDPs.

From the biased simulations, we calculated the fraction of successful anchoring events for each potential anchor residue (Figure 4.7). For this, we pooled the pulling simulations of peptides that share the amino acid of interest. For example, CDP **1**-**7** all have a leucine at position 8. Thus, we combined the results of pulling at leucine 8 from these seven peptides. Note that this approach neglects how non-anchoring residues affect the anchor probability. However, as shown in Figure 4.4C or Figure 4.11A.1, the non-anchoring side chains are distant from the interaction side and therefore unlikely to substantially impact anchoring. Interestingly, the anchoring pattern was very different when the peptides were in an 'open' or in a 'closed' state. In the 'closed' conformation, the phenylalanine residues were the best anchors. All other residues showed only a very low probability. In contrast, the phenylalanine residues were the weakest anchors in 'open' conformations, whereas the leucine residues at position 3 and 8 showed the highest probability. This difference can be explained by the different accessibility of the side chains in the two conformations. In the 'closed' conformation, the leucine residues form a continuous hydrophobic patch, thus a single leucine is less accessible, while the phenylalanine residues are oriented outwards (see also Chapter 3). In 'open' conformations, on the other hand, only leucine residues 1 and 6 form the continuous hydrophobic patch, while the leucines at position 3 and 8 are positioned outwards and accessible (Chapter 3). In general, the overall anchor probability of the 'open' conformations was higher. This is potentially due to the fact that 'open' peptides can form stabilizing hydrogen bonds more easily and is more flexible. Thus, interactions between the apolar residues and the gaps created by thermal fluctuations in the lipids result more often to stable anchoring and subsequent insertion into the membrane.

4.3.2 Effect of Membrane Composition on Head-Group Gaps

The probability of gaps to occur between the lipid head groups is expected to affect the anchoring rate of the CDPs. To investigate how this probability is modulated by the addition of cholesterol molecules (as in biological membranes), we performed MD simulations of a pure POPC membrane and a POPC membrane with 30% cholesterol (without any CDP). The distribution of head-group gaps was computed using Packmem.[190] In both cases, we observed an exponential decrease in head-group gap probability with increasing gap size (Figure 4.8). Gaps up to an area of $0.60\,\mathrm{nm}^2$ and $0.40\,\mathrm{nm}^2$ were observed for pure POPC and POPC+cholesterol, respectively. The minimum gap size needed for a leucine side chain is approximately $0.23\,\mathrm{nm}^2$ (see vertical line). Therefore, both lipid compositions lead to gaps that can accommodate amino acid side chains. However, very bulky anchor residues are likely not beneficial for the entry of the hydrophobic part of the membrane, although the stronger interactions formed by large anchors might compensate for this effect. We also find that head-group gaps are less frequent in the membrane with cholesterol. This is expected as experimental results show that cholesterol decreases the flexibility of a membrane.[191] Additionally, cholesterol is known to decrease membrane permeability.[192] Nevertheless, head-group gaps still occur and allow anchoring of the CDPs.

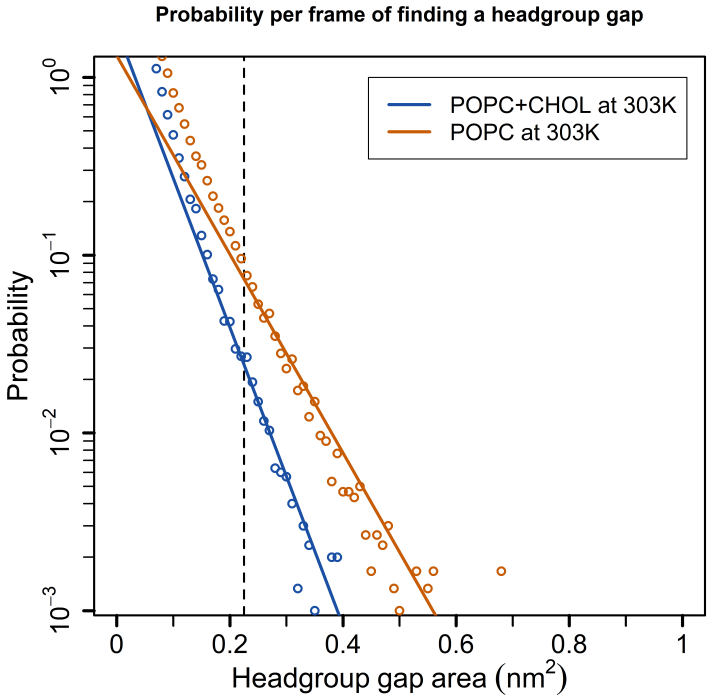


FIGURE 4.8: Probability of finding a head-group gap of a given size in a pure POPC membrane (orange) or a POPC membrane with 30% cholesterol (blue) at 303 K 303 K. The straight lines show a linear fit on the logarithmic probability values, omitting points below a probability of 1/1000 and below an area of 0.15 nm^2. The dashed line is at 0.23 nm^2, as an estimate of the area required by a leucine sidechain (as explained above).

4.3.3 Cyclic Peptides Occupy Two Distinct Orientations at Lipid Membranes

Starting from the anchored peptides, we performed elongated unbiased simulations to analyse how cyclic peptides behave inside lipid membranes. Figure 4.11A displays the entry pathway for a prototypical CDP (here on the example of the 'open' and 'closed' states of CDP **3**). The plots for all CDPs are shown in Figures 4.9 and 4.10. The coordinates of the peptide are projected to the distance of the peptide from the bilayer center (y-axis) and its orientation with respect to the membrane plane (x-axis). The blue heat map represents the density of simulations snapshots, i.e., the darker the color the more simulation points fall into that phase space. Note, however, that due to the rarity of the events we did not reach simulation equilibrium. Thus, the densities do not directly translate to free energies.

All peptides started anchored at the membrane (Figure 4.11A.1). In these starting structures, the COM of the peptides resided outside the membrane (distance to the membrane center >2.1 nm) and their longitudinal backbone axis was oriented nearly perpendicular to the membrane. The peptides stayed in this upright orientation while moving deeper into the membrane (Figure 4.11A.2), as it minimizes both the perturbation of the membrane and facilitates hydrogen bonding with the polar head groups (Figure 4.4). Once the head-group region was passed, the peptides started to rotate in one of two stable orientations. In both orientations, the peptides lay relatively parallel to the membrane plane at the interface between the head groups and the tails (Figure 4.11A 3.a and 3.b). A comparison of the entry pathway of peptides in the 'closed' and 'open' conformation revealed that 'closed' peptides penetrated deeper into the membrane (Figure 4.10) and

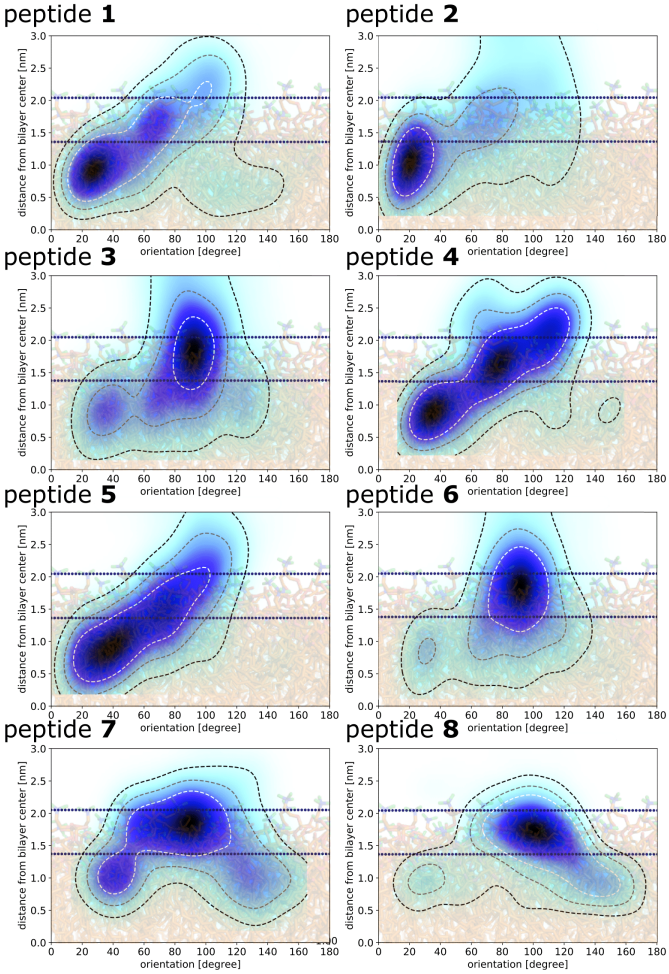


FIGURE 4.9: Positional phase space of CDPs **1**-**8** in the 'open' conformation. The coordinates of the peptides are projected onto its distance from the bilayer center and its orientation in respect to the membrane. The heatmap shows the distribution of the simulation time spent in this phase space. Darker color corresponds to more simulation time.

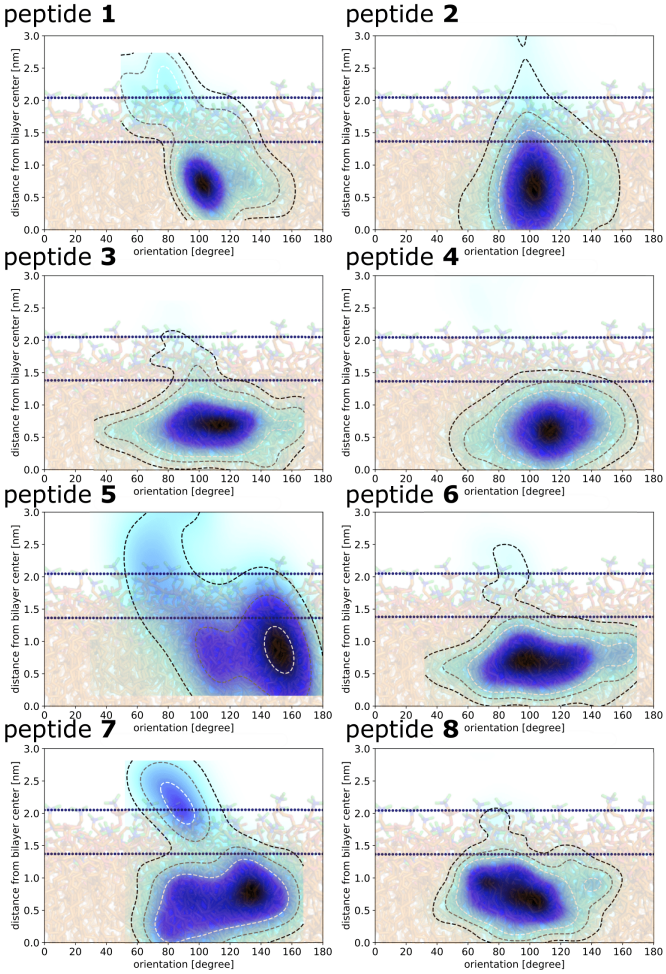


FIGURE 4.10: Positional phase space of CDPs **1-8** in the 'closed' conformation. The coordinates of the peptides are projected onto its distance from the bilayer center and its orientation in respect to the membrane. The heatmap shows the distribution of the simulation time spent in this phase space. Darker color corresponds to more simulation time.

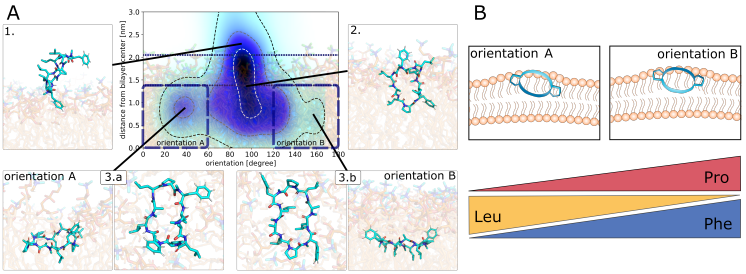


FIGURE 4.11: (**A**): Representation of how cyclic peptides insert into lipid membranes using the example of CDP **3**. The coordinates of the peptide are projected onto its distance from the bilayer center and its orientation with respect to the membrane plane. The heatmap shows the distribution of the simulation time spent in this phase space with darker color corresponding to more simulation time. Regions of interest are highlighted with simulation snapshots. The regions corresponding to orientation A and orientation B are marked with a dotted box. (**B**): Visualization of Table 4.3. The amino-acid composition of the CDPs determines their preference for orientation A or B; proline and phenylalanine residues favor orientation B, leucine residues favor orientation A.

only occupied one orientation, which we will term orientation B. Peptides in the 'open' conformation occupied both orientation A and B with a preference for orientation A. We also observed rotation from one orientation to the other. An example trajectory for such a rotation is shown in Figure 4.12.

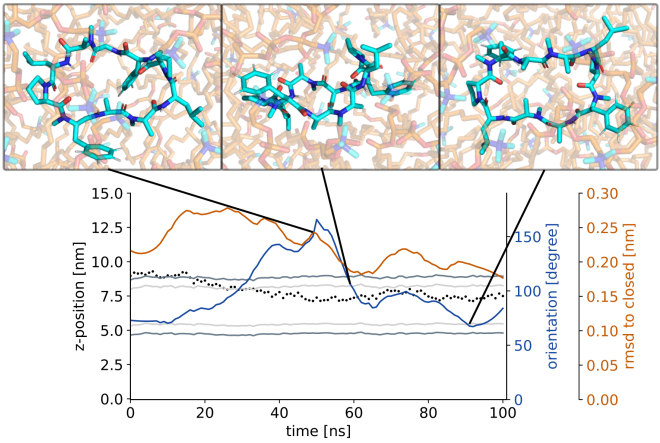


FIGURE 4.12: Rotation of CDP **6** from orientation B to A in the 'open' conformation. The peptide rotates along its major axis defined by the peptide backbone. Simulation snapshots of points of interest are shown. The z-position of the CDP and the membrane are indicated with black and grey dots, respectively. The angle between the normal vectors of the peptide and the membrane is shown in blue. The RMSD with respect to the 'closed' conformation of the CDP is shown in orange.

The difference between orientation A and B is highlighted in Figure 4.13. The peptide is depicted in a top view at the membrane. In comparison to orientation A, orientation B is rotated roughly 180° along the major axis of the peptide backbone. In both orientations, the leucine residues approximately align with the lipid tails. To distinguish the two orientations in the cartoon representation, we chose to depict the proline residues. The pro-

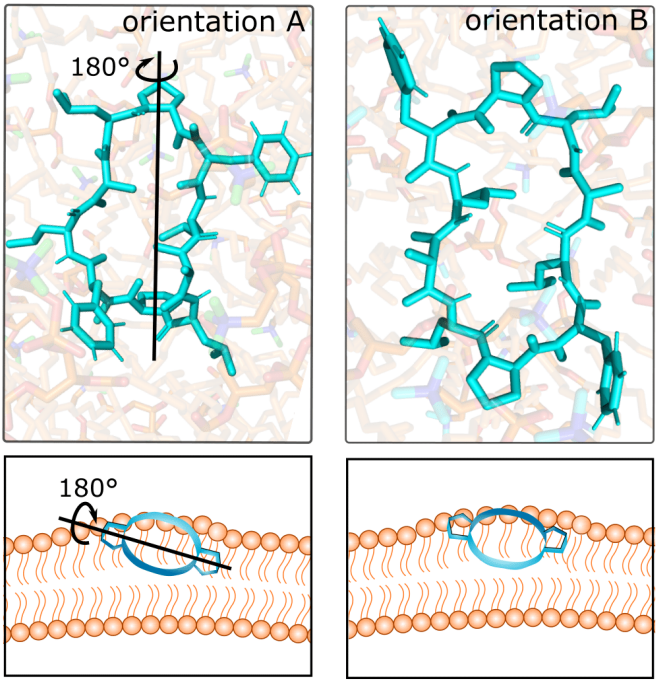


FIGURE 4.13: Representative snapshots of orientation A and B using the example of CDP **3**. The peptide is shown at the polar/apolar interface of the membrane and in a top view. To obtain orientation B from orientation A, the peptide rotates by roughly 180° around its major axis. In addition, the ϕ-angles of leucine residues 3 and 8 show a $\sim$160° shift such that the leucine side chains are approximately aligned with the lipid tails. A cartoon was added for visual guidance.

line ring is peaked towards the membrane middle in orientation A and towards the aqueous phase in orientation B.

At a first glance, it may be surprising that the CDPs orient parallel to the membrane, because this leads to a larger perturbed area in comparison to a perpendicular orientation. Interestingly, in Chapter 3, where we simulated CDPs at a water/chloroform interface, we observed the same two orientations and identified them as energetic minima (see Figure 4.14 for a comparison). We hypothesize that the parallel orientation is more favorable as the hydrophobicity profiles of the CDPs and the membrane match (see also panel 4 in Figure 4.1). These favorable interactions seem to outweigh the penalty of membrane perturbation. In addition, in both orientation A and B, the leucine residues align with the lipid tails, possibly reducing the entropic cost. Furthermore, the observed positioning of the CDPs at the interface of the polar head groups and apolar tails is also in line with the results of previous studies using enhanced sampling simulations that located the free-energy minimum in this region.[114]

4.3.4 Amino-Acid Composition Influences the Preferred Peptide Orientation

In Chapter 3, we found that the amino-acid composition determines the orientation preference of CDPs at a water/chloroform interface. Therefore, we tested whether this finding was similar in the lipid bilayer system. Table 4.3 lists the relative fraction of simulation time spent in either orientation A or B for peptides in their 'open' conformation.

Encouragingly, the numbers in Table 4.3 match the fractions found in the water/chloroform system well (see Figure 4.15 and Chapter 3). With the only exception of CDP **8**, where the frac-

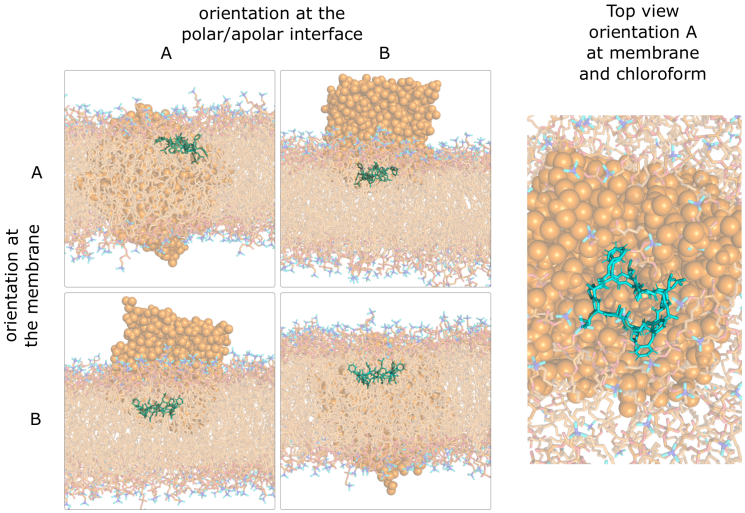


FIGURE 4.14: Comparison of the orientations and conformations found at a polar/apolar water/chloroform interface and at the POPC membrane. Orientations and conformations at the polar/apolar interface were taken from Chapter 3. Chloroform molecules are shown as orange balls. Water molecules are omitted for visual clarity. Orientation A at the interface agrees well with orientation A at the membrane. Similarly, orientation B at the interface agrees with orientation B at the membrane. This indicates that the peptides adopt comparable orientations and conformations in both systems.

TABLE 4.3: Relative fraction [%] of simulation time spent in orientation A and B after equilibration for CDPs **1**-**8** in the 'open' conformation.

CDP	1	2	3	4	5	6	7	8
orientation A	87	95	71	89	96	69	53	33
orientation B	13	5	29	11	4	31	47	67

tion differed by a factor of two, the mean relative difference was only 7.6 %. Three general trends can be observed in Table 4.3 and Chapter 3, which are summarized in Figure 4.11B. (1) The presence of proline in the peptide increases the fraction of orientation B. (2) The presence of phenylalanine also increases the fraction of orientation B. Here, we even observed an titratable effect. Peptides without phenylalanine in their sequence (e.g., CDP **5**) show the smallest fraction of orientation B, followed by peptides with one phenylalanine (e.g., CDP **4**). Peptides with two phenylalanines (e.g., CDP **3**) have the highest fraction of orientation B. (3) The presence of leucine increases the fraction of orientation A in a titratable manner (e.g., CDP **3** versus **6**/**7** versus **8**). Note that the percentages in Table 4.3 display the distribution after 100 ns simulation time. Thus, these numbers reflect the initial orientation distribution after membrane insertion. These clear sequence-specific differences in the orientation preferences of the peptides naturally raise the question of how the orientation influences the permeation process. Indeed, as we show below, the propensity of orientation B appears to be decisive for CDPs to interconvert to the 'closed' conformation in the membrane, which is a necessary prerequisite to cross the lipid bilayer. Therefore, introducing structural modifications to favor orientation B might be a valuable consideration when designing permeable cyclic peptides.

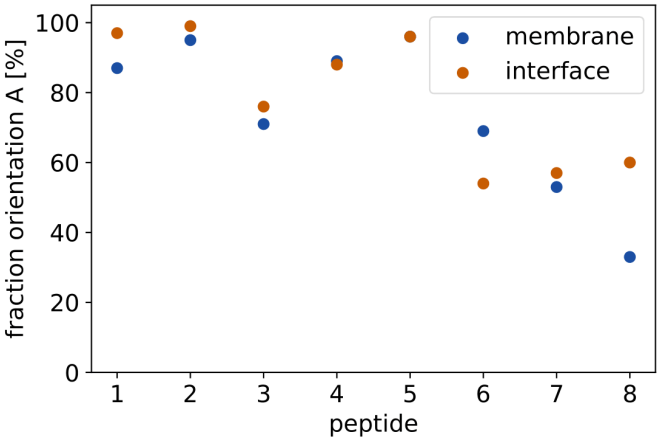


FIGURE 4.15: Fraction of simulation frames in orientation A at a water/chloroform interface (orange) and the POPC membrane (blue). The values for the membranes were taken from Table 2 in the main text. The values for the water/chloroform interface were taken from Chapter 3.

4.3.5 'open' Cyclic Peptides Can Close Inside the Lipid Membrane

In our simulations, we observed that CDPs can insert into lipid bilayers both in the 'closed' and 'open' conformation (see Figure 4.6). Interestingly, the CDPs retained their conformational flexibility inside the bilayer. Thus, also inside the membrane we observed multiple opening and closing events. Figure 4.16A displays a prototypic closing event within the membrane environment. The peptide enters the membrane in the 'open' conformation (RMSD to the 'closed' conformation of $> 0.2\,\text{nm}$). After the initial anchoring phase (till $\sim 18\,\text{ns}$), the peptide adopts orientation B, as indicated by its high orientation angle (till $\sim 30\,\text{ns}$). The displayed inlay at 21 ns also shows that the residues of the peptide are shifted by one position in comparison to their location in the 'closed' structure, which we called 'register shift'. The closing is initiated by the formation of a first hydrogen bond on one side leading to a 'half-closed' conformation ($\sim 40\,\text{ns}$). In this 'half-closed' conformation, the residues relocate their relative position and the register shift is resolved. In order to fully close, the formed hydrogen bond is broken again, thus leading to an 'open' conformation without a register shift. After further backbone torsional changes, the peptide is finally in the 'closed' conformation ($\sim 60\,\text{ns}$).

In total, we observed thirteen closing events and thirteen half-closing events across the different CDPs (see Table 4.2). Importantly, all closing and half-closing events started in orientation B. Although the peptides spent on average eight times more simulation time in orientation A, no closing event originating from orientation A was observed. Figure 4.16B illustrates this finding. The top panel shows the projected coordinates of all frames in the

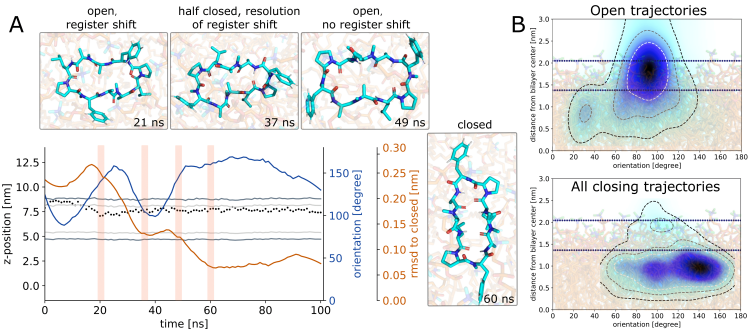


FIGURE 4.16: (**A**): Representative closing simulation using the example of CDP **6**. The peptide starts in the 'open' conformation and closes inside the membrane via a half-closed structure. The closing of the peptide is traced by its RMSD with respect to the 'closed' reference conformation (orange line). Inlays show selected simulation snapshots. Shaded areas correspond to the time point of the inlays. The dotted line indicates the z-position of the peptide. The membrane position is shown as a reference (grey). The peptide stays in orientation B for the whole simulation (blue line). (**B**): Heatmap comparison of the orientation/position of all simulation frames in the 'open' conformation (here for CDP **6**) versus the orientation/position of the twelve closing trajectories. Whereas 'open' peptides prefer orientation A, all 'open' peptides that close during our simulations originate from orientation B.

'open' conformation of CDP **6**. This analysis already reveals the preference for orientation A when the CDP is in an 'open' conformation. The bottom panel shows the projected coordinates of all frames in the 'open' conformations that subsequently close. They are exclusively found in orientation B. Furthermore, as described in Section 4.3.3 above, peptides that already enter the membrane in the 'closed' conformation exclusively occupy orientation B (see also Figure 4.10). In summary, these observations suggest a link between the orientation of the peptide in the membrane and its 'internal' conformational preferences. While orientation A appears to prohibit rearrangements to the 'closed' conformation, orientation B shifts the ensemble population towards it and closing events become favorable. Thus, we argue that the phase-space overlap between the 'open' peptides in orientation B and the 'closed' peptides may facilitate the closing. These conclusions again match the central findings from Chapter 3 at the water/chloroform interface. 'open' peptides that were in orientation B occupied a higher energy state and showed faster closing kinetics than peptides in orientation A. Additionally, Markov state models on that simulation data revealed that some CDPs with a low 'closed' population in water had a significantly higher 'closed' population at the interface. We hence hypothesize that the membrane has a similar 'catalytic' effect on the conformational behavior of the CDPs and can facilitate a population shift towards the 'closed' conformation. To test this hypothesis, we constructed a MSM for CDP **4** as it was the best sampled peptide. The implied timescale plot of the MSM and a comparison between the MSM for CDP **4** at the water/chloroform interface and the membrane are shown in Figure 4.17 and 4.18. Both models identified a 'closed' and a 'half-closed' metastable state that only occupied orientation B. The 'open' metastable states were split based on

their orientation. Only orientation B showed transitions into the 'closed' and 'half-closed' conformation. More extensive studies using enhanced sampling[193] and dynamic reweighting (Chapter 5) will be required to test the 'catalytic' hypothesis further. However, from a methodological point of view, our combined works so far indicate that highly valuable mechanistic insights can be collected by approximating the membrane interface with a water/chloroform system. This is particularly interesting from a technical perspective as the latter system is computationally significantly less demanding.

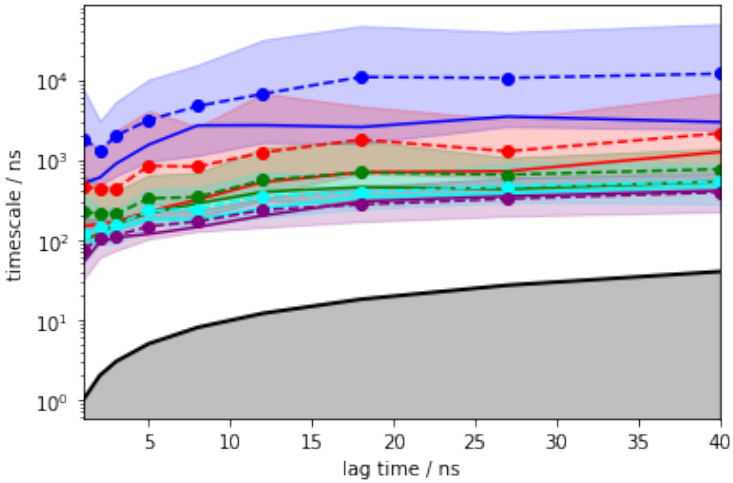


FIGURE 4.17: Implied timescales of the MSM of CDP **4**

Taken together our observations so far imply that cyclic peptides can cross membranes via two pathways: (1) In the 'pre-folding' pathway, peptides have a significant population of the 'closed' conformation already in the aqueous solution. In this conformation, they are able to insert into the membrane and cross it. (2) In addition, our simulations suggest that cyclic peptides

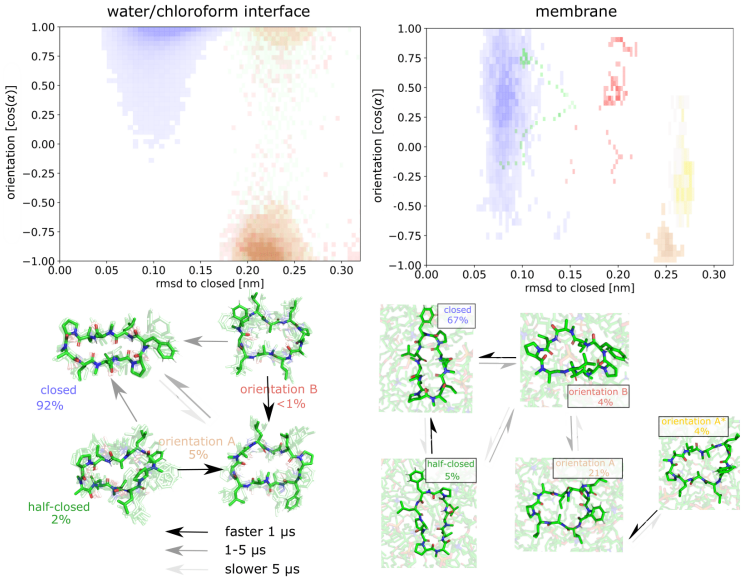


FIGURE 4.18: MSM of CDP **4** in the water/chloroform system (left) and at the membrane (right). Top panel: Metastable state assignment mapped to the orientation feature and the RMSD with respect to the 'closed' state. Bottom panel: Representative members of the metastable states, their equilibrium populations and mean first passage times between them.

can enter the membrane in an 'open' conformation and close inside the membrane. Thus, we showed that also peptides without pre-folding can possibly achieve passive permeability if closing is sufficiently favorable inside the membrane. This opens up a new realm of design considerations that focus on increasing the 'closed' population at the lipid interface in addition to pre-folding in water.

4.3.6 Phenylalanine Residues Can Act as a 'Lock' That Prevents Closing

We observed in our simulations that CDPs only closed when in orientation B. Although all closing events originated from orientation B, not all peptides in orientation B closed. Therefore, we investigated the difference between the 'closing' and 'non-closing' peptides in orientation B. We found that the residue position at the β-turn of the CDPs (position 5 and 10, in this CDP series either phenylalanine or alanine) was decisive for closing. The left panel of Figure 4.19 illustrates this difference. In the top snapshot, the two phenylalanine side chains at the β-turn both point in the same direction towards the aqueous solution. In this position, the peptide backbone is able to interconvert to the 'closed' conformation. In contrast, if the two residues point in different directions, this creates a 'lock' that prevents the peptide from closing (bottom snapshot). The right panel of Figure 4.19 shows a simulation where the peptide enters the membrane in an 'open' conformation and adopts orientation B. Initially, the two phenylalanine residues point in different directions ($\sim 14\,\mathrm{ns}$). In this specific example, phenylalanine at position 5 points towards the membrane center. Despite the bulkiness of the phenylalanine residue, it rotates away from the 'locked' position until both

phenylalanine residues face toward the aqueous phase ($\sim$ 29 ns). Shortly after this shift (and thus the release of the 'lock'), we observed the closing of the peptide. We validated the 'lock' hypothesis using biased simulations (see Methods Section). Starting from a 'locked' position, we applied a small force to the dihedral angle of the phenylalanine backbone torsion that pulled it to the unlocked position. Indeed, as shown in Figure 4.20, this led to releases of the 'lock' and subsequent closing events.

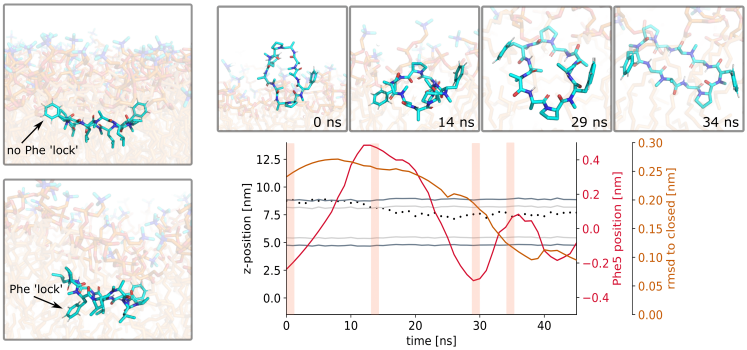


FIGURE 4.19: Phenylalanine can act as a 'lock' that prevents closing in the membrane. (**Left**): Phenylalanine can adopt two distinct positions in orientation B. In the unlocked position, both phenylalanine residues point towards the aqueous phase. In the 'locked' position, at least one phenylalanine residue is rotated and points towards the membrane center. All closing events originate from the unlocked position. (**Right**): Simulation of CDP **8** that shows an 'unlocking' and a closing event. The closing of the peptide is traced by the RMSD with respect to the closed reference (orange line). The red line indicates the relative position of phenylalanine at position 5 with respect to the ring plane of the peptide. Inlays show selected simulation snapshots. Shaded areas correspond to the time point of the inlays. The dotted line indicates the z-position of the peptide. The membrane position is shown as a reference (grey). The peptide first adopts orientation B in the 'locked' position after entering the membrane. After a rotation of phenylalanine residue 5 to the unlocked position, the peptide starts closing.

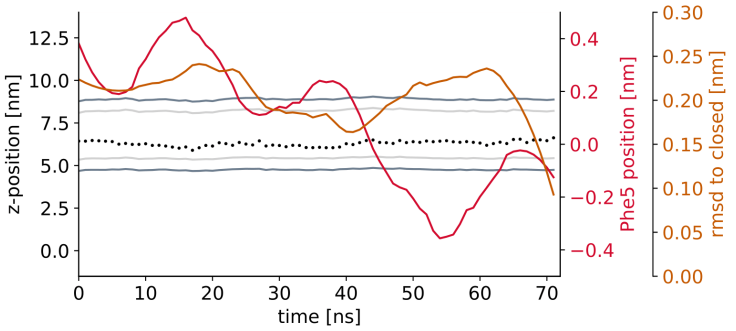


FIGURE 4.20: Phenylalanine can act as a 'lock' that prevents interconversion from the 'open' to the 'closed' conformation. A bias was applied on the ψ backbone torsional angle of phenylalanine in the position 5 to pull it from the 'locked' to the unlocked position. After unlocking, the peptide started closing. The closing of the peptide is traced by the RMSD with respect to the 'closed' reference conformation (orange line). The red line indicates the relative position of phenylalanine residue 5 with respect the the ring plane of the peptide. The dotted line indicates the z-position of the peptide. The membrane position is shown as a reference (grey).

These findings highlight the contextual role of bulky residues like phenylalanine at position 5 and 10. On the one hand, they are the most efficient anchors in the 'closed' conformation. On the other hand, their bulky nature can hinder conformational closure if they are in the 'locked' position. Consequently, this observation emphasizes again that for an optimal design strategy, it is crucial to know the prevalent permeability pathway of a cyclic peptide. Pre-folding peptides enter the membrane in the 'closed' conformation and thus rely on anchors at the β-turns. In addition, they do not need to close inside the membrane as they are already in the 'closed' conformation. In contrast, peptides, which are mostly in 'open' conformations in water, prefer anchors at different positions, but have to close inside the membrane. For such peptides, it might be beneficial to have less bulky residues at the β-turns.

4.3.7 Crossing From the Upper to the Lower Leaflet Requires Anchoring and Flipping

After entering the lipid membrane, the cyclic peptides have to cross from the upper leaflet to the lower leaflet in order to fully permeate through the membrane. Previous research has shown that this is associated with an energy barrier that can be higher than the barrier for entering the membrane.[114] Indeed, also in our simulations, leaflet crossing was the rarest of the three membrane permeation steps we observed (anchoring/entering, closing, and leaflet crossing). However, we were able to observe five permanent and three transient unbiased leaflet crossing events. Generally, only peptides in the 'closed' conformation penetrated deep enough into the bilayer to loose all water interactions (see Figure

4.21) and subsequently cross to the lower leaflet. This is in line with our observation that only 'closed' peptides fully enter the apolar phase in a polar/apolar interface system (Chapter 3).

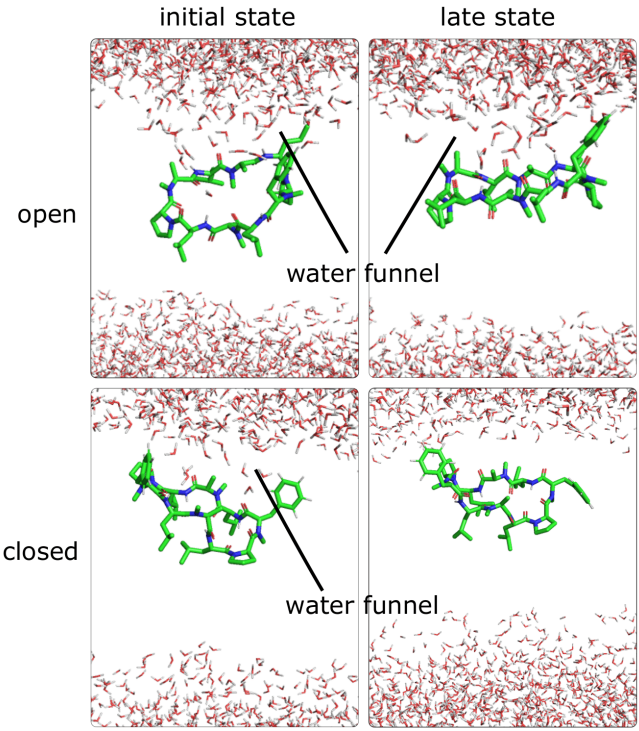


FIGURE 4.21: Water interactions of the peptide inside in the membrane in the 'open' and 'closed' orientation. The lipid molecules are not shown for visual clarity. In the 'open' conformation, the peptide stays in contact with the aqueous phase through a water funnel. In the 'closed' conformation, the peptide loses all water contacts and fully emerges into the membrane.

Figure 4.22A shows an unbiased leaflet crossing event for CDP **3**. The 'closed' peptide first enters the membrane in the typical upright position (0 ns to 5 ns). After penetrating deeper into the

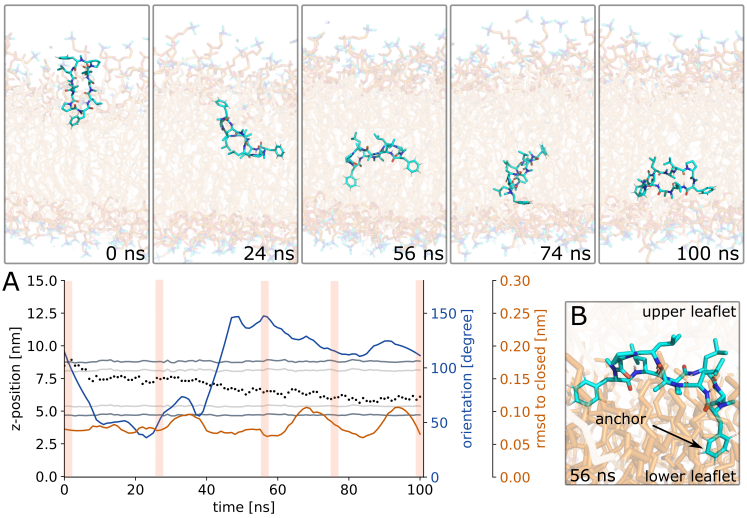


FIGURE 4.22: (**A**): CDP **3** crossing from the upper to the lower leaflet. Inlays show representative simulation snapshots. Shaded areas correspond to the time point of the inlays. Upon passing the membrane center, the peptide undergoes a flip along its major axis (blue line). (**B**): Zoom-in on the peptide anchoring in the lower leaflet. Lipid tails from the upper and lower leaflet are colored differently to help distinguishing the two leaflets.

lipid membrane, it adopts orientation B, which is the stable orientation for 'closed' peptides (24 ns). In this orientation, the proline residues point towards the upper aqueous phase and the leucine residues align with the lipid tails. Most of our simulations remain in this state for the entire simulation time. However, we also observed some rare leaflet crossing events that were associated with a flip of the peptide as shown in Figure 4.22 at 56 ns. During the movement into the lower leaflet, the peptide rotates roughly 180 ° along its major axis. Figure 4.22 shows a fast flipping event where the peptide performed the full flip within a few ns, while other simulations also showed a slower and more gradual flipping.

A closer inspection of the flipping process revealed two underlying principles. First, the flip was triggered by an anchoring event. Anchoring of one of the peptide residues in the lipid tails of the lower leaflet preceded the flip and leaflet crossing. Figure 4.22B shows a zoom-in on this anchoring. Here, the lipid tails of the upper and lower leaflet are colored differently to aid the visual inspection. Using the knowledge of the initial anchoring events, we re-seeded simulations from anchored positions and obtained 18 additional unbiased leaflet crossing events. Second, our orientation analysis showed that 'closed' peptides almost exclusively occupied orientation B (see Figure 4.10). In orientation B, the leucine residues point towards the membrane center and the phenylalanine and proline residues point towards the aqueous phase. In order to adopt this orientation in both leaflets, the peptide has to rotate 180° while passing through the membrane center. Thus, flipping is necessary to adopt the more favorable orientation B for 'closed' peptides at the lower leaflet. Indeed, in the three transient leaflet crossing events we observed no flipping and thus the peptides diffused back to the original leaflet after a few ns. Flipping along its major axis requires larger movements of

the peptide in the rather viscous and sterically hindered lipid-tail environment. We reason that the anchoring and the orientation change go hand in hand to overcome the leaflet crossing barrier.

To assess the anchor quality of the different amino acid residues for leaflet crossing, we again performed pulling simulations as described above. Using these simulations, we were able to increase the rate of leaflet crossing events by a factor of 140. The resulting probabilities are depicted in Figure 4.23. The results resemble the anchor probabilities for entering the membrane in the closed state (Figure 4.7) with the difference that the total values are higher for the leaflet crossing event. The phenylalanine residue is again the best anchor. As the local environment in the upper and lower leaflet are identical, there is no environmental change associated with the anchoring. Thus, the different anchor probabilities can be attributed to differences in accessibility, with the phenylalanine residue being the most exposed in the 'closed' conformation.

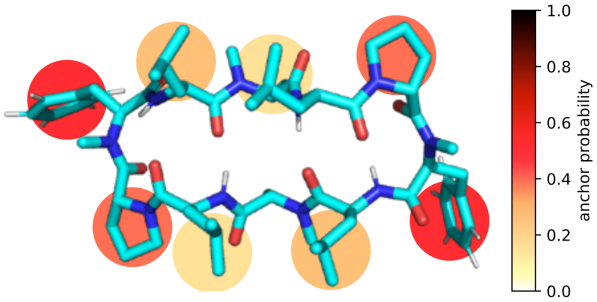


FIGURE 4.23: Ability of the different side chains of the CDPs for anchoring in the lower membrane leaflet when the peptide is inserted in the upper leaflet. Membrane crossings were observed only for peptides in the 'closed' conformation . The anchor probability was determined as the fraction of successfully anchoring events after pulling the CDPs on that respective side chain towards the lower leaflet.

4.3.8 Four Steps of Membrane Permeation

Figure 4.1 summarizes the four steps of passive membrane permeation that we identified for the series of CDPs. We found that the first step involves the anchoring of an exposed residue to a transient gap between the lipid head groups (Figure 4.4). This process is stabilized by hydrogen bonds between the backbone atoms of the peptide and the head groups. Interestingly, using machine learning techniques on a large compound library, Rzepiela *et al.*[161] found that an asymmetric accumulation of hydrophobicity on one side of a compound is predictive for highly permeable macrocycles. The authors presume that hydrophobic regions might enter the membrane first and thus catalyze the entry of the remaining molecule. This hypothesis is in line with the mechanistic insights gained from our simulations, where the apolar residues act as anchors and enter the membrane first. The CDPs could insert into the membrane in both 'open' and 'closed' conformations, with different residues as main anchors. Interestingly, 'open' conformations showed a higher total probability of entering the membrane due to an increased ability to form stabilizing hydrogen bonds with the lipid head groups.

The second step is insertion and orientation. In the bilayer, the peptides locate themselves at the interface between the apolar tails and the polar head groups. While peptides in the 'open' conformation can adopt two different orientations A and B that differ by a 180° rotation (Figure 4.13), 'closed' peptides occur only in orientation B. In line with our previous work, we found that the amino-acid composition modulates the fraction between orientation A and B for the 'open' peptides (Figure 4.11 and Table 4.3). If the peptide entered the membrane in an 'open' conformation, it has to interconvert to the 'closed' conformation to be

able to cross the membrane core. Importantly, only 'open' peptides in orientation B interconvert to the 'closed' conformation (Figure 4.16). Given the importance of orientation B for closing, optimizing CDPs to occupy orientation B might be of interest in designing peptides. In our simulations, both the presence of proline and phenylalanine as well as the absence of leucine enhanced the fraction of orientation B. However, both the phenylalanine and leucine residues have an ambivalent role. As shown in Figure 4.19, phenylalanine residues can also act as a 'lock' and prevent closing. Leucines favor orientation A but are also important anchors for peptides in 'open' conformations.

The fourth step is the crossing from the upper to the lower leaflet. We found that this process again involves anchoring of an exposed residue – this time to the lipid tails of the lower leaflet. In addition, the peptide flips when passing the membrane center (Figure 4.22). Thus, in each leaflet the CDP adopts the favorable orientation B within the membrane. Given the major lipid rearrangements necessary to allow the flip of molecules as large as the CDPs in this chapter, bulky amino acids are expected to reduce the flipping rate.

Taking into account all studied permeation steps and the various – and sometimes conflicting – roles amino acids play in those steps, it is not surprising that the effects of the amino-acid composition are highly contextual. Although the dataset is small, we can make some observations.

Leucines. Leucine residues are important anchors for 'open' peptides, but favor orientation A inside the membrane and are relatively bulky. Starting from CDP **3** and removing half of the leucine residues leads to CDPs **6** and **7**, which have an eight to fifteen fold increased permeability over CDP **3** (Table 4.1). The remaining two leucine residues appear to be sufficient to ensure

reasonable anchoring rates while the probability of orientation B is significantly increased (Table 4.3). However, if all leucine residues are removed as in CDP **8**, the permeability is decreased by a factor of 12 compared to CDP **3**, even though the fraction of orientation B is even higher (Table 4.3).

Phenylalanines. Phenylalanine residues showed a special significance throughout all four steps of permeation. For peptides in the 'closed' conformation, it showed the highest anchor probability for entering the membrane as well as for crossing to the lower leaflet. For peptides inserting in the 'open' conformation, the bulkyness of phenylalanine hinders the interconversion to the 'closed' permeable conformation ('lock'). In water, CDP **1** is found almost exclusively in 'open' conformations, whereas CDP **3** has a significant population of the 'closed' conformation ($< 1\%$ versus 45% 'closed' fraction, respectively).[89] Thus, the replacement of all phenylalanine residues affects these two peptides differently. The 'open' CDP **1** does not rely on phenylalanine to anchor to the membrane but benefits from the removal of the 'locks'. Thus, we observe a large 80 fold increase in permeability upon replacement of all phenylalanine residues resulting in CDP **2** (Table 4.1). In contrast, replacing all phenylalanine residues of CDP **3** leads to CDP **5**, which is only associated with a moderate five fold permeability increase (Table 4.1).

Based on these observations, we can draft schematic free-energy surfaces for the permeation process of the impermeable CDPs **1** and **8**, and how it may change upon modifications (Figure 4.24). The main energy barriers are associated with the interconvertion between the 'closed' and 'open' states in water and in the membrane, passing through the lipid head-group region, and leaflet crossing. As shown in our previous studies (Ref. 89

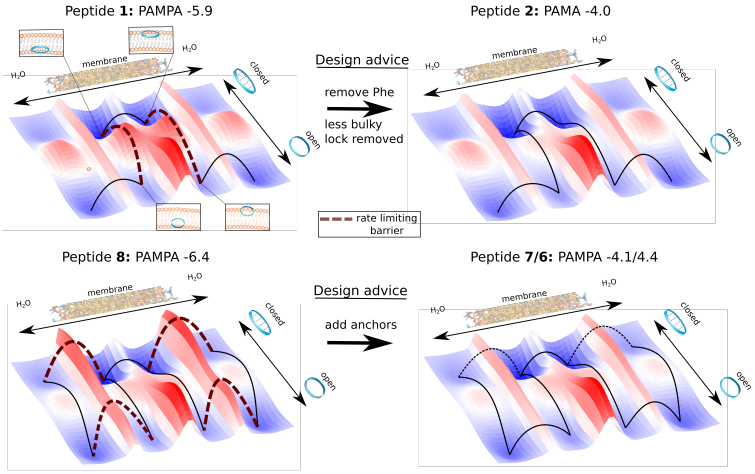


FIGURE 4.24: Schematic free-energy surfaces and the corresponding rate-limiting barriers for the impermeable CDPs **1** and **8** (conformational closure for CDP **1** and anchoring for CDP **8**). The free energies are projected to the conformation of the peptide and its position with respect to the membrane. Design advices for lowering barriers and thus improving the passive permeability are indicated.

and Chapter 3) and based on the fact that CDP **1** was the only peptide where not a single closing or half-closing was observed (see Table 4.2), CDP **1** suffers from both an unfavorable ratio between 'open'/'closed' states and high interconversion barriers. To improve interconversion within the membrane, our permeation model suggests to remove the bulky phenylalanine residue(s) that can lock the peptide in the 'open' conformation inside the membrane, leading to CDP **2** which has a significant higher permeability. For CDP **8**, in contrast, we observed a high interconversion dynamic with multiple closing and opening events (Table 4.2). However, CDP **8** contains no leucine residues. In our permeation model, this should decrease its anchoring potential, especially in

the 'open' conformation. Indeed, in this chapter not a single anchoring event was observed for CDP **8** (see Table 4.2). This highlights that different energy barriers can become rate limiting for different cyclic peptides. In this case, introducing residues with good anchoring ability is adviced by the model, leading to CDPs **6** and **7** with dramatically improved permeabilities.

4.4 CONCLUSION

In this study, we investigated the pathway and main steps for the passive membrane diffusion of conformationally flexible cyclic peptides. Based on extensive simulations, we identified four steps of the membrane permeation process of cyclic peptides: (1) anchoring with residues in transient gaps between lipid head groups, (2) insertion in the membrane and orienting parallel to the membrane plane in orientation B, (3) if the peptide enters in an 'open' conformation, interconversion to the 'closed' permeable conformation, and (4) leaflet crossing involving anchoring and rotation.

For the first step, the pulling simulations revealed that the main anchoring residues differ for 'open' and 'closed' conformations, due to the different accessibility of these residues in the two conformations. Given that the anchoring probability of amino acids is conformation dependent, characterizing the conformational behavior of cyclic peptides is therefore crucial for rational design. Knowing the predominant conformation of the peptide before entering the membrane and the exposed residues may help to optimize the amino-acid composition for membrane permeable CDPs. Considering the anchor potential of residues for permeability is a new design concept that might be transferable

to other compounds as well. A recent study by Morstein *et al.*[194] identified medium-chain lipid conjugation as a general modulator of cell membrane permeability. Investigations of whether these medium-length lipid chains act as 'anchors' will be part of future research.

In the second step, the peptides insert in the membrane and orient themselves parallel to the membrane plane. If the insertion occurs in an 'open' conformation, two orientations A and B are possible. If the peptides insert already in the 'closed' conformation (pre-folding), only orientation B was observed. This design consideration should be applicable to other macrocycles with asymmetrically oriented large and continuous hydrophobic surface patches. Only orientation B is favorable for subsequent interconversion to the 'closed' permeable conformation (step 3). We expect this to apply to chameleonic cyclic peptides in general.

For full permeation, the peptides have to cross the membrane center to reach the lower leaflet (step 4). An anchoring mechanism similar to the initial step was observed in connection with a rotation of the peptide to reach orientation B in the lower leaflet. Again, we reason that this step should be applicable to other macrocycles with asymmetrically oriented large and continuous hydrophobic surface patches.

The simulations show that the effect of amino acids (e.g. leucines, phenylalanines, prolines) can be highly contextual, i.e. it depends on the location of the amino acid in the peptide and on the other residues. The observations and hypotheses presented in this study are based on a small data set and limited influence parameters, but nevertheless provide important insights and a highly detailed mechanistic model of the membrane permeation process of flexible cyclic peptides. While biasing was applied in the initial peptide-membrane association with a pulling force, all

mechanistic conclusions were drawn from unbiased MD simulations. Previous simulation-based studies either have focused on significantly smaller and less flexible molecules and/or fully relied on biased simulations. Most importantly, our findings emphasize that solid understanding of the preferred conformation(s) of a peptide in solution is decisive for the success of a lead optimization campaign targeting permeability. With the here provided atomistic and dynamic insights into the permeation pathway of cyclic peptides, we aim to inspire and stimulate new design principles and predictive modeling approaches for bioavailable macrocyclic drugs.

5

Connecting Dynamic Reweighting Algorithms: Derivation of the Dynamic Reweighting Family Tree

*

"Wer ein gutes Gedicht schrieben kann, der kann auch programmieren" and *"Wer die Poesie der Algebra beherrscht, der kann auch programmieren"*

Michael Resch and
an anonymous physicist
This quote and chapter are dedicated to R. Gregor Weiß and his mastery of the poesy of physical equations

* This Chapter is reproduced in part from S. M. Linker, R. G. Weiß, S. Riniker, *J. Chem. Phys.*, 153, (2020) 234106, with the permission of AIP Publishing.

Thermally driven processes of molecular systems include transitions of energy barriers on the microsecond timescales and higher. Sufficient sampling of such processes with molecular dynamics (MD) simulations is challenging and often requires to accelerate slow transitions using external biasing potentials. Different dynamic reweighting algorithms have been proposed in the past few years to recover the unbiased kinetics from biased systems. However, it remains an open question if and how these dynamic reweighting approaches are connected. In this work, we establish the link between the two main reweighting types, i.e. path-based and energy-based reweighting. We derive a path-based correction factor for the energy-based dynamic histogram analysis method (DHAM), thus connecting the previously separate reweighting types. We show that the correction factor can be used to combine the advantages of path-based and energy-based reweighting algorithms: it is integrator independent, more robust, and at the same time able to reweight time-dependent biases. We can furthermore demonstrate the relationship between two independently derived path-based reweighting algorithms. Our theoretical findings are verified on a one-dimensional four-well system. By connecting different dynamic reweighting algorithms, this work helps clarify the strengths and limitations of the different methods and enables a more robust usage of the combined types.

5.1 INTRODUCTION

Molecular dynamics (MD) simulations play a central role in understanding molecular motion at atomic resolution across disciplines (see e.g. Refs. 19,195–198). Since the advent of MD simulations, the system size and complexity has considerably progressed. Nevertheless, MD is still limited at best to millisecond timescales.[199] However, free-energy landscapes of molecular systems may comprise barriers of several hundred thermal energies ($k_{\mathrm{B}}T$), which results in comparably rare transitions. Thus, sufficient sampling of complex systems is challenging or even impossible with unbiased MD setups.

To overcome these limitations, a wide variety of enhanced sampling techniques have been developed over the past decades. Methods with time-independent biasing such as umbrella sampling,[200,201] replica exchange,[202,203] and accelerated MD[204,205] as well as methods with time-dependent biasing such as local elevation,[49] conformational flooding,[206] and metadynamics[50] are among the most widely used enhanced sampling algorithms. These techniques have in common that a biasing potential or increased temperature is applied to the system to systematically enhance the occurrence of the slowest processes. Hence, by definition the enhanced sampling algorithms distort the trajectories and their information about the unbiased system's thermodynamics and kinetics.

The ensemble averages of the unbiased system can be recovered from biased trajectories with phase-space reweighting methods.[207–211] Note that these can recover thermodynamic properties (such as energetic differences), but not the unbiased system kinetics.

Only recently, dynamic reweighting methods have been proposed that have the ability to extract the original, unbiased kinetics from biased simulations.[6–8,45,46,212,213] A summary of the different dynamic reweighting methods is provided by Kieninger *et al.*[214] In general, these dynamic reweighting methods are applied to Markov state models (MSMs). MSMs are a widely used technique to obtain the slowest dynamic processes from simulated trajectories.[6,23,24,110,111,215] The application of MSMs is focused on, but not limited to, meta-stable state detection, equilibrium distribution calculation, kinetic rate recovery, and flux analysis (see e.g. Refs. 89,216–218 as recent examples). Dynamic reweighting methods reweight individual transitions. For this, individual transitions between states are tracked, reweighted, and subsequently collected in the so-called count matrix. This procedure is discussed in detail in Section 5.2.

One can distinguish two main classes of dynamic reweighting approaches: path-based and energy-based. Both types of approaches reweight the MSM propagator P_{ij}, i.e. the probability of being in state j after a timestep τ under the condition that we started in state i. Path-based reweighting methods are currently only applicable to stochastic dynamics (SD) and require the knowledge of both the bias energy and the random number for integration at each step. The relatively new path-based reweighting class has currently two members, Weber-Pande reweighting[45] and Girsanov reweighting.[8,213] Both methods yield good results for the kinetics, however, their implementation is usually nontrivial and depends on the chosen integration scheme. In contrast, the energy-based reweighting methods are independent of the chosen integration algorithm and only require the knowledge of the bias energy for each phase-space state, which is easily accessible from the MD trajectories. However, depending on the

system, these methods are less accurate because they do not explicitly consider the different paths that could lead from state i to state j. The energy-based reweighting class includes multiple members, see Refs. 6,7,46,212. In this chapter, we will focus on one of the earliest algorithms, the dynamic histogram analysis method (DHAM),[46] as a representative of this class. Other energy-based reweighting methods are either extensions of this method, i.e. DHAMed,[7] or have a similar maximum likelihood approach.[6] In addition, DHAM can reweight kinetics based on only enhanced sampling trajectories without the need for additional unbiased simulations. Therefore, DHAM was chosen as a representative.

The different dynamic reweighting algorithms were developed by different groups for different simulation setups and with different physical assumptions. Therefore, it remains an open question if and how the different approaches are connected, and if a "family tree" of dynamic reweighting can be established. In this chapter, we derive the link between the two main reweighting classes (path-based and energy-based). First, we demonstrate that the two path-based reweighting algorithms Weber-Pande[45] and Girsanov[8,213] can be deduced from the same "generalized" path reweighting idea. Therefore, they share the path reweighting factor but differ significantly in the phase-space reweighting. Next, we reformulate this generalized path reweighting idea to show that the energy-based reweighting method DHAM is a special case of the path-based reweighting methods. They are connected by a path correction factor, which can be used to combine the strengths of both reweighting types. Importantly, our new formulation does no longer require the knowledge of the integrator's random number. Hence, this approach can be used with arbitrary numerical schemes such that the scope of this type of reweight-

ing methods is extended to MD simulations with deterministic integrator schemes. Additionally, our formulation leads to multiple implementation advantages as well as increased parameter stability. The impact of the derivation assumptions as well as the previous and new formulations are illustrated using (stochastic) simulations of a Brownian particle in a one-dimensional four-well potential. This work clarifies strengths and limitations of the different reweighting approaches and will be crucial for the application of dynamic reweighting to more complex systems.

5.2 THEORY

5.2.1 Markov State Models

MSMs are a technique to analyze dynamic systems in a Master equation representation.[219,220] In the case of MD simulations, the conformational dynamics are discretized into n states for which a transition probability matrix is constructed. Particularly, the total count C_{ij} of transitions from state i to j can be expressed as,

$$C_{ij}(\tau) = \sum_{t=1}^{N_\tau} \chi_i(x_t)\chi_j(x_{t+\tau}), \tag{5.1}$$

where N_τ is the number of paths of length τ, and $\chi_i(x_t)$ are the indicator functions with $\chi_i(x_t) = 1$ if the configuration x_t at time t is in state i and $\chi_i(x_t) = 0$ otherwise. The individual transitions will be termed,

$$b_{ij}(t, \tau) = \chi_i(x_t)\chi_j(x_{t+\tau}), \tag{5.2}$$

The total counts C_{ij} are stored in the count matrix C, which yields the MSM, i.e., the $n \times n$ transition probability matrix P, upon row-normalization,

$$P_{ij} = \frac{C_{ij}}{\sum_{j=1}^{n} C_{ij}} \,. \tag{5.3}$$

The eigenvalues λ_i and eigenvectors ρ_i of the transition probability matrix P provide the dynamic information of the system. The first eigenvector ρ_1 represents the stationary probability distribution of the n states and its eigenvalue is $\lambda_i = 1$. The remaining eigenvectors $\rho_2, \ldots, \rho_n$ represent principal dynamic modes, i.e., probability fluxes, in the network of the n states. The corresponding eigenvalues $\lambda_2, \ldots, \lambda_n$ are smaller than one and quantify the associated rates of the probability fluxes. Hence, the relaxation timescales of the dynamic modes

$$t_i = -\frac{\tau}{\ln(\lambda_i)}, \tag{5.4}$$

are determined by the eigenvalues with $i > 1$.

5.2.2 DHAM Reweighting

The dynamic histogram analysis method (DHAM)[46] is based on the Master equation ansatz. More specific, it is based on the spatial discretization of the Smoluchowski diffusion equation derived by Bicout and Szabo.[221] This diffusion equation states that for short time steps and diffusive dynamics the unbiased transition probabilities P_{ij} are related with the biased ones $\tilde{P}_{ij}$ by,

$$P_{ij}^{unnorm} = \tilde{P}_{ij} \exp\left(-\frac{U_j - U_i}{2k_{\mathrm{B}}T}\right), \tag{5.5}$$

where U_i and U_j are the biasing potentials of discretized states i and j. The biasing potential is here defined as the potential subtracted from the system. The resulting unbiased transition probabilities P_{ij}^{unnorm} are not normalized, thus an additional normalization step is needed,

$$P_{ij} = \frac{P_{ij}^{unnorm}}{\sum_j P_{ij}^{unnorm}}. \tag{5.6}$$

It can be easily shown that this procedure is equivalent to reweighting every individual transition $\tilde{b}_{ij}(t,\tau)$ and, subsequently, perform a row normalization (Eq. (5.3)). Thus, DHAM is implemented using,

$$b_{ij}(t,\tau) = \tilde{b}_{ij}(t,\tau) \exp\left(-\frac{U_j - U_i}{2k_\mathrm{B}T}\right), \tag{5.7}$$

where $\tilde{b}_{ij}(t,\tau)$ is a transition of the biased simulation. DHAM is limited to time-independent biasing potentials because U_i and U_j are assumed to be time-independent. The reweighting type of DHAM can be considered 'energy-based' because Eq. (5.5) employs the biasing energies of the start and end states. This reweighting method was successfully applied to calculate membrane crossing rates for drug molecules.[222] Additionally, DHAM was extended to the assumption of the detailed balance kinetics, which was termed 'DHAM extended to detailed balance' (DHAMed),[7] which is similar to the earlier transition-based reweighting analysis method (TRAM).[6]

5.2.3 Girsanov Reweighting

Girsanov reweighting[8,213] is a path-based reweighting method, i.e. it reweights individual trajectory paths. Since different transition

paths can generally contribute to the same transition probability (exemplified in Figure 5.1A), such a path-based method explicitly accounts for the biasing contributions of different pathways. Girsanov reweighting is based on the Girsanov theorem.[223] It describes how the dynamics of a system varies if the probability measure is replaced by an *equivalent* measure.

In contrast to the Smoluchowski equation ansatz of DHAM, path-based reweighting algorithms are derived from the Langevin equation of motion for stochastic dynamics (SD),

$$m_l \frac{d^2 \boldsymbol{r}_l(t)}{dt^2} = \boldsymbol{F}_l(t) - \gamma_l m_l \frac{d\boldsymbol{r}_l(t)}{dt} + \sigma_l \boldsymbol{W}_l(t). \tag{5.8}$$

Here, $\boldsymbol{r}_l(t)$ denotes the position vector of the l^{th} particle with mass m_l. The force $\boldsymbol{F}_l = -\nabla_l V$ is the gradient of the potential energy V and γ_l denotes the static friction coefficient. The fluctuation-dissipation theorem determines the prefactor $\sigma_l^2 = 2k_\mathrm{B}T\gamma_l m_l$ of the random force $\boldsymbol{W}_l(t)$. For simplicity, we drop the particle index l on the constants in the following, i.e. $m_l = m$, $\gamma_l = \gamma$, and $\sigma_l = \sigma$. Additionally, in the subsequent equations the time is indicated by the index for all time-dependent variable such that, for instance, $\boldsymbol{F}_{lk} = \boldsymbol{F}_l(t_k)$.

For reweighting, the Girsanov theorem uses the product of the relative starting probability at configuration x_k and the relative path probability. Hence, Girsanov reweighting depicts the ratio of path probabilities in the biased and unbiased setup as,

$$\omega_{k,k+M_\tau} = \frac{\pi(x_k)}{\tilde{\pi}(x_k)} \cdot \frac{p(x_{k+1}, \ldots, x_{k+M_\tau}|x_k)}{\tilde{p}(x_{k+1}, \ldots, x_{k+M_\tau}|x_k)}\,. \tag{5.9}$$

Note that $M_\tau = \tau/\Delta t$ is the number of simulation frames in the path from $t = t_k$ to $t + \tau = t_k + M_\tau \Delta t$. The phase-space probability ratio can be determined by the Radon-Nikodym

derivative[224] as,

$$g(x_k) = \frac{\pi(x_k)}{\tilde{\pi}(x_k)} = \frac{\tilde{Z}}{Z} \exp(\frac{-U(x_k)}{k_B T}) \,. \tag{5.10}$$

where U denotes the biasing potential, and Z and $\tilde{Z}$ stand for the respective partition functions of the unbiased and biased system.

Similarly, the ratio of the path probability measures is determined by the Radon-Nikodym derivative. N is the number of dimensions.

$$\begin{aligned} \mu_{path} &= \frac{p(x_{k+1}, \ldots, x_{k+M_\tau} | x_k)}{\tilde{p}(x_{k+1}, \ldots, x_{k+M_\tau} | x_k)} \\ &= \exp\Big[\sum_{k}^{k+M_\tau} \left(\sum_{l=1}^{N} \frac{\nabla_l U|_k}{\sigma} \tilde{\boldsymbol{w}}_{lk} \sqrt{\Delta t} - \frac{1}{2} \frac{(\nabla_l U|_k)^2}{\sigma^2} \Delta t \right) \Big] \,. \end{aligned} \tag{5.11}$$

For convenience, we will from here on use the action difference defined by $\Delta S = -\ln(P/\tilde{P})$,

$$\Delta S = \sum_{k}^{k+M_\tau} \left(\sum_{l=1}^{N} \frac{-\nabla_l U|_k}{\sigma} \tilde{\boldsymbol{w}}_{lk} \sqrt{\Delta t} + \frac{1}{2} \frac{(\nabla_l U|_k)^2}{\sigma^2} \Delta t \right) \,. \tag{5.12}$$

Eqs. (5.10) and (5.11) compose the reweighting factor in Eq. (5.9). The reweighting factor is afterwards applied to every individual transition,

$$b_{ij}(t, \tau) = \tilde{b}_{ij}(t, \tau) \cdot \omega_{i,j} \,. \tag{5.13}$$

Note that the partition functions in Eq. (5.10) cancel for time-independent biasing potentials during the row normalization of the count matrix in Eq. (5.3). We will discuss the implication

of the partition function for time-dependent biasing potentials in Section 5.4.

Note that the random number and the force at every time step are required to calculate the relative path probability. Most MD software packages do not store random numbers and a write-out frequency of every step is very storage demanding. Therefore, for most use cases the Girsanov reweighting factor is calculated on the fly during the numerical integration of the MD engine. This requires implementation changes in the MD software in order to use this reweighting approach. In contrast, energy-based reweighting algorithms like DHAM,[46] TRAM,[6] or DHAMed[7] can be applied entirely in a post-processing step. On the plus side, the Girsanov algorithm was shown to be able to handle the time-dependent biasing potentials $U(t)$ during the build-up phase of metadynamics simulations.[8] Possible limitations of time-dependent biasing in conjunction with a rigorous consideration of the partition sums in Eq. (5.10) are discussed below in Section 5.4.

5.2.4 Weber-Pande Reweighting

Although among the first dynamic reweighting methods published, the algorithm proposed by Weber and Pande[45] remained largely unnoticed. In contrast to the rigorous mathematical derivation of Girsanov reweighting, Weber-Pande reweighting is derived via a physical approach based on action differences. Additionally, the phase-space probability reweighting of the start state is not considered, which leads to some important implications that are discussed in Section 5.4.

To follow the path-based reweighting arguments from Weber and Pande, we consider a simulation step $\Delta t = t_{k+1} - t_k$ from the same phase-space start point $\{\boldsymbol{r}_l(t_k), \boldsymbol{v}_l(t_k)\}$ to the same phase-

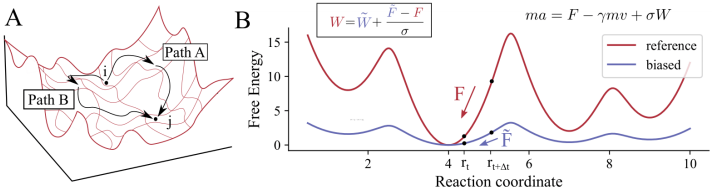


FIGURE 5.1: **A** Different paths can lead to a transition from state i to j. Their corresponding contributions to the reweighting are accounted for in path-based reweighting. **B** Illustration of the basic idea behind the Weber-Pande algorithm.[45] An identical simulation step from x_1 to x_2 is performed on both the original energy potential (red) and the biased potential (blue). The two systems experience different forces $\boldsymbol{F}$ and $\tilde{\boldsymbol{F}}$. As the simulations steps are defined identical, the acceleration and velocity terms cancel from the Langevin equation of motion (Eq. (5.8)). The resulting relationship between the two random numbers and the force terms is $\boldsymbol{W} = \tilde{\boldsymbol{W}} + [\tilde{\boldsymbol{F}} - \boldsymbol{F}]\sigma^{-1}$.

space end point $\{\boldsymbol{r}_l(t_{k+1}), \boldsymbol{v}_l(t_{k+1})\}$ in the unbiased and biased energy landscape (Figure 5.1B).

In order to make the same step length in both the unbiased and biased energy landscapes, the net forces in both cases must be equal such that

$$\begin{aligned} \boldsymbol{F}_{lk} - \gamma m \boldsymbol{v}_{lk} + \sigma \boldsymbol{W}_{lk} &= \tilde{\boldsymbol{F}}_{lk} - \gamma m \boldsymbol{v}_{lk} + \sigma \tilde{\boldsymbol{W}}_{lk} \\ \boldsymbol{W}_{lk} &= \tilde{\boldsymbol{W}}_{lk} + [\tilde{\boldsymbol{F}}_{lk} - \boldsymbol{F}_{lk}]\sigma^{-1} \,. \end{aligned} \tag{5.14}$$

where quantities in the biased trajectory are labeled with tilde, e.g. $\tilde{\boldsymbol{F}}$. Note that $\boldsymbol{v}_{lk}$ is identical in the biased and unbiased simulations because we demand that both systems end and start at the same phase-space points.

Eq. (5.14) is used to determine the relative path probability of the biased and unbiased simulations. The relative path probability is defined as the probability to sample the sequence of random

vectors required for the unbiased simulation $[\boldsymbol{W}_{1k}, \boldsymbol{W}_{2k}, \ldots, \boldsymbol{W}_{lk}, \ldots, \boldsymbol{W}_{N(k+M_\tau)}]$ in comparison to the sequence $[\tilde{\boldsymbol{W}}_{1k}, \tilde{\boldsymbol{W}}_{2k}, \ldots, \tilde{\boldsymbol{W}}_{lk}, \ldots, \tilde{\boldsymbol{W}}_{N(k+M_\tau)}]$ of the actually performed biased simulation.

The path reweighting term depends on the numerical representation of $\boldsymbol{W}_{lk}$. Due to the discretization the reweighting becomes integrator dependent. In their publication, Weber and Pande chose $\boldsymbol{W}_{lk} \approx \boldsymbol{w}_{lk}\sqrt{\Delta t}^{-1}$, where $\boldsymbol{w}_{lk}$ is sampled from a Gaussian distribution with zero mean and unit variance.[45] This discretization is based on the Brunger-Brooks-Karplus (BBK) integrator.[225,226] The BBK integrator assumes weak coupling between the system and the surrounding.[226] Therefore, in contrast to Girsanov reweighting, Weber-Pande is derived for the low friction limit. Insertion into Eq. (5.14) yields,

$$\boldsymbol{w}_{lk} = \tilde{\boldsymbol{w}}_{lk} + [\tilde{\boldsymbol{F}}_{lk} - \boldsymbol{F}_{lk}]\sigma^{-1}\sqrt{\Delta t}\,. \tag{5.15}$$

Since both $\boldsymbol{w}_{lk}$ and $\tilde{\boldsymbol{w}}_{lk}$ are Gaussian random numbers with zero mean and unit variance, the relative probability of $\boldsymbol{w}_{lk}$ and $\tilde{\boldsymbol{w}}_{lk}$ is given by,

$$\mu_{path} = \exp\left(\sum_{k}^{k+M_\tau}\sum_{l=1}^{N} -\frac{\left(\boldsymbol{w}_{lk}^2 - \tilde{\boldsymbol{w}}_{lk}^2\right)}{2}\right), \tag{5.16}$$

and consequently is the action difference of the biased and unbiased transition,

$$\Delta S_{k,k+M_\tau} = \sum_{k}^{k+M_\tau}\sum_{l=1}^{N} \frac{\left(\boldsymbol{w}_{lk}^2 - \tilde{\boldsymbol{w}}_{lk}^2\right)}{2}\,. \tag{5.17}$$

This means that the path reweighting factor in the Weber-Pande scheme is the ratio of transition probabilities. Note that

the relative phase-space probability of the initial state as in Eq. (5.10) was not considered. This leads to some implications which are discussed in Section 5.4. The reweighting of the MSM is performed on the individual transitions such that

$$b_{ij}(t,\tau) = \tilde{b}_{ij}(t,\tau) \cdot \mu_{i,j} \,. \tag{5.18}$$

Weber-Pande reweighting has similar advantages and limitations as the Girsanov algorithm, i.e., it can reweight time-dependent biases, but has to be incorporated in the numerical integration of the MD engine. A detailed comparison of their relative advantages and disadvantages is given in Section 5.4.

5.3 CONNECTION BETWEEN WEBER-PANDE AND GIRSANOV REWEIGHTING

We will formulate a generalized path-based reweighting algorithm inspired by the work of Weber and Pande.[45] In contrast to Weber-Pande reweighting, we will assume high frictional diffusion processes and use Ito diffusion, an alternative formulation of the Langevin equation,

$$\gamma m \; d\boldsymbol{r}_l(t) = \boldsymbol{F}_l(t)dt + \sigma d\boldsymbol{W}_l(t). \tag{5.19}$$

Here, $\boldsymbol{dr}_l(t)$ denotes the infinitesimal position vector shift of the l^{th} particle with mass m. $d\boldsymbol{W}_l(t)$ is the increment of a Wiener process, which has the following properties: $d\boldsymbol{W}_l(t)$ is Gaussian with zero mean, i.e. $\langle d\boldsymbol{W}_l(t)\rangle = \boldsymbol{0}$ and variance $\langle d\boldsymbol{W}_l(t) d\boldsymbol{W}_l(t)\rangle = dt$.

Additionally, the increments are independent of the past values and independent across particles.

Following the reasoning of Weber and Pande, we consider an infinitesimal step dt in the unbiased and biased energy landscape that leads to the same displacement $d\boldsymbol{r}_l(t)$ such that

$$d\boldsymbol{W}_l(t) = (\tilde{\boldsymbol{F}}_l(t) - \boldsymbol{F}_l(t))\sigma^{-1}dt + d\tilde{\boldsymbol{W}}_l(t)\,. \tag{5.20}$$

The instantaneous relative action differential of the biased and unbiased simulation is given by,

$$dS_{inst} = \sum_{l=1}^{N} \frac{\left(d\boldsymbol{W}_l(t)^2 - d\tilde{\boldsymbol{W}}_l(t)^2\right)}{2\langle d\boldsymbol{W}_l(t) d\boldsymbol{W}_l(t)\rangle}\,. \tag{5.21}$$

Inserting the connection between the random numbers from Eq. (5.20) in Eq. (5.21) yields a relationship that only contains parameters, which are either directly obtained from the biased system (i.e. $\tilde{\boldsymbol{F}}_l(t)$ and $d\tilde{\boldsymbol{W}}_l(t)$) or defined by the unbiased potential (i.e. the unbiased force $\boldsymbol{F}_l(t)$). We abbreviate $\Delta\boldsymbol{F}_l(t) = \tilde{\boldsymbol{F}}_l(t) - \boldsymbol{F}_l(t)$ such that the relative action reads,

$$\begin{aligned} dS_{inst} &= \sum_{l=1}^{N} \left(\frac{(\Delta\boldsymbol{F}_l(t)\sigma^{-1}dt + d\tilde{\boldsymbol{W}}_l(t))^2 - d\tilde{\boldsymbol{W}}_l(t)^2}{2\langle d\boldsymbol{W}_l(t) d\boldsymbol{W}_l(t)\rangle} \right) \\ &= \sum_{l=1}^{N} \left(\frac{2\Delta\boldsymbol{F}_l(t)\sigma^{-1}dt d\tilde{\boldsymbol{W}}_l(t) + \Delta\boldsymbol{F}_l(t)^2\sigma^{-2}dt^2}{2dt} \right) \\ &= \sum_{l=1}^{N} \left(\frac{\Delta\boldsymbol{F}_l(t)}{\sigma} d\tilde{\boldsymbol{W}}_l(t) + \frac{\Delta\boldsymbol{F}_l(t)^2}{2\sigma^2} dt \right)\,. \end{aligned} \tag{5.22}$$

To compare Eq. (5.22) with Girsanov reweighting, we need to discretize, in particular $d\boldsymbol{W}_l(t)$ and dt. Note that this discretization is integrator dependent. Girsanov reweighting was derived

for the Euler-Maruyama integrator, therefore we will use the same integrator here. In the special case of the Euler-Maruyama integrator, one uses $d\boldsymbol{W}_l(t_k) = \boldsymbol{w}_{lk}\sqrt{\Delta t}$, where $\boldsymbol{w}_{lk}$ is the discrete random vector drawn at time t_k. Similarly, the force at time t_k is termed $\boldsymbol{F}_{lk}$. Inserting the Euler-Maruyama discretization into Eq. (5.22) yields the relative action for a single step,

$$\Delta S_{step} = \sum_{l=1}^{N} \left(\frac{\Delta \boldsymbol{F}_{lk}}{\sigma} \tilde{\boldsymbol{w}}_{lk} \sqrt{\Delta t} + \frac{(\Delta \boldsymbol{F}_{lk})^2}{2\sigma^2} \Delta t \right) \tag{5.23}$$

$$= \sum_{l=1}^{N} \left(-\frac{\nabla_l U|_k}{\sigma} \tilde{\boldsymbol{w}}_{lk} \sqrt{\Delta t} + \frac{(\nabla_l U|_k)^2}{2\sigma^2} \Delta t \right) \tag{5.24}$$

In a final step, we use the fact that the force difference between the biased and unbiased systems is exactly determined by the biasing potential, i.e., $\Delta \boldsymbol{F} = -\nabla U$. Thus, the formulation of the relative probability in Eq. (5.24) is identical to the relative path probability in the Girsanov formulation (Eq. (5.11)). Therefore, Girsanov and Weber-Pande reweighting can be deduced from the same principles.

An interesting coincidence is that, even though Weber-Pande reweighting was derived in the low and Girsanov reweighting in the high friction limit, both yield the same transition reweighting term. In particular, insertion of Eq. (5.15) into Eq. (5.17) yields,

$$\Delta S_{path} = \sum_{k}^{k+M_\tau} \sum_{l=1}^{N} \left(\frac{\Delta \boldsymbol{F}_{lk}}{\sigma} \tilde{\boldsymbol{w}}_{lk} \sqrt{\Delta t} + \frac{(\Delta \boldsymbol{F}_{lk})^2}{2\sigma^2} \Delta t \right) , \tag{5.25}$$

which is identical to Eq. (5.23) that leads to the Girsanov reweighting formulation. This observation is discussed in more detail in Section 5.7.1.

The two path-based reweighting algorithms Weber-Pande and

Girsanov have been developed independently by different groups. The first method was derived by comparing the relative probabilities S of Gaussian random forces in the biased and unbiased cases, whereas the latter is based on the Girsanov theorem. However, both methods can be traced back to a common path-reweighting concept and therefore share some similarities.

5.4 DIFFERENCES BETWEEN WEBER-PANDE AND GIRSANOV REWEIGHTING

The main difference between Girsanov and Weber-Pande reweighting lies in Eq. (5.10), namely the ratio of the phase-space probability of the starting configuration. This term is present in Girsanov reweighting but not in the Weber-Pande approach. The physical intuition behind the phase-space probability term is depicted in Figure 5.2A.1. Imagine that one combines a phase-space region, in which phase-space points were biased differently, into a single cluster n. We will now compare two of these phase-space points r_1 and r_2 within cluster n that have a weak and a strong bias, respectively. The path reweighting term of Weber-Pande and Girsanov will correctly reweight the conditional transition probability (Eq. (5.9)). In other words, the transition probability under the condition that the path started in r_1 or r_2 is reweighted correctly. However, the relative probability of starting a simulation in r_1 or r_2 changes due to their different biases. In our example, r_2 has a higher equilibrium distribution on the biased energy surface, because it is biased more strongly. Thus, more

simulations will start in r_2 then in r_1 in the biased case, but not in the unbiased one. This can create an imbalance in the count matrix, if the conditional transition probabilities of r_1 or r_2 differ. Eq. (5.10) accounts for this imbalance by reweighting the probability to start in r_1 or r_2. The Weber-Pande approach on the other hand does not reweight the relative starting probability within the cluster. Instead, it implicitly assumes that the biasing potential $U_i(t)$ is identical for each member of the cluster. In this case, the phase-space reweighting factor can be moved in front of the path summation (Eq. (5.1)) and cancels during the row normalization. This assumption is approximately true only for very fine MSM clusters. Note that DHAM uses the same approximation and thus, also requires very fine clustering. In contrast, Girsanov reweighting is not limited to small cluster sizes.

For time-dependent biases, we have to consider two effects. First, Eq. (5.10) assumes that the starting probabilities at configuration x_k are Boltzmann distributed. This assumption can be violated for time-dependent biases that push the system out of equilibrium. This problem can be minimized using biases that only lead to minimal non-equilibrium perturbations. However, even in the case of Boltzmann distributed starting probabilities a second problem occurs. The biased partition function $\tilde{Z}$ in Eq. (5.10) becomes time-dependent. Thus, $\tilde{Z}$ does not cancel in the row-normalization of the count matrix,

$$\begin{aligned} P_{ij} &= \frac{C_{ij}}{\sum_{j=1}^{n} C_{ij}} \\ &= \frac{\sum_{t=1}^{N_\tau} \tilde{b}_{ij}(t,\tau) \cdot \mu_{path} \cdot \frac{\tilde{Z}(t)}{Z} \exp(\frac{-U_i(t)}{k_B T})}{\sum_{j=1}^{n} \sum_{t=1}^{N_\tau} \tilde{b}_{ij}(t,\tau) \cdot \mu_{path} \cdot \frac{\tilde{Z}(t)}{Z} \exp(\frac{-U_i(t)}{k_B T})} \,. \end{aligned} \tag{5.26}$$

Therefore, for each time step the knowledge of the partition

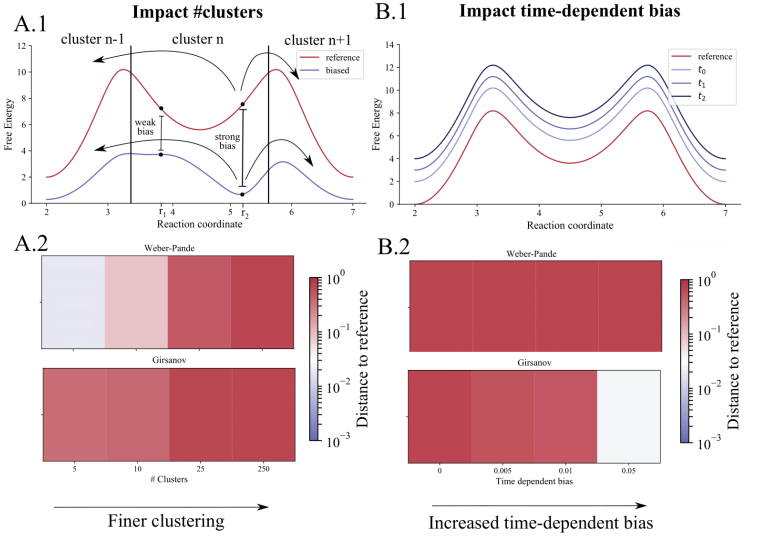


FIGURE 5.2: Relation between phase-space reweighting and clustering as well as time-dependent biases. (**A.1**) Phase-space reweighting corrects for differently biased start states in the same cluster. Due to row normalization the phase-space reweighting factor cancels if all cluster members have the same bias. However, the coarser the clustering and steeper the potential, the greater the phase-space reweighting correction. (**A.2**) Heatmaps showing the distance factor D between the reweighted MSMs and the reference in the limiting case of very coarse clustering. The phase-space reweighting factor of the Girsanov algorithm is able to correct the imbalance imposed by the coarse clustering. Weber-Pande lacks this factor, which leads to inaccurate reweighting results. The inaccuracy of Girsanov reweighting for low cluster numbers is due to MSM errors and not the reweighting error. (**B.1**) Representation of the time-dependent but position-independent biasing potential $U(t) = const \cdot t$. (**B.2**) Heatmaps showing the distance factor D between the reweighted MSMs and the reference in the limiting case of a strong time-dependent biasing potential. The neglected time-dependency of the Girsanov phase-space reweighting factor leads to inaccurate reweighting results for strong time-dependent biases.

function $\tilde{Z}$ would be necessary to reweight accurately. In the Girsanov reweighting approach, however, it is assumed that the partition function cancels in the row normalization,[8] which is only true for time-independent biases. The implication of this second issue is depicted in Figure 5.2B.1. For didactic purposes, we chose a biasing potential that is time- but not position-dependent $U(t) = const \cdot t$. Thus, the energy surface is shifted along the y-axis for every step. The starting probabilities remain Boltzmann distributed and the first issue does not apply. Therefore, this bias can be used to observe the effect of the second issue only. This bias neither changes the system's kinetics nor its equilibrium distribution. However, due to the phase-space reweighting factor in Eq. (5.10), the Girsanov formulation gives higher weights to transitions with a higher bias (i.e., transitions that occur later in the simulation). This again creates an imbalance in the count matrix. The effect of the imbalance can be minimized by choosing time-dependent biases that decay over time, e.g., well-tempered metadynamics.[227]

In summary, the phase-space reweighting factor Eq. (5.10) has two consequences. It removes the intra-cluster imbalance in the case of differently biased sub-states, but creates an imbalance for time-dependent biases. By choosing fine-grained clusters or applying well-suited enhanced sampling techniques, these effects can be minimized. One has to carefully choose if one includes the phase-space reweighting factor based on the combination of the system and biasing technique.

5.5 AN INTEGRATOR INDEPENDENT REWEIGHTING SCHEME AND THE CONNECTION TO DHAM

Both Girsanov and Weber-Pande reweighting used a specific discretization scheme and are therefore integrator dependent. We start from the discretization independent form in Eq. (5.22) to derive an integrator independent formulation for path-based reweighting. Additionally, we show that it intrinsically contains the DHAM formalism.

We start our derivation from the action difference formulation in Eq. (5.22). Note that this formulation is only valid in the overdamped regime (i.e., high friction), which is a good assumption for molecules in aqueous solution,

$$dS_{inst} = \sum_{l=1}^{N} \left(\frac{\Delta \boldsymbol{F}_l(t)}{\sigma} d\tilde{\boldsymbol{W}}_l(t) + \frac{\Delta \boldsymbol{F}_l(t)^2}{2\sigma^2} dt \right) .$$

Furthermore, the assumption of an overdamped system is also part of the derivation of the Smoluchowski diffusion equation and therefore an intrinsic feature of DHAM.[46] Solving Eq. (5.19) for $d\tilde{\boldsymbol{W}}_l(t)$ and introducing $\Delta \boldsymbol{F}_l(t) = \tilde{\boldsymbol{F}}_l(t) - \boldsymbol{F}_l(t)$ back into Eq. (5.22)

leads to

$$
\begin{aligned}
dS_{inst} &= \sum_{l=1}^{N} \frac{\tilde{\boldsymbol{F}}_l(t) - \boldsymbol{F}_l(t)}{\sigma^2}(\gamma m \; d\boldsymbol{r}_l(t) - \tilde{\boldsymbol{F}}_l(t)dt) \\
&+ \sum_{l=1}^{N} \frac{(\tilde{\boldsymbol{F}}_l(t) - \boldsymbol{F}_l(t))^2}{2\sigma^2}dt \\
&= \sum_{l=1}^{N} \frac{\tilde{\boldsymbol{F}}_l(t) - \boldsymbol{F}_l(t)}{\sigma^2}\gamma m \; d\boldsymbol{r}_l(t) \\
&+ \sum_{l=1}^{N} \frac{(\tilde{\boldsymbol{F}}_l(t) - \boldsymbol{F}_l(t))^2 - 2\tilde{\boldsymbol{F}}_l(t)(\tilde{\boldsymbol{F}}_l(t) - \boldsymbol{F}_l(t))}{2\sigma^2}dt \\
&= \sum_{l=1}^{N} \frac{\tilde{\boldsymbol{F}}_l(t) - \boldsymbol{F}_l(t)}{\sigma^2}\gamma m \; d\boldsymbol{r}_l(t) + \sum_{l=1}^{N} \frac{\boldsymbol{F}_l(t)^2 - \tilde{\boldsymbol{F}}_l(t)^2}{2\sigma^2}dt \\
&= \sum_{l=1}^{N} \frac{\Delta \boldsymbol{F}_l(t)}{2k_{\mathrm{B}}T}d\boldsymbol{r}_l(t) + \sum_{l=1}^{N} \frac{\boldsymbol{F}_l(t)^2 - \tilde{\boldsymbol{F}}_l(t)^2}{2\sigma^2}dt. \qquad (5.27)
\end{aligned}
$$

For the last step, the fluctuation-dissipation theorem determines $\sigma^2 = 2k_{\mathrm{B}}T\gamma m$. This formulation of the action difference between biased and unbiased system is independent of the random force realizations $d\tilde{\boldsymbol{W}}_l(t)$. In particular, the random term is mapped by the position shift $d\boldsymbol{r}_l(t)$ and thus, Eq. (5.27) is independent of the numerical integrator.

This leads to three important technical implications:
(1) The requirement to know the random number at every time step is currently one of the main limitations of path-based reweighting algorithms. Many higher-level SD integrators work with multiple random numbers, or with random numbers that add to the velocities (random impulses) and not to the forces.[228,229] Consequently, the path-dependent reweighting algorithms have to be derived for each integrator separately. In some cases, it might even be impossible to find a suitable modification. Therefore, a

random-number independent path-based reweighting algorithm would be preferable.
(2) The random number is uncorrelated between two time steps, which makes it necessary to reweight at the same time step as the integration. In contrast, the force $\boldsymbol{F}$ may be correlated between time steps. Thus, it might be possible to write out the forces, like the positions, at a lower frequency. However, if and how much the write-out frequency can be reduced is still speculative and subject to future research.
(3) In relation to the first point, path-based reweighting algorithms are currently limited to SD simulations. Eq. (5.27) can also be applied with MD simulations. In that case, an approximation of γ is required which is a problem-specific but straightforward task in MD simulation.[230–232] The potential of this path correction will be explored in future work.

In a final step, we can relate the formulation of Eq. (5.27) to DHAM. We take DHAM's assumption of a time-independent biasing potential U. In order to obtain the relative action for the whole path, we integrate over the instantaneous actions,

$$\begin{aligned}\Delta S &= \int_{path} dS_{inst} \\ &= \sum_{l=1}^{N} \int_{\boldsymbol{r}_l(t)}^{\boldsymbol{r}_l(t+\tau)} \frac{\Delta \boldsymbol{F}_l(\boldsymbol{r})}{2k_\mathrm{B}T} \mathrm{d}\boldsymbol{r} + \sum_{l=1}^{N} \int_{t}^{t+\tau} \frac{\boldsymbol{F}_l(t)^2 - \tilde{\boldsymbol{F}}_l(t)^2}{2\sigma^2} dt.\end{aligned} \tag{5.28}$$

As we have defined before, the force at time t_k is termed F_{lk} and $\Delta F_{lk} = -\nabla_l U|_k$ is exactly the negative gradient of the biasing potential of particle l at time t_k. Hence, for deterministic (non-stochastic) processes one can obtain from Eq. (5.28) directly

the discrete form,

$$\Delta S = \frac{U(r_{k+M_\tau}) - U(r_k)}{2k_\mathrm{B}T} + \sum_{l=1}^{N}\sum_{k}^{k+M_\tau} \frac{\boldsymbol{F}_{lk}^2 - \tilde{\boldsymbol{F}}_{lk}^2}{2\sigma^2}\Delta t. \quad (5.29)$$

If the configuration of $\{\boldsymbol{r}_{lk}\}$ lies in the (MSM) state i and the terminal configuration $\{\boldsymbol{r}_{lk+M_\tau}\}$ lies in a different state j, we obtain,

$$\frac{P_{k,k+M_\tau}}{\tilde{P}_{k,k+M_\tau}} = \exp\left(-\frac{U_j - U_i}{2k_BT}\right) \times \exp\left(\sum_{l=1}^{N}\sum_{k}^{k+M_\tau} -\frac{\boldsymbol{F}_{lk}^2 - \tilde{\boldsymbol{F}}_{lk}^2}{2\sigma^2}\Delta t\right). \quad (5.30)$$

For stochastic processes, the displacement $\mathrm{d}\boldsymbol{r}$ consists of an deterministic and stochastic proportion. Based on the principle of superposition, one can split these proportions apart, i.e., $\mathrm{d}\boldsymbol{r} = \mathrm{d}\boldsymbol{r}_{determ} + \mathrm{d}\boldsymbol{r}_{stoch}$. The integral of the force over the deterministic displacement will again yield the energies. For small contributions of the stochastic displacement $\mathrm{d}\boldsymbol{r}_{stoch}$ one obtains,

$$\Delta S = \frac{U_j - U_i}{2k_\mathrm{B}T} + \sum_{l=1}^{N}\sum_{k}^{k+M_\tau} \frac{\boldsymbol{F}_{lk}^2 - \tilde{\boldsymbol{F}}_{lk}^2}{2\sigma^2}\Delta t + \mathcal{O}(\int \Delta\boldsymbol{F}\mathrm{d}\boldsymbol{r}_{stoch}). \quad (5.31)$$

Eq. (5.31) illustrates that path-based reweighting and DHAM share the same (path independent) energy term, but differ in the path-dependent term and the stochastic displacement term. The magnitude of the deviation will be determined by the specific path taken, the length of the path, and the (random) forces encountered along this path.

Starting from the path reweighting term, we showed that the

reweighting can be split into a path-independent and a path-dependent term (Eq. (5.27)). For deterministic processes, the path-independent term is identical to the DHAM reweighting formula, whereas the path-dependent term can be viewed as a correction factor. The contribution of the path correction increases with the number of steps M_τ in a path, which is investigated with a four-well test system in Section 5.7.3. Note however, that the phase-space probability reweighting of the initial state has to be performed in addition to the path reweighting as in Eq. (5.10).

5.6 METHODS

The reweighting performance of the different algorithms was assessed using a simple one-dimensional (1D) four-well system. The system contains three distinct energy barriers, which correspond to the slowest kinetics of the system. In order to accelerate barrier crossing, the system was perturbed by scaling the potential energy with a factor of 0.2. This corresponds to a biasing potential of $U(x) = 0.8V(x)$, where $V(x)$ is the potential energy of the system. The ability of a reweighting algorithm to recover both the equilibrium population as well as the equilibrium kinetics was evaluated. Reweighting was performed with the Weber-Pande (Eq. (5.16)) and Girsanov (Eq. (5.11)) reweighting algorithms as well as with the original DHAM (Eq. (5.5)) and DHAM with path correction (Eq. (5.30)). Additionally, the reweighting was carried out for different MSM parameters to assess the (in)dependence of the chosen parameter set.

The energy surface of the system is defined by
$V(x) = -4\ln(e^{-2(x-1.5)^2-2}+e^{-2(x-4)^2}+0.6e^{-2(x-7)^2}+e^{-2(x-9)^2-1})$.

The simulations were performed using stochastic dynamics with a BBK integrator[225] and a time step of $\Delta t = 0.005$. The parameters for the Langevin equation of motion were set as follows: $\gamma = 50$, $m = 1$, and $k_B T = 1.5$. All quantities are in arbitrary units. The parameters were chosen to fulfill the requirement $\gamma \Delta t < 1$ for the BBK integrator and to yield reasonably short simulation times. In general, dynamic-reweighting methods are stable for different parameter choices. However, if the bias is very large in comparison to σ, the relative path probability becomes very small, which can lead to numerical instabilities. Forty trajectories with a length of 400'000 steps each were simulated for the biased systems, starting at 40 equally spaced points on the energy surface. 200 reference simulations with 4'000'000 steps each were performed for the original system. The simulation trajectories were clustered in 250 microclusters based on their x-coordinate. For the MSM transition matrix, counts were obtained using a sliding window approach. A lagtime of 4'000 steps was chosen for Weber-Pande and Girsanov reweighting. For DHAM reweighting, a lagtime of 40 steps was used. The counts were reweighted by each of the dynamic reweighting algorithms. Source code to perform both the simulations and reweighting is provided on Github (see Section 5.6.1).

The distance factor D describes how accurate the reweighting algorithm recovers the reference kinetics. To calculate D, we use the MSM relaxation timescale of the slowest process t_2^{ref} and $t_2^{reweight}$ of the reference and the reweighted system, respectively. The MSM relaxation timescale is defined in Eq. (5.4). Thus, D is computed using,

$$D = \frac{min(t_2^{ref}, t_2^{reweight})}{max(t_2^{ref}, t_2^{reweight})}. \tag{5.32}$$

The distance factor is always $\leq$ 1, where a value of 1 corresponds to perfect agreement between reference and reweighted kinetics.

5.6.1 Data and Software Availability

Source code to perform the simulations and reweighting of the 1D potentials as well as to reproduce all figures of this manuscript is available on GitHub (`https://www.github.com/rinikerlab/Dynamic_Reweighting_FamilyTree`).

5.7 RESULTS AND DISCUSSION

5.7.1 Transition Matrix Comparison Between Weber-Pande and Girsanov Reweighting

As shown in Section 5.3, the Weber-Pande and Girsanov algorithms share the same path-reweighting factor (Eqs. (5.11), (5.25)) but differ in the phase-space reweighting factor (Eq. (5.10)). In the following, we showcase the discussed theoretical implications on a simple test system. First, we tested the deviation between Weber-Pande and Girsanov reweighting on the level of the transition matrix. An MSM with fine clustering was built on the biased simulations and reweighted to the original energy surface (Figure 5.3A) by either the Weber-Pande or Girsanov approach. The resulting reweighted transition matrices were compared. Figure 5.3B depicts the difference between the transition matrices if only the path reweighting term of Weber-Pande and Girsanov is used. Every square of the 250x250 heat map represents a transition probability between a start state i and an end state j. The

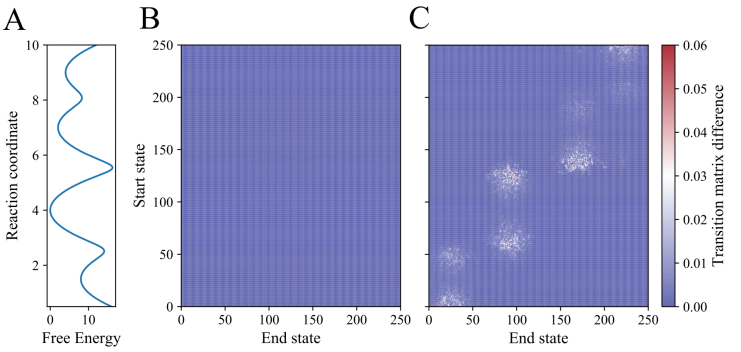


FIGURE 5.3: Potential-energy surface of the four-well test system (**A**), and the difference between the transition matrices of Weber-Pande and Girsanov reweighting without (**B**) and with (**C**) the phase-space reweighting factor for Girsanov. The transition matrix is visualized as a heatmap with elements for the 250 start states i and end states j. Blue colors encode no difference, red colors encode high differences.

color indicates the difference between the Weber-Pande and Girsanov transition matrices. Blue regions correspond to zero deviations, white regions indicate absolute differences of up to 3% and red regions absolute differences of up to 6%. As expected by theory, there are no differences in the transition matrix in Figure 5.3B. This confirms that the path-reweighting terms are identical. In contrast, Figure 5.3C shows the difference between the transition matrices for the full reweighting approach, i.e. it includes the phase-space reweighting factor g for Girsanov. Most regions show still no difference between the two methods. Only transitions that start from a high-energy cluster deviate around 3%. In our simple test system, clusters comprising high-energy states have a higher intra-cluster variance in the bias energy. In other words, not all members of the start cluster i have exactly the same bias energy. As the bias energy contributes exponentially

to the phase-space reweighting factor, a small relative change in the bias energy may lead to a large change in the likelihood factor g. However, for sufficiently fine-grained clustering the difference between Weber-Pande and Girsanov is small. In the presented case, we observed only minor deviations around 3%, which were limited to a few regions.

Second, to explore the differences between Weber-Pande and Girsanov reweighting further, we simulated extreme cases where the phase-space reweighting factor g is expected to have the highest impact. For this, we tested the reweighting of an MSM as a function of (a) the number of clusters, and (b) the magnitude of a time-dependent biasing potential (but coordinate-independent, i.e., $U(t) = const \cdot t$). Figure 5.2A.2 shows the distance between the reweighted MSMs and the reference for increasingly finer clustering (with a time-independent bias). Red color indicates good agreement with the reference, blue color indicates larger deviations. As expected by theory, Weber-Pande reweighting is inaccurate for a small (coarse) number of clusters, because the bias for the different members of the cluster is not homogeneous, which leads to an imbalance in the count matrix. Girsanov reweighting, on the other hand, accounts for this imbalance with the phase-space reweighting factor, and thus the reweighted results are close to the reference.

Figure 5.2B.2 shows the distance between the reweighted MSMs and the reference for increasing time-independent bias (and fine-grained clustering). With increasing magnitude of the bias, Girsanov reweighting becomes more inaccurate, because the phase-space reweighting factor g leads to a stronger contribution of transitions that have a higher bias, i.e occur late in the simulation. This leads again to an imbalance in the count matrix. In contrast, the Weber-Pande approach without the phase-space reweighting

factor gives accurate reweighting results.

We want to stress that Figure 5.2 explores extreme cases to emphasis the impact of the phase-space reweighting factor g on reweighting. For most use cases, the difference between Weber-Pande and Girsanov reweighting is likely marginal. Nevertheless, it is important to keep in mind the difference between the two path-reweighting algorithms when choosing a reweighting approach.

5.7.2 Reweighting Results for Weber-Pande, Girsanov, and DHAM

The performance of different dynamic reweighting algorithms (Weber-Pande, Girsanov, and DHAM) was tested on a four-well system (Figure 5.4A). The ability to obtain both accurate equilibrium populations as well as kinetics from biased simulations was assessed. The original potential (red) contained high energy barriers that were significantly lowered in the biased potential (blue). The histograms from the 40 separate simulations on the biased energy surface are shown in Figure 5.4B. An MSM was built from these simulations and reweighted by Weber-Pande, Girsanov, and DHAM, respectively. A second MSM was built on the 200 unbiased simulations. The kinetics of this reference system are displayed in Figure 5.4C ("Reference"). Three distinct slow processes can be observed that correspond to the three barrier crossings in the four-well system. The numerical values of the timescales are given in Table 5.1. As expected, the MSM from the biased simulations yielded kinetics that were up to 300 fold faster due to the reduced barrier heights (Figure 5.4C: "Not reweighted"). The reweighted kinetics (Figure 5.4C: "Weber-Pande", "Girsanov", "DHAM") are able to recover the original

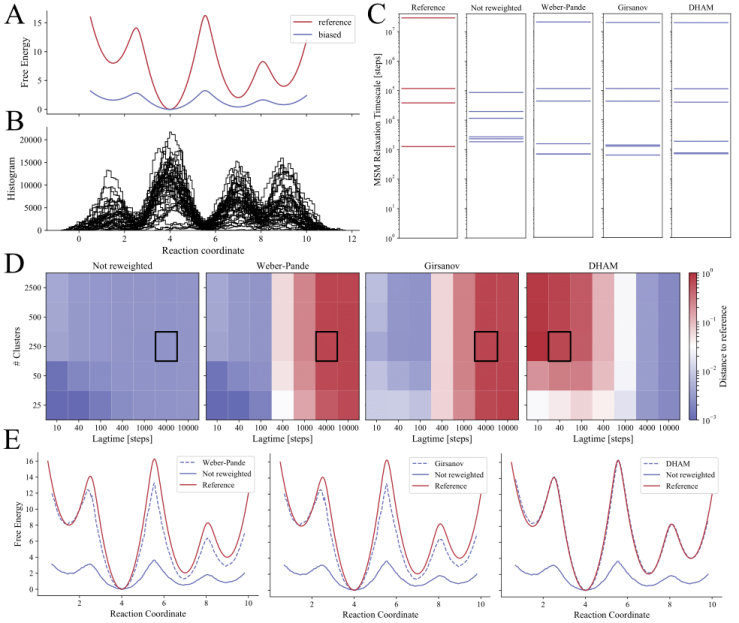


FIGURE 5.4: Comparison of Weber-Pande, Girsanov and, DHAM dynamic reweighting on a four-well 1D test system. (**A**) Free-energy landscape of the unbiased system (red) and the biased system (blue). (**B**) Frequency of visited states along the reaction coordinate for each of the 40 biased simulations. (**C**) Kinetic spectra of the unbiased MSM (Reference), of the MSM from the biased simulations (Not reweighted), and of the MSMs obtained by reweighting with Weber-Pande, Girsanov, and DHAM. (**D**) Heatmaps showing the distance factor D between the reweighted MSMs from Weber-Pande, Girsanov, and DHAM with the reference MSM as a function of the number of clusters of the MSM and the lagtime (red: good agreement, blue: no agreement). Parameter sets used for **C** and **E** are highlighted with a black box. (**E**) Reweighted free-energy landscape (blue, dotted) in comparison with the reference (red) and the biased landscape (blue).

kinetic pattern. The two path-based reweighting methods, Weber-Pande and Girsanov, give very similar results and differ only by a factor 1.4 from the reference (see also Table 5.1). In contrast, reweighting results obtained with DHAM are slightly different. The kinetic results are also within 1.4 of the reference, however, a much shorter lagtime was necessary to obtain good results.

This issue becomes more evident when varying the MSM input parameters to test the stability of the model for different parameter choices. Figure 5.4D shows the results of this parameter test for the slowest process using heatmaps. The number of clusters of the MSM was varied between 25 and 2'500 and the lagtime between 10 and 10'000 frames. These extreme ranges were chosen to test the limits of the reweighting algorithms. All reweighting methods showed improvement over the raw biased MSM ("Not reweighted"). The path-based algorithms were again nearly identical and also showed similar stability. They only deviate from each other for small numbers of clusters, where Girsanov reweighting outperforms Weber-Pande. This is in line with the observations from theory and Section 5.7.1. For large numbers of clusters, on the other hand, the contribution of the phase-space reweighting factor g (Eq. (5.10)) approximately cancels out in the transition matrix. Thus, similar results are obtained in these cases. Furthermore, both path-based algorithms performed poorly at small lagtimes (<400 steps) and well at high lagtimes (>1'000 steps). DHAM, on the other hand, showed the reversed lagtime dependency, i.e. it performed well for small lagtimes (<1'000 steps) and poorly at high lagtimes (>4'000 steps). Additionally, a sufficiently large number of clusters is necessary for good performance . It is known that DHAM needs relatively fine-grained clustering in order to keep the discretization error of the MSM small.[214] For our test system, DHAM showed the best per-

formance of the three reweighting algorithms, within 1% of the reference for 250 clusters and a lagtime of 10.

In addition to the kinetics, we also tested the performance of the reweighting algorithms to recover the equilibrium distribution (Figure 5.4E). All reweighting algorithms performed reasonably well. Again the path-based approaches gave very similar results, however, the reweighted potential-energy surface showed distinct deviations from the reference one. In contrast, DHAM reweighting recovered the reference potential-energy surface nearly perfectly, even though the kinetic results were similar to those from Weber-Pande and Girsanov.

In summary, the path-based reweighting methods performed very similar to each other as shown already in sSection 5.2. For DHAM, we observed a different parameter dependency than for the path-based methods. When choosing the optimal MSM-parameter combination for each reweighting method, the results from DHAM reweighting were closest to the reference for the chosen four-well test system. This is especially remarkable as DHAM only uses information about the first and last step of the path (in contrast to the path-based methods, where the reweighting factor has to be calculated for every simulation step within the path). Thus, DHAM reweighting requires less input data and user effort. However, the best choice for a reweighting method will depend of course on the use case and the enhanced sampling method used.

5.7.3 Path Correction Expands DHAM to Long Lagtimes

In order to test the connection between DHAM and Weber-Pande, we additionally reweighted the MSM with Eq. (5.30), which we will term “DHAM + path correction”. We named it path correc-

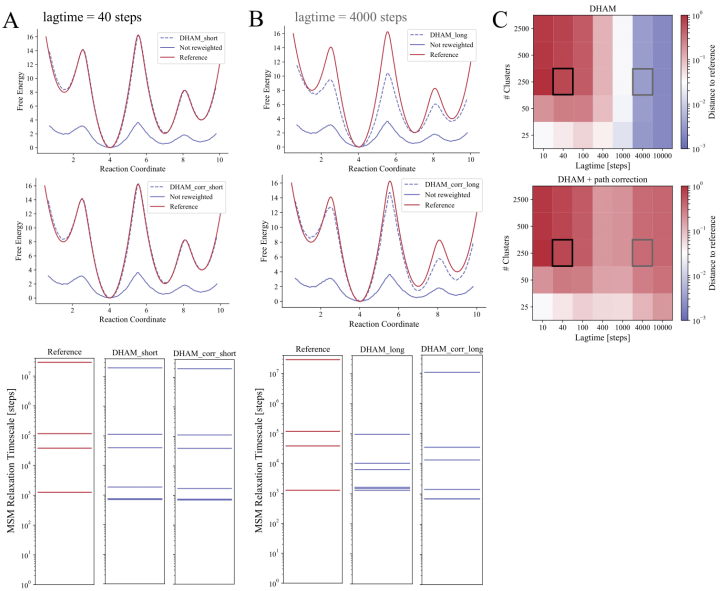


FIGURE 5.5: Reweighting performance at a lagtime of 40 steps (**A**) and 4'000 steps (**B**). Top panel: DHAM reweighted free-energy landscape (blue, dotted) in comparison with the reference (red) and the biased landscape (blue). Middle panel: "DHAM + path correction" reweighted free-energy landscape (blue, dotted) in comparison with the reference (red) and the biased landscape (blue). Bottom panel: Kinetic spectra of the reference MSM, the DHAM reweighted MSM, and the "DHAM + path correction" reweighted MSM. (**C**): Heatmaps showing the distance factor D between the reweighted MSMs with DHAM (top) and "DHAM + path correction" (bottom) and the reference as a function of the number of clusters and the lagtime (red: good agreement, blue: no agreement). Parameter sets used for **A** and **B** are highlighted with a black and grey box, respectively.

TABLE 5.1: MSM relaxation timescales of the three slowest processes (the crossings of the three distinct energy barriers in the four-well system) of the reference MSM, the MSM from the biased simulations, and the reweighted MSMs with the Weber-Pande, Girsanov, and DHAM algorithms.

Relaxation process	Reference timescale [steps]	Biased timescale [steps]	Weber-Pande timescale [steps]	Girsanov timescale [steps]	DHAM timescale [steps]
1	$2.90 \cdot 10^7$	$7.78 \cdot 10^4$	$2.03 \cdot 10^7$	$2.01 \cdot 10^7$	$1.97 \cdot 10^7$
2	$1.17 \cdot 10^5$	$1.75 \cdot 10^4$	$1.12 \cdot 10^5$	$1.15 \cdot 10^5$	$1.13 \cdot 10^5$
3	$3.82 \cdot 10^4$	$1.07 \cdot 10^4$	$4.22 \cdot 10^4$	$4.29 \cdot 10^4$	$3.96 \cdot 10^4$

tion, because it adds a correction to DHAM based on the specific path chosen. We compared the reweighting results from DHAM and "DHAM + path correction" for short and long lagtimes. For short lagtime, DHAM results match the reference both in terms of the potential energy and the kinetics, and thus the application of the path correction does not change these results (Figure 5.5A). This is expected as the path correction (i.e. a sum over the transition path) does not contribute much to the reweighting for short lagtimes.

As already discussed in Section 5.7.2, DHAM becomes inaccurate for longer lagtimes. Figure 5.5B depicts the reweighting results for a lagtime of 4'000 steps. The energy surface is underestimated and the kinetics differ by a factor of 300 from the reference. For long lagtimes, the path correction contributes significantly, and thus the reweighted results with "DHAM + path correction" improve (Figure 5.5B). The reweighted energy surface is as close to the reference one as with Weber-Pande or Girsanov for the same lagtime. The kinetics deviate from the reference only by a factor of 3. However, the kinetic results have a larger deviation from the reference than with Weber-Pande or Girsanov. We speculate that this is because we assumed an overdamped

(high friction) system in the derivation of the path correction. In the test system, the assumption $|\gamma m \frac{dr}{dt}| >> |m \frac{d^2r}{dt^2}|$ is not true for all simulation steps. Therefore, the path correction does not perfectly resemble the Weber-Pande results. It remains to be tested how well the path correction performs for more complex, biological systems. This will be the subject of future work.

In Figure 5.5C, the performance of DHAM and "DHAM + path correction" is compared for a wide range of MSM input parameters. In contrast to the other three algorithms, "DHAM + path correction" is applicable to a wide range of parameters and is not limited to either short or long lagtimes. Additionally, note that the application of the path correction does never lead to a decrease in the performance, even at small lagtimes.

In their original publication, Rosta *et al.*[46] proposed a different reweighting term for long lagtimes,

$$P_{ij}^{unnorm} = \tilde{P}_{ij} \exp\left(-\frac{U_j}{k_B T}\right). \tag{5.33}$$

In contrast to Eq. (5.5), the reweighting with Eq. 5.33 is independent of the bias energy of the start state U_i. The authors argue that with increasing lagtime, the memory of the start state vanishes. To compare this approach with the path correction developed in the present work, we reweighted the test system with Eq. (5.33), termed "DHAM for long lagtimes" in the following (Figure 5.6). This variant of DHAM gave more accurate reweighting results for lagtimes in the range of 400-1'000 steps. However, the results for short lagtimes (<400) showed significantly reduced accuracy. Furthermore, we observed a high parameter sensitivity for "DHAM for long lagtimes", and even for the best parameter combinations the kinetics results deviated around one order of magnitude from the reference. Therefore, we conclude that

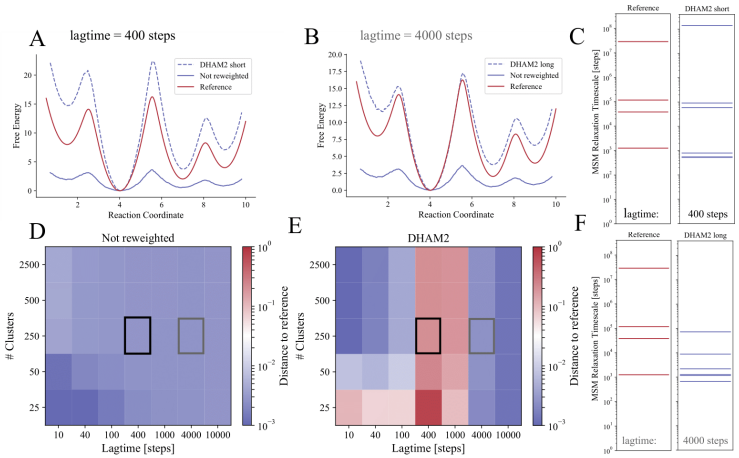


FIGURE 5.6: Reweighting performance at a lagtime of 400 steps (**A** and **C**) and 4'000 steps (**B** and **F**). **A** and **B** Free-energy landscape reweighted by "DHAM for long lagtimes" (blue, dotted) in comparison with the reference (red) and the biased landscape (blue) at lagtimes of 400 and 4'000 steps, respectively. **C** Kinetic spectra of the reference MSM and the MSM reweighted with "DHAM for long lagtimes" with a lagtime of 400 steps. (**D**): Heatmap showing the distance factor D between the biased MSMs and the reference as a function of the number of clusters and the lagtime (red: good agreement, blue: no agreement). Parameter sets used for **A**/**C** and **B**/**F** are highlighted with a black and grey box, respectively. (**E**): Same as **D** but for the results with "DHAM for long lagtimes". (**F**): Kinetic spectra of the reference MSM and the MSM reweighted with "DHAM for long lagtimes" with a lagtime of 4'000 steps.

"DHAM + path correction" is more accurate as well as less parameter sensitive than "DHAM for long lagtimes".

In summary, we showed that the application of the path correction is able to extend DHAM to longer lagtimes, with improved parameter stability compared to the other tested reweighting methods. With this approach, it was possible to recover the same potential-energy landscape as with the path-based reweighting algorithms. For long lagtimes, the kinetics were still a factor of two less accurate than Weber-Pande and Girsanov reweighting, but a factor of 100 more accurate than DHAM. Future work will test the performance and stability of "DHAM + path correction" for more complex systems to obtain more general conclusions.

5.8 CONCLUSION AND OUTLOOK

In this chapter, we presented the connection between path-based reweighting and energy-based reweighting methods and shown a strong similarity among the path-based reweighting algorithms. We could show that Weber-Pande and Girsanov reweighting share the same path-reweighting term (Eqs. (5.11), (5.25)), but differ in the phase-space reweighting term g, which is only present in the Girsanov approach. The inclusion of g leads to improved performances for MSMs with of a low number of clusters. In contrast, g introduces an imbalance for time-dependent biases, and thus decreases the performance in these cases. It therefore depends on the system, the MSM input parameters, and the biasing potential if the inclusion of g is beneficial. Despite the small difference due to the phase-space reweighting factor, the two path-based reweighting approaches yield very similar results for our simple

test system, especially with large numbers of clusters.

In addition, we tested the dependency of the different reweighting methods on the MSM input parameters. Both path-based reweighting algorithms performed well at long lagtimes but poorly at short lagtimes. Further research is needed to clarify if the need for long lagtimes is due to (implicit) physical assumptions in the derivation or due to numerical reasons like precision problems.

In the second part of this chapter, we showed that the path-based reweighting methods can be linked to the DHAM approach. The only assumptions required to connect the approaches were a high-friction system and time-independent biases. Especially the latter assumption illustrates that path-based reweighting can be applied to more different types of biases. However path-based reweighting is currently limited to SD, whereas energy-based reweighting can be applied to both SD and MD. Based on our derivation, we found that path-based methods contain the path-independent DHAM equation plus a path-dependent term, which we termed "path correction". While the original DHAM showed good performance for short lagtimes and poor performance for long lagtimes, the application of the path correction led to good performances over the full range of lagtimes. "DHAM + path correction" combines the strength of high numerical stability at short lagtimes and the consideration of path dependency for long lagtimes. This will be especially important for more complex systems and slow dynamics, where it is inevitable to use long lagtimes.

From a technical point of view, "DHAM + path correction" has a number of key advantages. First of all, the formulation does not depend on the random number, in contrast to the path-based reweighting methods Weber-Pande and Girsanov. The need to know the random number at every time step is currently one of the main limitations of path-based reweighting. Thus, the path

correction can be applied with any SD integrator. Secondly, the path correction depends on the force of every time step, which may be correlated between time steps (in contrast to the random number). Therefore, it may be possible to increase the write-out frequency of the forces or even use the path correction for post processing. Thirdly, the "DHAM + path correction" approach is in principle no longer restricted to SD like the other path-based methods but can be applied to MD as well. However, a good approximation for γ will be necessary in this case. The potential of the path correction approach and its applicability to more complex systems will be explored in future work.

This chapter led to new insights into the existing dynamic reweighting algorithms and for the first time connected path-based and energy-based approaches. In addition, a combined method that joins the advantages of path-based and energy-based reweighting was proposed. Nevertheless, many open questions about dynamic reweighting remain: In the presented chapter, we focused on the theoretical aspects and the mathematical derivation of the connection between path- and energy-based reweighting methods. We chose a simple, easy to understand 1D model system to test our theoretical derivations and to highlight important results. Future work is needed to validate the presented findings on more complex atomistic systems of practical relevance such as drug-like molecules or proteins. Additionally, in this chapter we focused on DHAM as a representative of the energy-based reweighting methods. As mentioned in the Introduction, the energy-based reweighting class includes more members, e.g. DHAMed[7] and TRAM.[6] The performance of these methods was not explicitly tested in our study. It will be of special interest to test if these methods do overcome the lagtime dependence of DHAM.

Future work will therefore need to systematically compare the performance of all available dynamic reweighting algorithms on both model systems and molecular systems. In addition to testing different reweighting algorithms, it will be important to include a variety of enhanced sampling techniques and bias strengths to determine the best balance between speed-up and reweighting performance. The main difficulty is thereby that the different reweighting methods are not implemented in a single MD code and were not tested on the same molecular systems.

6 Revisiting the Kinetics of Cyclosporin A

"Proceed, With Fingers Crossed"

TeX

This quote and chapter are dedicated to Alzbeta Kubincova and her lessons on resilience and bravery

Cyclic peptide based drugs can be increasingly found in pharmaceutical drug pipelines. Their size and large number of potential hydrogen bond donors and acceptors allow them to bind to targets complementary to conventional small molecule drugs. However, out of the same reasons, they often suffer from poor membrane permeability. The natural product cyclosporin A (CycA) is known for its relatively high passive permeability given its size that originates from a complex conformational behavior. Thus, CycA has been considered a model system to study permeable cyclic peptides. We revisit the conformational dynamics of CycA and explore whether dynamic reweighting techniques can be helpful to facilitate sampling. By improving and correcting previous kinetic models for CycA, we gain new insights into its permeable conformational states, but observe also challenges and pitfalls associated with kinetic models. Our results show that a vast amount of simulations are necessary to sample the rare cis/trans isomerization in the peptide backbone of CycA. As an alternative to this massive sampling, we show how enhanced sampling and dynamic reweighting can be applied to this system. We compare the reweighted equilibrium and kinetic information as well as the robustness of the kinetic models to our massive sampling reference. In addition, we give guidelines on best practices and highlight obstacles for enhanced sampling and dynamic reweighting.

6.1 INTRODUCTION

Cyclic peptides belong to the beyond rule-of-5 (bRO5) class of compounds and expand the chemical space beyond traditional small molecule drugs.[119–121,158–161,183,233] With their relatively large size, polar surface area, hydrogen bonding potential, and often inherent flexibility they offer unique binding qualities and are able to bind currently undruggable targets.[234] Specifically, cyclic peptide drugs hold the potential to bind to large and flat bindings sites, for example on the surface of a protein or at the interface of a protein-protein interaction.[61,118,123,124,163,235–237] However, cyclic peptides come with their own unique challenges related to solubility and permeability. As permeability decreases exponentially with increasing molecular volume,[70] the relatively large cyclic peptides often suffer from poor membrane permeability.[69,70,94,97] To leverage the clinical potential of cyclic peptides, the drug development community is striving to understand their structure-permeability relationship.[61,70,114,126,127,176,233] Efforts from numerous groups have revealed the effects on permeability from backbone N-methylations,[84,85,128,177,178] side-chain modifications,[108,182] stereocenters,[106,179,180] and amphiphilicity.[181] However, these effects are highly non-linear and context specific.[98,99,106,108,115,129–132,1] Thus, we have not yet reached a level of understanding where cyclic peptide permeability can be considered predictable.

The natural product cyclosporin A (CycA) ranks among the best studied cyclic peptides. It is well known for its oral availability despite its relatively high molecular weight of 1202.6 $\mathrm{g\,mol^{-1}}$, and is used as an immunosuppressive drug in transplantation medicine since 1983.[13,76–78,238] The discovery of the surprisingly high permeability of CycA triggered a cascade of research aim-

ing to understand the origin of its permeability leap.[14,15,79–82] A distinct feature of CycA caught the attention of the community early on: CycA is a molecular chameleon and can adapt to different environments.[12,100–103] The chameleonic property of CycA originates from its conformational flexibility that allows the macrocycle to adopt different conformational states with different lipophilicity depending on the environment.[12,16,95,102–104,185] In one conformation, which is referred to as the 'closed' state, CycA can form three transannular hydrogen bonds and consequently shield its polar surface area from the surroundings. The conformer thus appears apolar from the outside and is favored in an apolor environment. In contrast, in so-called 'open' conformations, the polar atoms are exposed. These conformers appear more polar from the outside and can form favorable interactions with a polar environment.[187,239] The chameleonic property of CycA allows the molecule to adapt to both polar and apolar environment, thus leading to increased solubility and permeability.

Computational and experimental approaches have worked hand in hand to increase our understanding of the origin of the chameleonic property of CycA and to extract design principles for new macrocycles with equal permeability. Passive membrane permeability is measured either using cell-based assays (Caco-2 or MDCK)[240,241] that resemble the *in vivo* membrane compositions or using an parallel artificial membrane permeability assay (PAMPA).[188] However, both approaches do not permit to gain atomistic insights into the permeability process or monitor the conformational changes during permeation. Addressing this gap, computer simulations using molecular dynamics (MD) can investigate the conformational and dynamic behavior of cyclic peptides in different polar and apolar environments (e.g., chloroform and water) or in pres-

ence of an explicit membrane.[12,89,96,104,108,114,115,189,242] In addition, machine learning approaches have been explored to predict the structure of permeable macrocycles.[186] Especially Markov state models (MSMs) – kinetic models based on MD simulations – have proven useful to characterize the dynamic properties of macrocycles.[16,89,96,106,108,115] MSMs provide the metastable conformational states, their equilibrium populations, and the interconversion kinetics between them.[22,152–154] Thus, they are a powerful tool to gather insights into the conformations relevant for permeability. However, constructing MSM models has some pitfalls and thus, a high validity model often requires expert knowledge and proper manual user input, e.g. relevant input features, appropriate clustering of the states, and a Markovian lagtime.

Previously, our group reported a kinetic model of CycA that explored its conformational space in water and chloroform and revealed that similar conformational states exist in both environments (termed congruent states).[16] The slowest interconversion process between any conformer was reported to be around 100 ns, however, as noted in Ref. 16 transition rates obtained from MSMs are only rough estimates of the real time scales due to discretization errors (e.g., from clustering) or force-field inaccuracies. This is especially true if sampling is not fully converged. Revisiting that data set, we discovered a slow cis/trans isomerization that was highly undersampled – only one transition was observed in the whole 10 µs data set. Thus, the intrinsic assumptions of MSMs were violated. However, the conventional MSM quality checks were passed, suggesting that the system was sufficiently sampled.

In this work, we aim to improve the kinetic model of CycA and to construct guidelines to better control the quality and robustness of MSMs. In order to achieve sufficient sampling of the

rare cis/trans isomerization, the overall simulation time had to be increased by one order of magnitude. As such massive sampling is not always feasible, we explored whether enhanced sampling simulations together with dynamic reweighting methods[6,46] can be used as an alternative. This work offers an improved kinetic model for CycA, new insights into a transient 'closed' state with all trans amide bonds, and introduces exemplary workflows for rare-event sampling. It also highlights the challenges and best practices associated with MSMs and dynamic reweighting.

6.2 COMPUTATIONAL DETAILS

6.2.1 Starting Conformations

A set of 100 diverse conformations were obtained from Ref. 16 as starting points for the MD simulations. The protocol to generate these conformations is described in detail in Ref. 16. In short, crystal structures of CycA in the 'closed' and in one 'open' state were obtained from the Cambridge Structural Database (CSD) and the Protein Data Bank (PDB), respectively.[14,243] Simulations at 400 K in water and chloroform, partially with modified potential-energy terms, were performed to enhance conformational sampling. The different enhanced sampling conditions were (a) no backbone dihedral angle terms, (b) repulsive potential on atoms forming H-bonds, (c) no charges on atoms forming H-bonds, (d) reduced charges (q = 0.1 e) on atoms forming H-bonds, or (e) no potential-energy modifications. The obtained enhanced sampling simulations were reweighted to physical conditions and only the low energy conformations were considered fur-

ther. Together with conformations from unbiased simulations at 300 K, the conformations were clustered based on the root-mean-square deviations (RMSD) of the backbone positions relative to the 'closed' or 'open' crystal structures. The centroids of these clusters were selected in the pool of 100 diverse conformations.

6.2.2 MD Simulations

Starting from the set of 100 diverse conformers, MD simulations were carried out using the Groningen Machine for Chemical Simulations (GROMACS) 2020.5 software package.[143] For compatibility with the results from Ref. 16, the GROMOS 54A7 force field[244] was employed. A time step of 2 fs together with the leapfrog integration scheme[32] was used, and periodic boundary conditions were applied. The peptide and the solvent were coupled to two different thermostats at 300 K using a weak coupling scheme[37] and a relaxation time of 0.1 ps. Pressure of the system was controlled using an isotropic Parrinello-Rahman barostat[145] at 1.0 bar with a coupling constant of 2.0 ps and a compressibility of $4.5 \times 10^5\,\mathrm{bar}^{-1}$. Long-range electrostatic interactions were treated using the particle mesh Ewald algorithm.[146] The linear constraint solver (LINCS)[147] algorithm was utilized to constrain the bond lengths.

In total, five separate simulations with different random seeds for the velocity generation and a 200 ns length each were started from each of the conformers. This resulted in a total simulation time of 100 µs. To increase the sampling of the 'cis state' of CycA (i.e., a cis amide bond between residues 9 and 10 as found in the small-molecule crystal structure), additional ten 200 ns simulations were performed from five different cis starting conformers. This provided an additional simulation time of 10 µs.

To explore the potential of dynamic reweighting methods, we performed enhanced sampling simulation that were subsequently reweighted to the physical conditions. To increase the sampling of the cis/trans isomerization of the peptide bond in CycA, the dihedral energetic contribution of the ω dihedral between residue 9 and 10 was modified. The different modification modes are summarized in Table 6.1.

Table 6.1: Simulation Setups for the Enhanced-Sampling MD Simulations

Condition	ω 9-10 dihedral term [kJ/mol]
Unbiased	$33.5 \cdot (1 + \cos(2\phi - 180))$
1	$30 \cdot (1 + \cos(2\phi - 180))$
2	$27 \cdot (1 + \cos(2\phi - 180))$
3	$23 \cdot (1 + \cos(2\phi - 180))$
4	$19 \cdot (1 + \cos(2\phi - 180))$
5	$16.25 \cdot (1 + \cos(2\phi - 180))$
6	$12 \cdot (1 + \cos(2\phi - 180))$
7	$8 \cdot (1 + \cos(2\phi - 180))$
8	$5.86 \cdot (1 + \cos(2\phi - 180))$

For conditions 2 and 5, eleven simulation with 50 ns each were performed for each of the 100 conformers, giving a total biased simulation time of 55 µs. For all other biased conditions, two simulations with 50 ns each were performed for each of the 100 conformers (i.e., total biased simulation time of 10 µs).

6.2.3 Markov State Model and Analysis

Torsional angles analysis on the raw trajectories was performed using MDTraj.[149] In order to extract the kinetic and equilibrium properties of the system, a MSM[152–155] was constructed to determine the long-time properties from multiple short – and parallel

– simulations.[22] For this, we extracted the ϕ, ψ, and ω backbone torsional angles of CycA for each simulation using the PyEmma package.[148] This resulted in 66 input features for the model. The dimensionality of the input features was reduced to six dimensions with time-lagged independent component analysis (TICA) dimensionality reduction.[150] Subsequently, the feature vectors were discretized. To assess the influence of the clustering method on the MSM, two different algorithms (hierarchical density-based Sittel-Stock clustering[151] and the PyEmma intrinsic k-means clustering) were compared. For the k-means clustering, the number of clusters was set to 100. For Sittel-Stock clustering, the minimum number of cluster members was set to 50.

Based on the discretized trajectories, the transition matrix of the MSM was constructed using a lagtime of 2.5 ns. The kinetic spectrum as well as equilibrium energetics and cluster populations were calculated. To kinetically group microstates into metastable sets, we performed robust Perron cluster-cluster analysis (PCCA+)[156] using five or six metastable states. The resulting states were visualized using PyMol[245] and the mean first passage times (MFPTs) between the metastable sets was computed.

6.2.4 Dynamic Reweighting

Dynamic reweighting was performed using the DHAM[46] algorithm in the short-time and long-time approximation (own implementation) or the TRAM[6] algorithm (provided in the PyEmma[148] package). For each frame of the trajectory, the difference between the unbiased and corresponding biased dihedral term of Table 6.1 was calculated. The DHAM-reweighted transition matrix was used as an input for the PyEmma package to extract equilibrium distributions and kinetics. The TRAM algorithm directly returns

the equilibrium distribution.

6.3 RESULTS AND DISCUSSION

Previous kinetic models revealed that the cis/trans isomerization of the peptide bond between residues 9 and 10 (ω_{9-10}, see also Figure 6.1F) is the kinetic bottleneck to sample the full conformational space of CycA.[16] To ensure that the rare cis/trans isomerization was sufficiently sampled in our data set, we plotted the ω_{9-10} distribution for all simulation trajectories (Figure 6.1A), and tracked the time spent in each state as well as the transitions between the states. In total, 31 transitions from cis to trans as well as five transitions from trans to cis were observed. Although the amount of trans to cis transitions is on the lower end, it should still be sufficient to obtain a reasonable MSM. All backbone dihedral angles (ϕ, ψ, ω) were used as input features and reduced using the time-lagged independent component analysis (TICA) dimension reduction approach.[150] Figure 6.1B shows the data distribution mapped to the two slowest transitions, roughly separated into four distinct clusters. To assess how the slowest transition is linked to the 'open' and 'closed' states of CycA, we projected the data to the first TICA coordinate and the RMSD to the 'closed' crystal structure (Figure 6.1C). Interestingly, 'closed' states (i.e., RMSD $\sim$ 0.1 nm, marked with asterisks) were observed at both extremes of the first TICA coordinate. To get an intuition which dihedral transitions contributed to the slowest dynamics of CycA, we looked at the correlations between the input features and the TICA coordinates. Not surprising, this analysis revealed that the slowest interconversion process of CycA is

dominated by the ω_{9-10} dihedral (time correlation = 0.995, see also Figure 6.1E), which defines the cis and trans states of CycA. This correlation can also be seen in Figure 6.1D when the ω_{9-10} dihedral angle is plotted against the first TICA coordinate. The second slowest process is dominated by the ϕ backbone dihedral angle of residue 6 (time correlation = 0.962) and the third slowest process by the ϕ backbone dihedral angle of residue 4 (time correlation = -0.630). The correlations between the three slowest processes and all input features are shown in Figure 6.1E. The three most dominating dihedrals are highlighted in Figure 6.1F. Taking together our observations that the cis/trans isomerization defines the first TICA coordinate and that 'closed' states were observed at both extremes of this coordinate, this suggests that 'closed' conformations exist with the ω_{9-10} peptide bond in both cis and trans configuration. We verified these observations at the level of the raw data. Figure 6.2 shows trajectories visiting the cis-closed as well as the trans-closed state. In addition, examples of cis to trans and trans to cis transitions are shown. The trans-closed state showed frequent short opening events. In contrast, the cis-closed state was very stable over the length of the simulation.

The MSM of CycA using the density-based Sittel-Stock clustering algorithm is shown in Figure 6.3. Panel 6.3A shows the implied timescales plot of a Bayesian MSM. The slowest process of the system relaxes with a timescale of 2.2 µs. The kinetic spectrum contains two spectral gaps (between the first and second as well as between the sixth and seventh slowest process, Figure 6.3B). The equilibrium population predicted by the MSM showed that the microstate corresponding to the cis-closed conformation makes up 1.6±0.3% of the population (errors denote the standard deviation based on the statistical error of the Bayesian MSM).

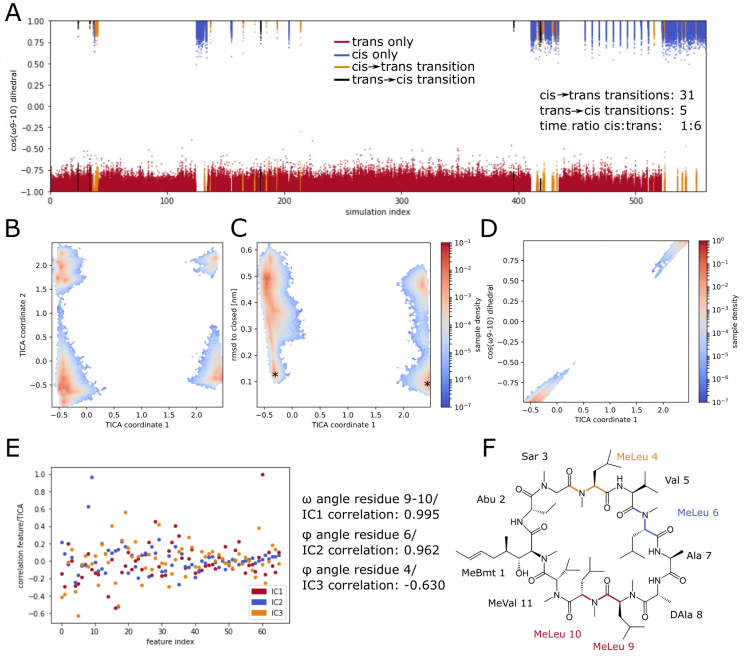


FIGURE 6.1: Description of the unbiased data set of CycA. (**A**): ω_{9-10} dihedral distribution for the individual trajectories plotted against the corresponding simulation index. Data points are colored based on the visited state of the ω_{9-10} dihedral angle: only trans (red), only cis (blue), only cis to trans transitions (orange), and trans to cis transitions (and potentially an additional cis to trans transition) (black). (**B**): Phase space of cyclosporin A porjected to the two slowest TICA coordinates. (**C**): Phase space of cyclosporin A projected to the slowest TICA coordinate and the RMSD to the cis-closed state. (**D**): Slowest TICA coordinate plotted against the cosine of ω_{9-10}. (**E**): Correlation between the input features and the three slowest TICA coordinates. The highest correlating features for each TICA coordinate are highlighted in **F**.

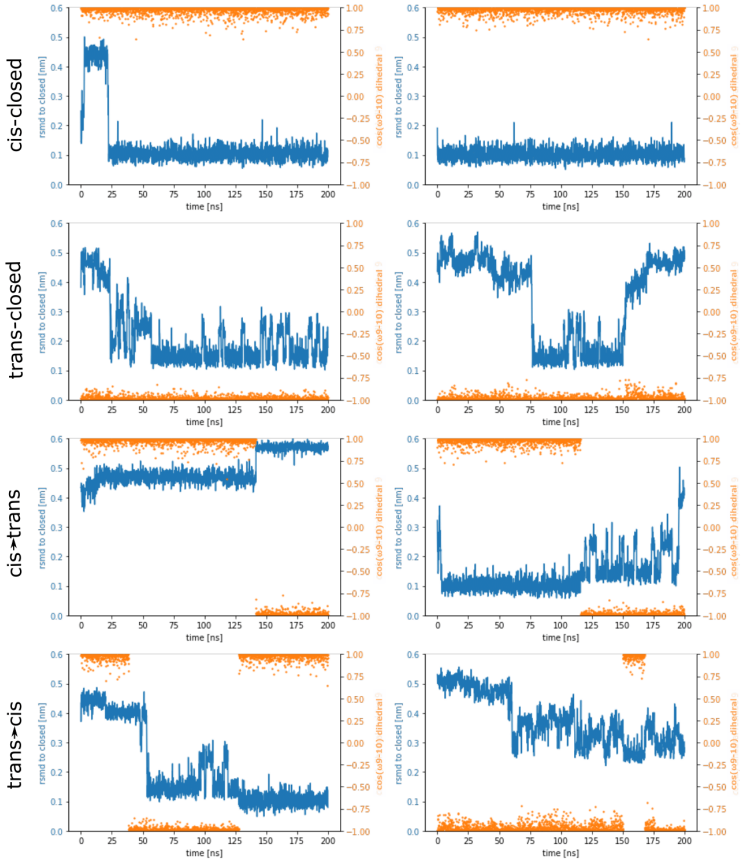


FIGURE 6.2: Example trajectories for cis-closed and trans-closed states as well as transitions between cis and trans of the peptide bond between residues 9 and 10. The RMSD to the cis-closed state is shown with blue lines, the cosine of the ω_{9-10} dihedral angle is shown in orange.

In comparison, the trans-closed microstate has a population of 3.0±0.2%. Thus, the trans-closed state has a two-fold higher population. Figure 6.3C plots the equilibrium population of all microstates as a function of the RMSD to the 'closed' crystal structure. The inlay shows the trans-closed conformation. With the exception of the ω_{9-10} dihedral angle, the trans-close state is identical to the cis-closed state and also contains three intramolecular hydrogen bonds. The estimated free-energy surface of the conformational states of CycA is shown in panel 6.3D. The two clusters in the cis state (i.e., on the positive extreme of the first TICA coordinate) only contain one minimum each. In contrast, the two clusters in the trans state contain multiple minima.

Based on the MSM, we split the conformational landscape of CycA into six metastable states (Figure 6.4). While one of the metastable states is the cis-closed state, none of the metastable states resembles the trans-closed conformation. This is in line with our observations from the raw data that the trans-closed state is unstable. However, it has a higher equilibrium population than the cis-trans state. This raises the question of how this state contributes to the permeability of CycA. On the one hand, the life time of the trans-closed state might be too short to be relevant for the permeability process. On the other hand, the relatively low energy barrier for interconversion to 'open' conformations may be important for resolvation of the peptide after passing through the apolar membrane part. Further analysis of the trans-closed state will be necessary to answer these questions. However, we find it intriguing that CycA is not only conformationally flexible but also 'kinetically flexible', i.e., it exhibits both a slow interconversion between open and closed conformations (in the cis isomer) as well as a fast interconversion between open and closed (in the trans isomer). The MFPT between the cis-closed state and the highest

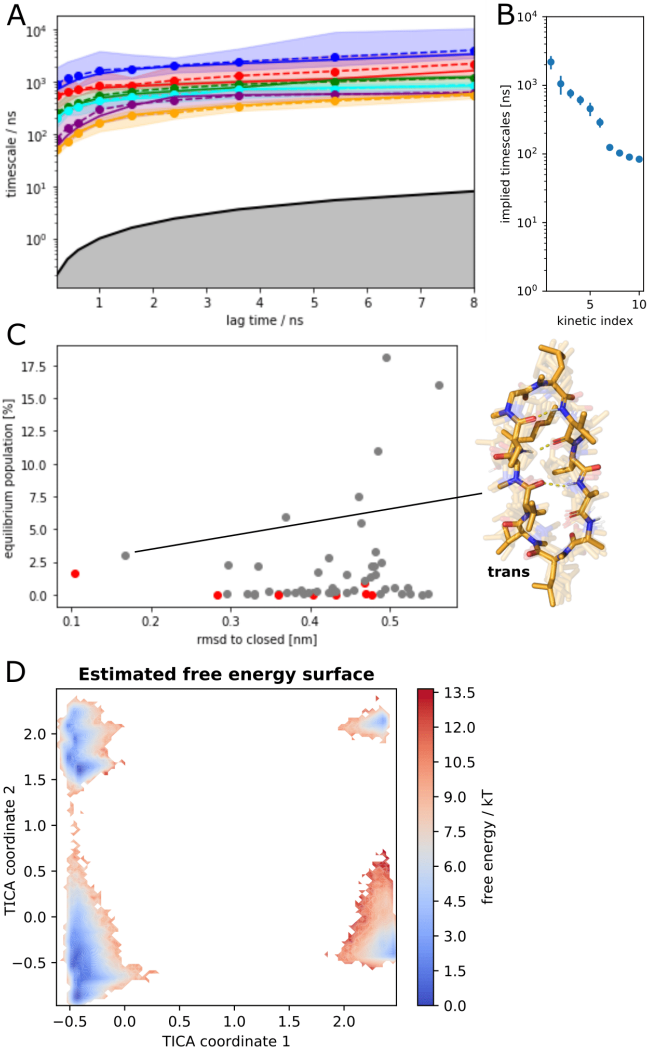


FIGURE 6.3: MSM based on the Sittel-Stock clustered data. (**A**): Implied timescales as a function of the lagtime. (**B**): Kinetic spectrum of the system. (**C**): Equilibrium population of the microstates as a function of the average RMSD to the cis-closed conformation. Clusters in the cis state are colored in red, clusters in the trans state are colored in grey. The structure of the trans-closed state is highlighted in the inlay. (**D**): Estimated free-energy profile of the MSM.

populated open state was found to be 1.1 µs for the cis to trans transition and 44.2 µs for the trans to cis transition.

The MSM based on the k-means clustering is shown in Figures 6.5 and 6.6. The kinetics agree reasonably well with the MSM based on the Sittel-Stock clustering (slowest timescale 1.5 µs versus 2.2 µs). Interestingly, the k-means clustering breaks the 'closed' states up into multiple sub-states. Counting together all cis-closed and trans-closed sub-states, the population of the trans-closed state exceeds again the cis-closed population (3.7±0.1% versus 2.0±0.1%, errors denote the standard deviation based on the statistical error of the Bayesian MSM). In addition, the fact that only the cis-closed state and not the trans-closed state is metastable is in line with the results discussed above.

To assess the robustness of the MSMs, we chose a variational approach, i.e. the input parameters for the MSM were varied and the influence on the model output was tracked. A well converged model with sufficient sampling should result in stable predictions,[218] whereas insufficiently sampled data results in model predictions that vary heavily with the input parameters. For this, we compare the kinetics of the slowest process of the model because this often shows the highest variation between MSMs. Figure 6.7 shows the robustness analysis for MSMs based on the density-based Sittel-Stock clustering and k-means clustering. To learn whether the input data was sufficiently sampled, MSMs were built on a random subset of the total data. The fraction of data used to build the model with respect to the total 110 µs simulation time is shown in the y-axis of the heatmap in Figure 6.7. In addition, the input parameter for the clustering was varied. The Sittel-Stock clustering did not converge for subdatasets with a fraction ≤ 0.3. Thus, no data is shown in that regime. In general, the variational approach revealed high discrepancies be-

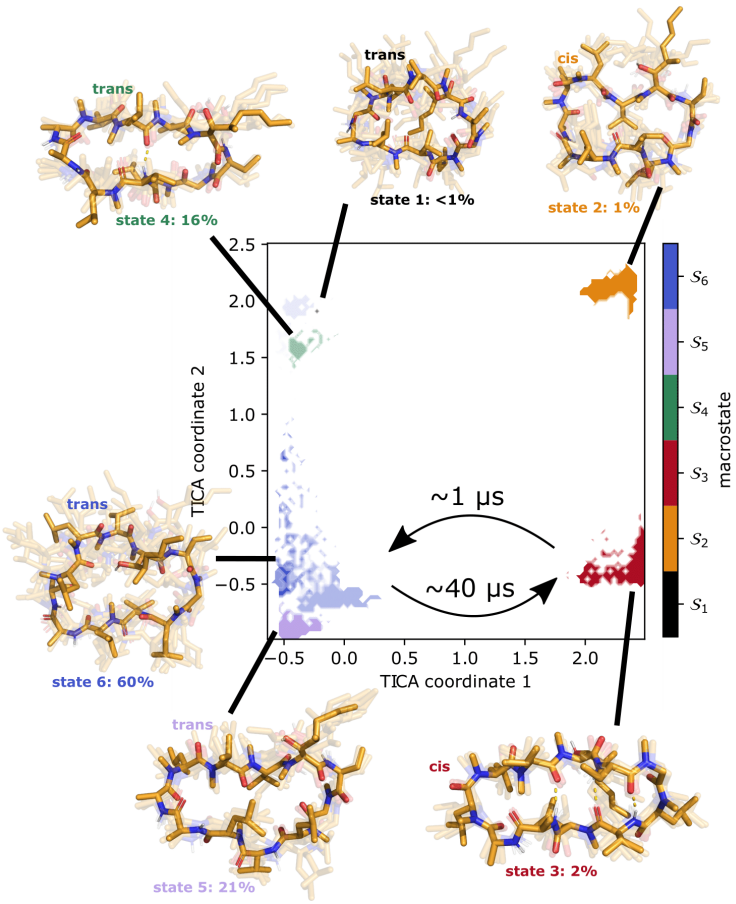


FIGURE 6.4: Metastable state assignment of the Sittel-Stock clustered MSM. The distribution of the six metastable sets and their position in the TICA phase space are plotted. For each metastable set, the conformational ensemble is provided. MFPTs between the cis-closed state and the highest populated open state are indicated.

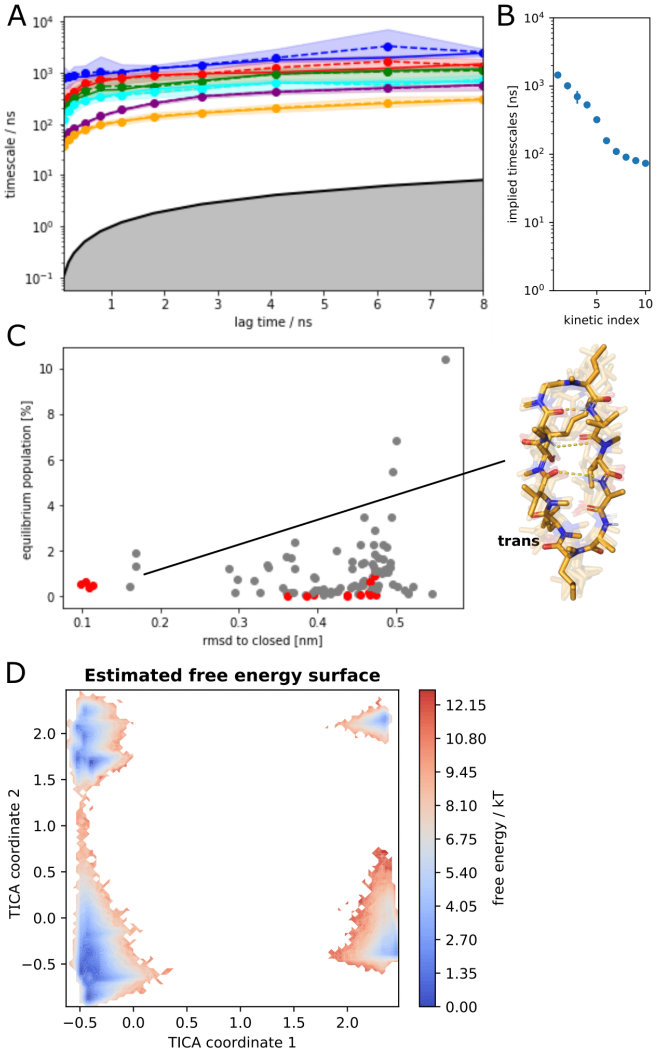


FIGURE 6.5: MSM based on the k-means clustered data. (**A**): Implied timescales as a function of the lagtime. (**B**) Kinetic spectrum of the system. (**C**): Equilibrium population of the microstates as a function of the average RMSD to the cis-closed conformation. Clusters in the cis state are colored in red, clusters in the trans state are colored in grey. The structure of the trans-closed state is highlighted in the inlay. (**D**): Estimated free-energy profile of the MSM.

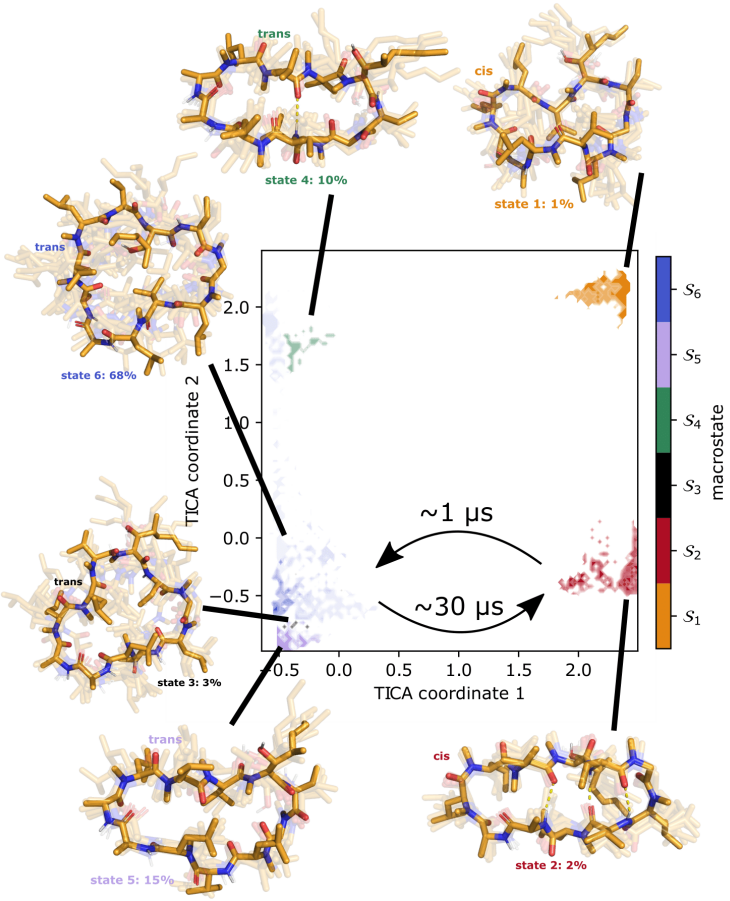


FIGURE 6.6: Metastable state assignment of the k-means clustered MSM. The distribution of the six metastable sets and their position in the TICA phase space are plotted. For each metastable set, the conformational ensemble is provided. MFPTs between the cis-closed state and the highest populated open state are indicated.

tween the predicted slowest kinetics with differences of multiple orders of magnitudes. Although the models based on Sittel-Stock clustering seemed slightly more robust, both MSMs cannot be considered fully converged. Thus, more sampling is necessary to improve the MSMs and their robustness. This is concerning as the model is already build on ten fold more simulation time than previously published models (110 µs versus 10 µs of simulation time).[16] As such extensive sampling is not feasible for screening and comparison of multiple cyclic peptides, we decided to evaluate alternatives. In specific, we tested whether the combination of enhanced sampling techniques and dynamic reweighting of the MSM can be used to recover the kinetic and equilibrium information about CycA.

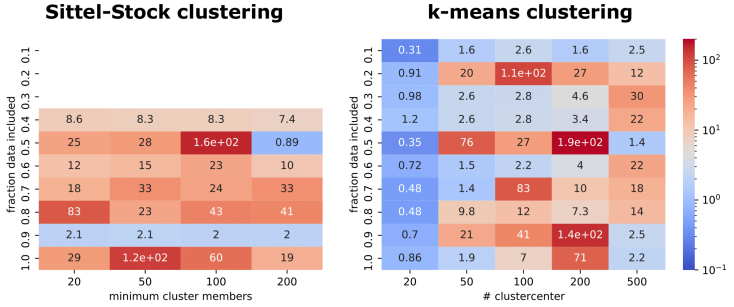


FIGURE 6.7: Robustness analysis of the MSMs based on Sittel-Stock clustering (left) and k-means clustering (right) by varying the clustering input parameter (x-axis) and the fraction of included simulation trajectories (y-axis). The slowest relaxation timescale for every combination is shown as a heatmap. Units are displayed in µs.

The Holy Grail for enhanced sampling is to significantly increase the frequency of desired transitions without disturbing the remaining system. The ω dihedral angles are characterized by a high force constant of 33.5 kJ mol^{-1} resulting in high transi-

tion barriers. An apparently straightforward way to increase the cis/trans transition frequency is to reduce the dihedral force constant of the peptide bond of interest. Similar approaches have been applied to cyclic peptides and proteins before.[80,189,246] To assess the effects of reducing the dihedral force constant, we carried out a total of 5 µs simulations for different scaling factors. Figure 6.8 shows the distribution of the ω_{9-10} dihedral angle for the different force constants. Gaussian functions were fitted to the distributions to obtain the peak positions of the distributions. In the unbiased simulations, two distinct peaks at 0 rad (cis) and $-\pi$ rad (trans) can be observed. Lowering the transition barrier leads first to a slight shift of the cis peak. Below a force constant of 23 kJ mol^{-1}, both the cis and trans peaks shift substantially and a third peak (additional state) appears. Using a 50% reduced force constant (= 16.75 kJ mol^{-1}) or smaller, the previous energy minima at 0 rad and $-\pi$ rad become increasingly unfavourable. The dramatic shift of the energy minima of the dihedral angle can be explained by the treatment of the 1-4 nonbonded interactions of the force field. These interactions act in combination with the dihedral-angle term, resulting in the desired effective torsional profile. However, this means that scaling the dihedral force constant without scaling the corresponding 1-4 nonbonded interactions disturbs the balance between the terms, causing the shifts of the energy minima. Alternative biasing approaches should therefore be explored in the future, e.g., by addition of a phase-shifted dihedral-angle term or by an additional scaling of the 1-4 nonbonded interactions.

Figure 6.8 highlights the importance of properly monitoring the bias applied by enhanced sampling. Choosing a too small force constant leads to an insufficient overlap between the original and the biased phase space. In order to have sufficient overlap,

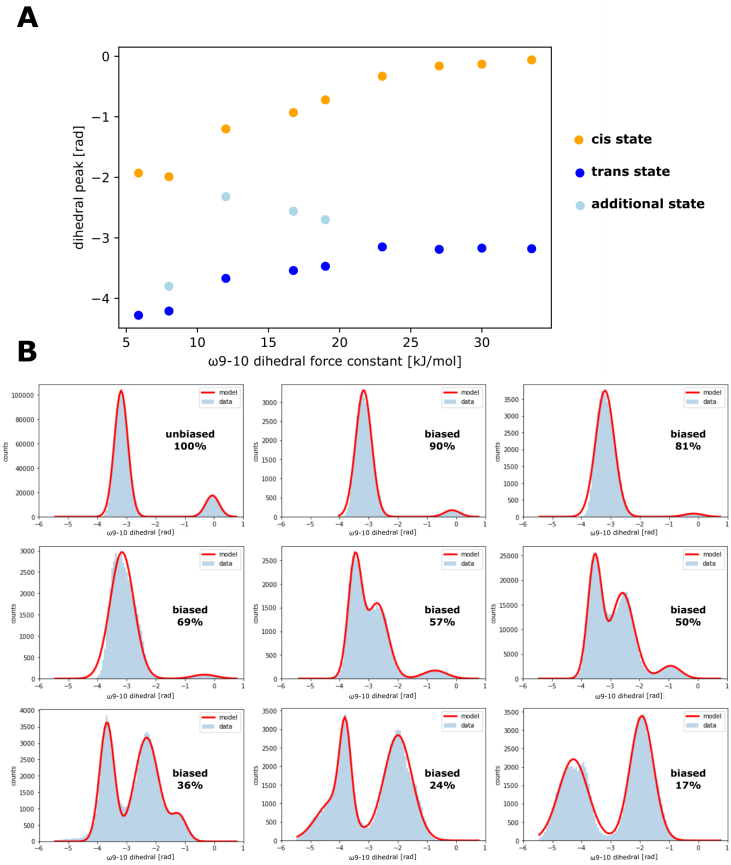


FIGURE 6.8: Effect of varying the force constant of the ω_{9-10} dihedral-angle term. (**A**): Peak positions of the ω_{9-10} dihedral-angle distribution as a function of the force constant. (**B**): Distributions of the ω_{9-10} dihedral angle for different values of the force constant.

the biased force constant was set to $27\,\mathrm{kJ\,mol^{-1}}$ in this work. The resulting data set will be called 'moderately biased' below. In addition, we wanted to test the effect of insufficient overlap on the reweighting approach and also performed biased simulations with a force constant of $16.75\,\mathrm{kJ\,mol^{-1}}$. This data set will be referred to as 'highly biased' below. For both conditions, a total simulations time 50 µs biased simulations was collected.

Figure 6.9 summarizes the 'moderately biased' data set. The data was projected to the TICA coordinates from the unbiased reference to facilitate comparison (Figure 6.9A). Interestingly, the cluster assigned to the cis-open states (top-right corner) is not visible in the biased ensemble. Investigation of the raw data revealed that the cis-open states are sampled in some simulations, however, they were highly unstable and directly relaxed into a trans-open state. Thus, although these states are metastable in the unbiased data set, the reduction of the force constant leads to a destabilization and subsequent disappearance from the phase space. This is problematic for the reweighting methods as a state that is not observed in the biased data set cannot be reweighted. Thus, an additional biasing term that stabilizes the cis-open states might be necessary to correctly sample all minima in the phase space. Figure 6.9B shows some example raw trajectories. The cis and trans states are still well distinguishable and are distributed around 0 and -π rad, respectively. The corresponding cis/trans transition plot in Figure 6.9C shows that the reduction of the force constant to $27\,\mathrm{kJ\,mol^{-1}}$ is sufficient to significantly increase the cis to trans (90 events) and trans to cis (83 events) transition rates. However, as already observed above, the cis state becomes more unstable. Simulations where the 9-10 peptide bond remains in the cis-configuration become rare and less simulation time is spent in the cis states in general.

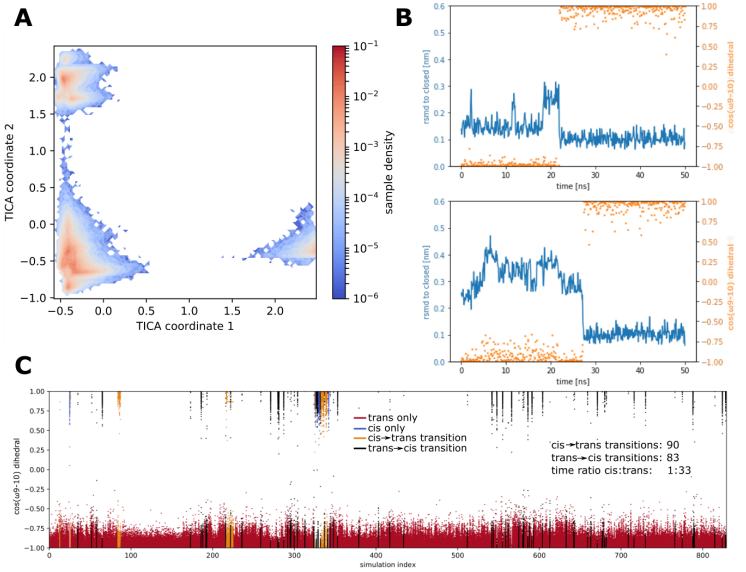


FIGURE 6.9: Description of the moderately biased data set. (**A**): Phase space visited in the simulations projected to the TICA coordinates of the unbiased reference system. (**B**): Example trajectories of the moderately biased simulations. The RMSD to the cis-closed state is shown with blue lines, the cosine of the ω_{9-10} dihedral angle is shown in orange. (**C**): Distribution of the ω_{9-10} dihedral angle for the individual trajectories as a function of the corresponding simulation index. Data points are colored based on the state of the ω_{9-10} dihedral angle: only trans (red), only cis (blue), only cis to trans transitions (orange), trans to cis transitions (and potentially an additional cis to trans transition) (black).

To highlight the contrast between the moderately biased and the highly biased simulations, Figure 6.10 shows the same plots when the force constant is reduced to 50% or less. With an increasing bias, the dihedral term is increasingly softened and the effective dihedral-angle distribution shifts. Especially when reducing the force constant to 24%, the ω_{9-10} dihedral angle is sampling the full range without clear cis/trans transitions anymore.

Figure 6.9C showed that moderate biasing results indeed in the desired higher frequency of the cis/trans isomerization of the 9-10 peptide bond. To recover the original energy landscape and kinetics, we applied different dynamic reweighting strategies: DHAM[46] in the short-time and long-time approximation and TRAM.[6] Figure 6.11 shows the reweighting results on the MSM constructed using Sittel-Stock clustering. The unbiased reference model is shown in Figure 6.11A. Without reweighting (Figure 6.11B), the minima are 'smeared out' in comparison to the reference and the top-right cluster is not sampled. Although all reweighting methods (Figure 6.11C-E) were able to slightly reduce this 'smearing', none was able to fully refocus the positions of the minima position compared to the reference. As expected, none of the reweighting methods could recover the lost top-right cluster. Looking at the kinetic spectrum, none of the reweighting methods leads to a convincing result. In the case of DHAM with long-time approximation, the kinetics were still underestimated by 30% (in comparison to a 50% underestimation in the not-reweighted data). DHAM with short-time approximation even led to a higher discrepancy with the reference. Note that TRAM reweighting only returns the reweighted equilibrium distribution without the kinetics.

Similar results were obtained with reweighting the MSM con-

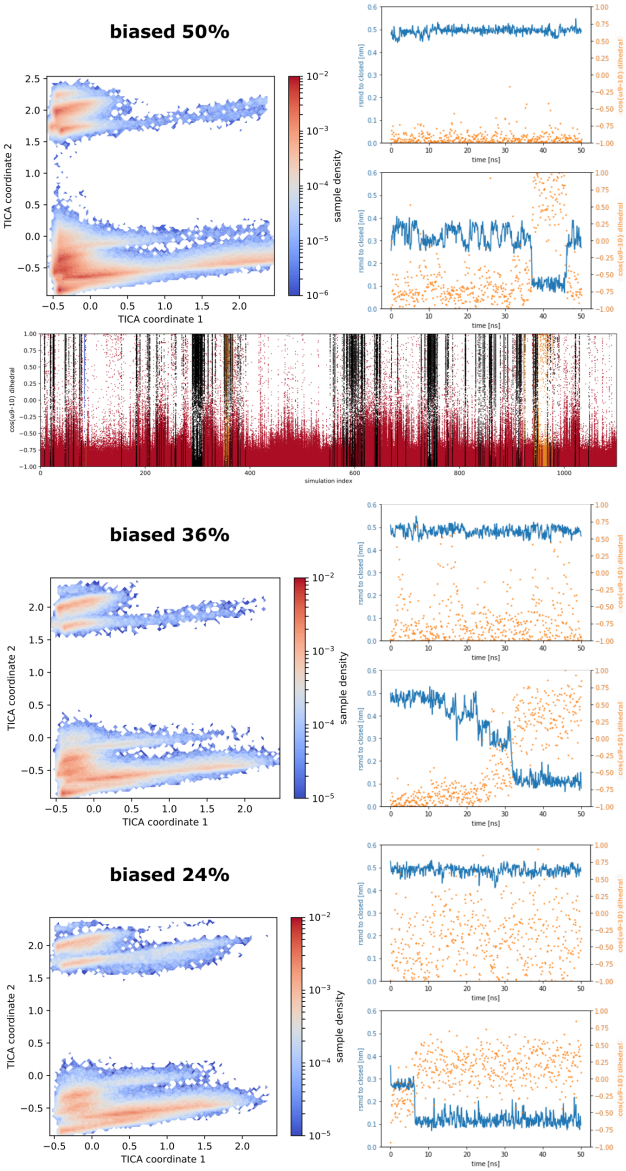


FIGURE 6.10: Description of the highly biased data set (reduction to 50%, 36%, and 24%). For each bias, the phase space visited in the simulations projected to the TICA coordinates of the unbiased reference system is shown as well as example trajectories. The RMSD to the cis-closed state is shown with blue lines, the cosine of the ω_{9-10} dihedral angle is shown in orange. In addition, the cis/trans transition plot is provided for the 50% biased data set.

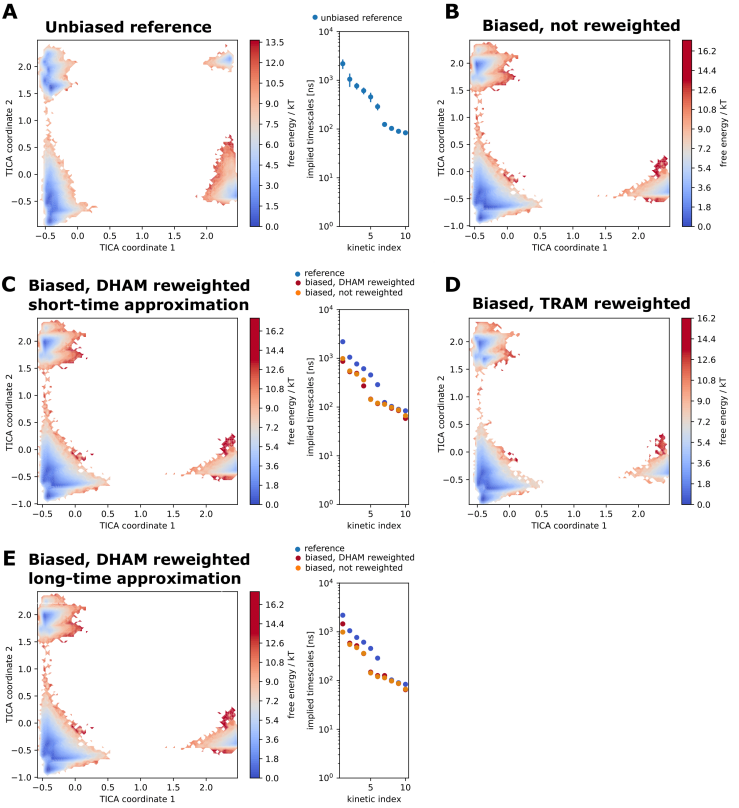


FIGURE 6.11: Reweighting results for MSMs using Sittel-stock clustering with moderately biased data. (**A**): Unbiased free-energy distribution projected to the TICA coordinates and reference kinetic spectrum. (**B**): Free-energy distribution of the moderately biased simulations projected to the unbiased phase space. (**C**): Free-energy distribution and kinetic spectrum of the DHAM reweighted (short-time approximation) biased data. Reweighted kinetics are shown in blue, not reweighted kinetics in orange. (**D**): Free-energy distribution of the TRAM reweighted biased data. (**E**): Free-energy distribution and kinetic spectrum of the DHAM reweighted (long-time approximation) biased data. Reweighted kinetics are shown in blue, not reweighted kinetics in orange.

structed using k-means clustering (Figure 6.12). The reweighted kinetics were only marginally better than the not-reweighted benchmark and 'minima smearing' was still an issue in the energy plot.

On a first glance, the reweighted results on the highly biased force constant (50% reduction) seem more successful (Figure 6.13 for Sittel-stock clustering and Figure 6.14 for k-means clustering). The kinetics of the DHAM in the short-time approximation are identical to the reference when using the Sittel-Stock clustering and slightly off with k-means clustering. In addition, the model is able to identify the transition region between the cis and trans state as a high energy region. However, the minima remain still 'smeared', and the seemingly good results on the kinetics are misleading. Considering the fact that the cis and trans distributions in the highly biased data set are shifted towards unnatural dihedral angles, one would expect much higher energies and slower kinetics. For example, the former cis state at 0° became centered around an -57°'angle. The unbiased kinetics leading to that state should be extremely slow and not on the same magnitude as a conventional cis/trans isomerization. Thus, also in the highly biased case we can conclude that reweighting did not recover the expected kinetics.

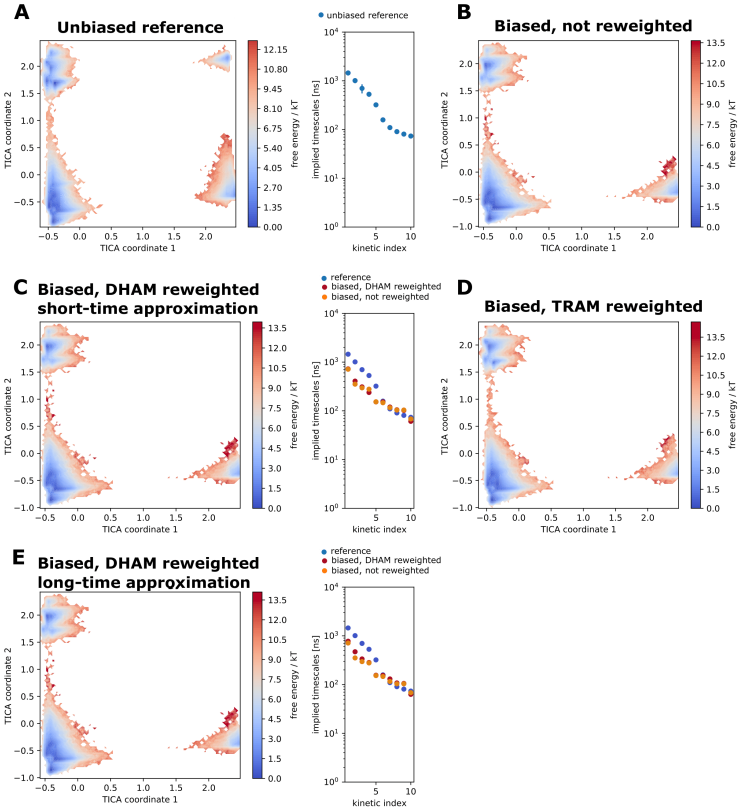


FIGURE 6.12: Reweighting results for MSMs using k-means clustering and moderately biased data. (**A**): Unbiased free-energy distribution projected to the TICA coordinates and reference kinetic spectrum. (**B**): Free-energy distribution of the moderately biased simulations projected to the unbiased phase space. (**C**): Free-energy distribution and kinetic spectrum of the DHAM reweighted (short-time approximation) biased data. Reweighted kinetics are shown in blue, not reweighted kinetics in orange. (**D**): Free-energy distribution of the TRAM reweighted biased data. (**E**): Free-energy distribution and kinetic spectrum of the DHAM reweighted (long-time approximation) biased data. Reweighted kinetics are shown in blue, not reweighted kinetics in orange.

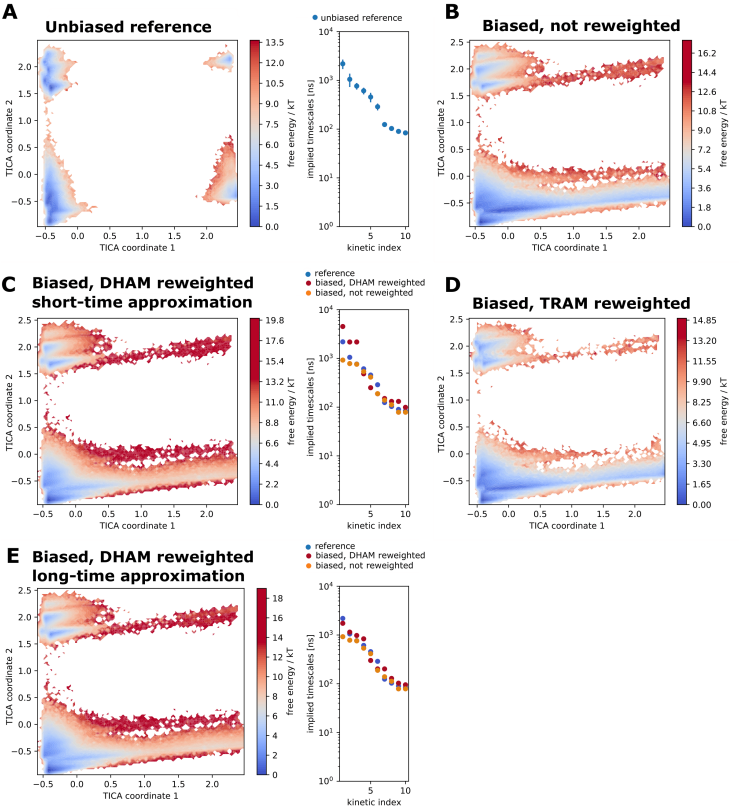


FIGURE 6.13: Reweighting results for MSMs using Sittel-stock clustering and highly biased data. (**A**): Unbiased free-energy distribution projected to the TICA coordinates and reference kinetic spectrum. (**B**): Free-energy distribution of the moderately biased simulations projected to the unbiased phase space. (**C**): Free-energy distribution and kinetic spectrum of the DHAM reweighted (short-time approximation) biased data. Reweighted kinetics are shown in blue, not reweighted kinetics in orange. (**D**): Free-energy distribution of the TRAM reweighted biased data. (**E**): Free-energy distribution and kinetic spectrum of the DHAM reweighted (long-time approximation) biased data. Reweighted kinetics are shown in blue, not reweighted kinetics in orange.

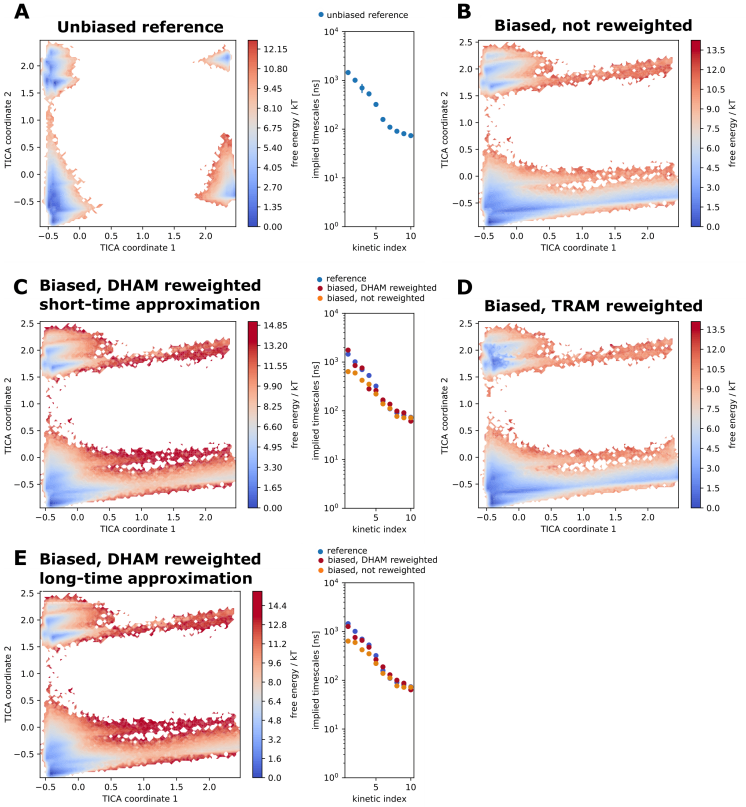


FIGURE 6.14: Reweighting results for MSMs using k-means clustering and highly biased data. (**A**): Unbiased free-energy distribution projected to the TICA coordinates and reference kinetic spectrum. (**B**): Free-energy distribution of the moderately biased simulations projected to the unbiased phase space. (**C**): Free-energy distribution and kinetic spectrum of the DHAM reweighted (short-time approximation) biased data. Reweighted kinetics are shown in blue, not reweighted kinetics in orange. (**D**): Free-energy distribution of the TRAM reweighted biased data. (**E**): Free-energy distribution and kinetic spectrum of the DHAM reweighted (long-time approximation) biased data. Reweighted kinetics are shown in blue, not reweighted kinetics in orange.

6.4 CONCLUSION

Revisiting the dynamics of CycA showed that extensive sampling is necessary to sufficiently sample the rare cis/trans isomerization in the peptide backbone between residues 9 and 10. Such a a large amount of accumulated sampling time was not reached by previous published models.[16,80] Due to insufficient sampling and discretization errors, these models significantly underestimated the transition kinetics of CycA (i.e., around 100 ns). Our improved model corrected these timescales to $> 1\,\mu s$, a difference of one order of magnitude. In addition, by analyzing the MSM beyond metastable states, we were able to identify an additional closed state of CycA with all trans peptide bonds (termed trans-closed). In contrast to the well known cis-closed state (corresponding to the small molecule crystal structure), this trans-closed state has fast interconversion kinetics with the open states and a higher equilibrium population. Thus, CycA is not only conformationally flexible, but also 'kinetically' flexible, i.e., it exhibits two very different transition rates between the open and closed states: a slow one on the order of µs and a fast one on the order of ns. It is possible that this kinetic flexibility provides advantages for the passive permeation of CycA through a lipid membrane. For example, the fast interconversion rates might be favorable for the re-solvation of the peptide after passing the apolar core of the membrane. Kinetic models of cyclosporin E (CsE), which differs from CycA by a single backbone methylation but has a ten fold decreased permeability, have shown that CsE has a metastable trans-closed state.[96] Thus, in contrast to the permeable CycA, the impermeable CsE is not kinetically flexible. Further analysis studying kinetic flexibility in a more complex environment,

i.e., a polar/apolar interface or a lipid membrane, will reveal its significance for permeability.

In addition to improving the kinetic model of CycA, this chapter explored the usage of enhanced sampling techniques and dynamic reweighting to obtain MSMs with less computational resources. The rare cis/trans isomerization of CycA requires sampling times above 100 µs. Thus, approaches that sufficiently sample this transition in a fraction of the time would be highly valuable. We showed that an increasing reduction of the dihedral force constant for the ω_{9-10} peptide bond can lead to shifts in the energy minima due to the increasing dominance of the 1-4 nonbonded interactions. Thus, simultaneous reduction of the 1-4 nonbonded interactions, addition of phase-shifted dihedral terms or local elevation coupled with umbrella sampling (LEUS)[247] may be more suitable ways to increase the transition frequency of the cis/trans isomerization. Compared to the not reweighted benchmark, the dynamic reweighting methods led to slightly improved but unfortunately not accurate enough results. Major problems were 'smearing' of the free-energy minima and underestimation of the kinetics. In addition, a considerable amount of human fine-tuning was needed for the approach with enhanced sampling and dynamic reweighting. Thus, we conclude that dynamic reweighting methods in the current form do not provide the desired speed-up when working with small systems like cyclic peptides. Should future developments lead to significant improvements in the usability and performance of these methods, the usage of dynamic reweighting should be reassessed.

We want to conclude this chapter with highlighting pitfalls and best practices for MSMs as well as enhanced sampling and reweighting.

Markov State Models

1. **Feature selection and dimension reduction.** This first step of building the MSM defines which kinetics can be considered. Some clustering methods require a strong dimension reduction (e.g., the Sittel-Stock clustering usually handles only five to six input dimensions). The user has to carefully check if this massive reduction is still able to reflect the complexity of the system. The variance explained by the chosen dimensions can be a good measure.

2. **Clustering.** Improper clustering can lead to false transitions, falsely connected states, and thus to incorrect kinetics. A good example is the previous model for CycA, where the rarest transition was not sufficiently sampled and thus states were disconnected. However, improper clustering led to false transitions and thus suggested a connected model. Some examples of improper clustering are displayed in Figure 6.15. The user has to carefully test the model for proper clustering. In general, we advice to use the more robust density-based clustering methods such as Sittel-Stock.

3. **Verification on raw data.** We further advise too double check the results from the high level MSM on the level of the raw trajectory data. This can help to track down the problems described in points 1 and 2.

4. **Occurrence of non-metastable states.** As the example of the kinetic flexibility of CycA has shown, the most important insights are not always found on the level of the metastable states. The user should evaluate the results on all levels of the MSM to gain an comprehensive understanding of the system.

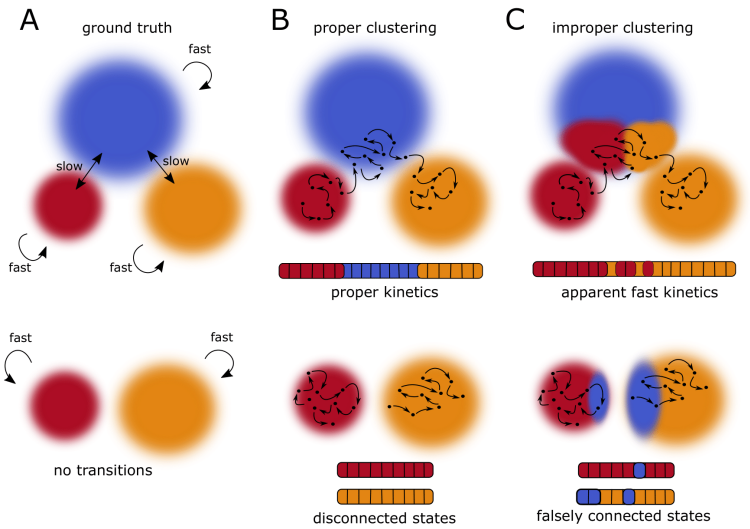


FIGURE 6.15: Examples of how misleading MSM results can emerge from improper clustering.

Dynamic Reweighting

1. **Proper biasing.** As extensively discussed in this chapter, the applied bias should speed up the desired transition without disturbing the remaining system or shifting energy minima.

2. **Usability.** In their current form, most dynamic reweighting approaches need an implementation by the user, which limits the dissemination and usage of these techniques. A package containing efficient implementations of different reweighting approaches would be desirable.

3. **Performance.** As discussed above, the results of dynamic reweighting still lack accuracy. Thus, the application of dynamic reweighting methods should be considered well for

each system.

4. **Metastable state extraction.** In their current form, none of the dynamic reweighting approaches yields metastable states. Due to the reweighting, the detailed balance assumption is violated, which prevents the use of PCCA-based to assign metastable sets. As the metastable states are often of particular interest, techniques to recover metastable sets from reweighted methods should be developed.

7

Outlook

"If I had asked people what they wanted, they would have said faster horses"

Henry Ford

This quote and chapter are dedicated to Frederik Zwilling and his visionary mind and his calm confidence

Chapters 2 - 6 focused on the permeability analysis of cyclic peptides with a special focus on rare events that dominate this kinetics. In Chapters 3 and 4, we extended the current picture of cyclic peptide permeability, while in Chapters 5 and 6 we explored new methods to extend the boundaries of kinetic models for rare events.

7.1 DESIGN OF PERMEABLE CYCLIC PEPTIDES

A low permeability – and thus a poor oral bioavailability – is a problem not only limited to cyclic peptides or macrocyclic drugs. New insights in the permeability process and guidelines for permeable drugs will benefit the whole drug discovery field. Chapter 2 introduced the challenges associated with permeability analysis and how to approach them using the combined power of computational and experimental techniques. We hope that this review will inspire new collaborations between experts with an experimental and computational background. We believe that the combination of sound experimental parameters and the atomistic insights from computer simulations is necessary to lead to a holistic view of passive permeability, extend current models, and thus advance the field. An example of how the combination of molecular dynamics simulations together with experimental permeability values can lead to the extension of current models is given in Chapters 3 and 4. Chapter 3 highlights the importance of considering the local environment during the permeation process using kinetic models. Our work in agreement with previous findings in the literature showed that cyclic peptides prefer to be located at the interface

of a polar and apolar environment (e.g., the polar head group and the apolar tail regions).[114,115] Thus, the kinetics at such an interface are likely decisive for the whole permeation process. In addition, such an interface leads to an anisotropic and directional environment, which causes the splitting into two metastable orientations. We expect that the interface system we introduced will be useful to study the dynamics of other series of cyclic peptides or drug molecules in general. This simple system mimics the condition at a lipid membrane but has the advantage of much faster kinetics, thus less simulation time is needed to reach convergence. To explore the transferability of the interface results to a more realistic lipid membrane system, Chapter 4 presents atomistic insights into the permeation pathway at a lipid membrane. We identified four permeation steps: (1) anchoring with residues in transient gaps between lipid head groups, (2) insertion in the membrane and orienting parallel to the membrane plane, (3) if the peptide enters in an 'open' conformation, interconversion to the 'closed' permeable conformation, and (4) leaflet crossing involving anchoring and rotation. The membrane model confirmed the preference of cyclic peptides to locate at the interface and showed highly transferable results to the interface system with respect to the cyclic peptide orientation. The anchoring observed in steps (1) and (4) were exclusively found in the lipid membrane system. This is no surprise as only the sterically hindered environment at a lipid membrane generates the need for an anchor. Anchors that help the molecule to insert into the membrane are a new design concept that might be of value also for other compounds. A recent experimental paper highlighted the occurrence of acyl chains in natural products that increase membrane permeability.[194] Based on the findings from of our computer simulations, we propose that these acyl chains act as anchors to enable mem-

brane entrance. Together, the insights from these experiments and our simulations generate a holistic view on the processes involved and suggest that attaching anchors to drug compounds can increase permeability and thus oral bioavailability. Multiple additional influence factors are expected to modulate the permeability process. For example, Sugita *et. al.*[189] showed that adding cholesterol to the lipid membrane shifts the energy landscape of permeation. Furthermore, the lipid composition of the membrane will likely influence the occurrence of headgroup gaps and the dynamics of the cyclic peptides inside the membrane. Thus, future research is needed to compare different lipid compositions. In addition, membrane proteins are expected to perturb the membrane. Therefore, it would be interesting to study if membrane insertions happen preferentially close to proteins. In addition, the glycocalyx covering the cell membrane has to be considered to mimic cellular systems. One can hypothesise that the glycocalyx either sterically hinders the cyclic peptide to reach the cell or, on the contrary, can act as a net and keep the peptide close to the membrane. Last but not least, although quite challenging, it might be interesting to study the influence of the chemical gradient. In the presented chapters, we solely performed equilibrium simulations. However, the chemical gradient is expected to act as a driving force to facilitate permeability.

7.2 APPLICATION OF DYNAMIC REWEIGHTING METHODS AND KINETIC MODES

Due to the slow kinetics of the four permeation steps, it is currently not feasible to build a full kinetic model of a cyclic peptide permeating through a lipid membrane. Thus, in the second part of this thesis we explored techniques that can extract unbiased kinetic and equilibrium information from biased simulations – termed dynamic reweighting techniques – and thereby expand the scope of kinetic models. In Chapter 5, we compared current approaches in dynamic reweighting and derived their physical connection. On a one dimensional test system, we were able to compare their performance and show limitations. Additionally, we propose an improved reweighting algorithm that can overcome some of these limitations. However, the insights gained were limited on theoretical considerations and a one dimensional system. Thus, further work will be necessary to compare the performance for different enhanced sampling and reweighting applications on a more complex system. This causes an obstacle especially for the so-called path-based reweighting methods that require modifications on the level of the numerical integration. Limited usability and accessibility, low robustness, and the need for manual curation of the models are currently the main hurdles for dynamic reweighting techniques. Thus, future software improvements are most significant to advance the field. First steps towards the assessment of dynamic reweighting techniques for more complex and atomistic systems are given in Chapter 6. There, we applied dynamic reweighting techniques on a cyclic peptide test system

to sample a rare cis/trans isomerization in the peptide backbone. Our findings show limitations of previous kinetic models on the same system and give guidelines on the application of kinetic models and dynamic reweighting techniques.

Abbreviations

1SS	one starting state
API	application programming interface
ATB	automated topology builder
COM	Center of Mass
CSMM	Core-set Markov Model
DAD	Diode Array Detection
EMMA	Emma's Markov Model Algorithms
GROMACS	Groningen Machine for Chemical Simulations
GROMOS	Groningen Molecular Simulation
LC-MS	Liquid Chromatography–Mass Spectrometry
LC-UV	Liquid Chromatography-Ultraviolet Spectroscopy
LINCS	Linear Constraint Solver
MD	Molecular Dynamics
MFPT	Mean First Passage Time
MSM	Markov State Model
NMR	Nuclear Magnetic Resonance
PCCA+	Robust Perron Cluster Cluster Analysis
PME	Particle Mesh Ewald
POPC	1-Palmitoyl-2-oleoylphosphatidylcholine
RMSD	Root-mean Square Deviation
SPC	Simple-Point-Charge
TICA	Time-lagged Independent Component Analysis
UPLC	Ultra-Performance Liquid Chromatography

Bibliography

[1] R. O. Dror, R. M. Dirks, J. Grossman, H. Xu, D. E. Shaw, *Biomolecular Simulation: A Computational Microscope for Molecular Biology* , Annu. Rev. Biophys. 41 (2012) 429–452.

[2] D. A. Schuetz, W. E. A. de Witte, Y. C. Wong, B. Knasmueller, L. Richter, D. B. Kokh, S. K. Sadiq, R. Bosma, I. Nederpelt, L. H. Heitman, E. Segala, M. Amaral, D. Guo, D. Andres, V. Georgi, L. A. Stoddart, S. Hill, R. M. Cooke, C. De Graaf, R. Leurs, M. Frech, R. C. Wade, E. C. M. de Lange, A. P. IJzerman, A. Müller-Fahrnow, G. F. Ecker, *Kinetics for Drug Discovery: An Industry-Driven Effort to Target Drug Residence Time* , Drug Discov. Today 22 (2017) 896–911.

[3] F. Paul, C. Wehmeyer, E. T. Abualrous, H. Wu, M. D. Crabtree, J. Schöneberg, J. Clarke, C. Freund, T. R. Weikl, F. Noé, *Protein-Peptide Association Kinetics Beyond the Seconds Timescale from Atomistic Simulations* , Nat. Commun. 8 (2017) 1095.

[4] A. Mitsutake, Y. Mori, Y. Okamoto, *Enhanced Sampling Algorithms* , in: L. Monticelli, E. Salonen (Eds.), *Biomolecular Simulations: Methods and Protocols* , Humana Press, Totowa, NJ, 2013, pp. 153–195.

[5] R. C. Bernardi, M. C. Melo, K. Schulten, *Enhanced Sampling Techniques in Molecular Dynamics Simulations of Bi-*

ological Systems , Biochim. Biophys. Acta Gen. Subj. 1850 (2015) 872–877.

[6] H. Wu, F. Paul, C. Wehmeyer, F. Noé, *Multiensemble Markov Models of Molecular Thermodynamics and Kinetics*, Proceed. Nat. Acad. Sci. USA 113 (2016) e3221.

[7] L. S. Stelzl, A. Kells, E. Rosta, G. Hummer, *Dynamic Histogram Analysis To Determine Free Energies and Rates from Biased Simulations* , J. Chem. Theory Comput. 13 (2017) 6328–6342.

[8] L. Donati, B. G. Keller, *Girsanov Reweighting for Metadynamics Simulations* , J. Chem. Phys. 149 (2018) 072335.

[9] W. E, E. Vanden-Eijnden, *Transition-Path Theory and Path-Finding Algorithms for the Study of Rare Events* , Annu. Rev. Phys. Chem. 61 (2010) 391–420.

[10] M. A. Rohrdanz, W. Zheng, C. Clementi, *Discovering Mountain Passes via Torchlight: Methods for the Definition of Reaction Coordinates and Pathways in Complex Macromolecular Reactions* , Annu. Rev. Phys. Chem. 64 (2013) 295–316.

[11] A. L. Hopkins, C. R. Groom, *The Druggable Genome* , Nature Rev. Drug Discov. 1 (2002) 727–730.

[12] E. Danelius, V. Poongavanam, S. Peintner, L. H. E. Wieske, M. Erdélyi, J. Kihlberg, *Solution Conformations Explain the Chameleonic Behaviour of Macrocyclic Drugs* , Chem. Eur. J. 26 (2020) 5231–5244.

[13] A. Rüegger, M. Kuhn, H. Lichti, H.-R. Loosli, R. Huguenin, C. Quiquerez, A. von Wartburg, *Cyclosporin A, ein Immunsuppressiv Wirksamer Peptidmetabolit aus Trichoderma*

Polysporum (LINK ex PERS.) Rifai , Helv. Chim. Acta 59 (1976) 1075–1092.

[14] H. Loosli, H. Kessler, H. Oschkinat, H. Weber, T. J. Petcher, A. Widmer, *Peptide Conformations. Part 31. The Conformation of Cyclosporin A in the Crystal and in Solution* , Helv. Chim. Acta 68 (1985) 682–704.

[15] H. Kessler, M. Köck, T. Wein, M. Gehrke, *Reinvestigation of the Conformation of Cyclosporin A in Chloroform* , Helv. Chim. Acta 73 (1990) 1818–1832.

[16] 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 (2016) 1547–1562.

[17] O. Guvench, A. D. MacKerell, Comparison of Protein Force Fields for Molecular Dynamics Simulations, in: A. Kukol (Ed.), Molecular Modeling of Proteins, Humana Press, Totowa, NJ, 2008, pp. 63–88.

[18] P. E. M. Lopes, O. Guvench, A. D. MacKerell, Current Status of Protein Force Fields for Molecular Dynamics Simulations, in: A. Kukol (Ed.), Molecular Modeling of Proteins, Springer New York, New York, NY, 2015, pp. 47–71.

[19] S. A. Hollingsworth, R. O. Dror, *Molecular Dynamics Simulation for All* , Neuron 99 (2018) 1129–1143.

[20] D. Frenkel, B. Smit, *Free Energies and Phase Equilibria*, in: D. Frenkel, B. Smit (Eds.), *Understanding Molecular Simulation* , 2nd Edition, Academic Press, San Diego, 2002, pp. 1–6.

[21] T. J. Lane, D. Shukla, K. A. Beauchamp, V. S. Pande, *To Milliseconds and Beyond: Challenges in the Simulation of Protein Folding* , Curr. Opin. Struct. Biol. 23 (2013) 58–65.

[22] V. S. Pande, K. Beauchamp, G. R. Bowman, *Everything You Wanted to Know About Markov State Models But Were Afraid to Ask* , Methods 52 (2010) 99–105.

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

[24] B. E. Husic, V. S. Pande, *Markov State Models: From an Art to a Science* , J. Am. Chem. Soc. 140 (2018) 2386–2396.

[25] J. D. Chodera, W. C. Swope, J. W. Pitera, K. A. Dill, *Long-Time Protein Folding Dynamics from Short-Time Molecular Dynamics Simulations* , Multiscale Model. Simul. 5 (2006) 1214–1226.

[26] P. Debye, *Näherungsformeln für die Zylinderfunktionen für große Werte des Arguments und unbeschränkt veränderliche Werte des Index* , Math. Ann. 67 (1909) 535–558.

[27] M. R. Hestenes, E. Stiefel, *Methods of Conjugate Gradients for Solving Linear Systems*, J. Res. Natl. Inst. Stand. Technol. 49 (1952) 409–436.

[28] W. K. Hastings, *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*, Biometrika. 57 (1970) 97–109.

[29] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, *Equation of State Calculations by Fast Computing Machines*, J. Chem. Phys. 21 (1953) 1087–1092.

[30] I. B. Cohen, A. Whitman, J. Budenz, *The Principia: Mathematical Principles of Natural Philosophy*, 1st Edition, University of California Press, 1999.

[31] L. Verlet, *Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules* , Phys. Rev. 159 (1967) 98–103.

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

[33] C. Runge, *Über die numerische Auflösung von Differentialgleichungen*, Math. Ann. 46 (1895) 167–178.

[34] M. Kutta, *Beitrag zur Näherungsweisen Integration Totaler Differentialgleichungen* , Zeitschrift für Mathematik und Physik 46 (1901) 435—453.

[35] W. F. van Gunsteren, H. J. C. Berendsen, *Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry*, Angew. Chem. Int. Ed. 29 (1990) 992–1023.

[36] J. W. Gibbs, *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics* , Cambridge Library Collection - Mathematics, Cambridge University Press, 1902.

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

[38] S. Nosé, *A Molecular Dynamics Method for Simulations in the Canonical Ensemble*, Mol. Phys. 52 (1984) 255–268.

[39] S. Nosé, *A Unified Formulation of the Constant Temperature Molecular Dynamics Methods*, J. Chem. Phys. 81 (1984) 511–519.

[40] W. G. Hoover, *Canonical Dynamics: Equilibrium Phase-Space Distributions*, Phys. Rev. A. 31 (1985) 1695–1697.

[41] G. J. Martyna, M. L. Klein, M. Tuckerman, *Nosé-Hoover Chains: The Canonical Ensemble via Continuous Dynamics*, J. Chem. Phys. 97 (1992) 2635–2643.

[42] M. Parrinello, A. Rahman, *Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method*, Int. J. Appl. Phys. 52 (1981) 7182–7190.

[43] H. C. Andersen, *Molecular Dynamics Simulations at Constant Pressure and/or Temperature*, J. Chem. Phys. 72 (1980) 2384–2393.

[44] S. Nosé, M. Klein, *Constant Pressure Molecular Dynamics for Molecular Systems*, Mol. Phys. 50 (1983) 1055–1076.

[45] J. K. Weber, V. S. Pande, *Potential-Based Dynamical Reweighting for Markov State Models of Protein Dynamics* , J. Chem. Theory Comput. 11 (2015) 2412–2420.

[46] E. Rosta, G. Hummer, *Free Energies From Dynamic Weighted Histogram Analysis Using Unbiased Markov State Model* , J. Chem. Theory Comput. 11 (2015) 276–285.

[47] Y. Sugita, Y. Okamoto, *Replica-Exchange Molecular Dynamics Method for Protein Folding*, Chem. Phys. Lett. 314 (1999) 141–151.

[48] S. Piana, A. Laio, *A Bias-Exchange Approach to Protein Folding* , J. Phys. Chem. B 111 (2007) 4553–4559.

[49] T. Huber, A. E. Torda, W. F. van Gunsteren, *Local Elevation: A Method for Improving the Searching Properties of Molecular Dynamics Simulation* , J. Comput.-Aided Mol. Des. 8 (1994) 695–708.

[50] A. Laio, M. Parrinello, *Escaping Free-energy Minima* , Proceed. Nat. Acad. Sci. USA 99 (2002) 12562–12566.

[51] G. Hummer, I. G. Kevrekidis, *Coarse Molecular Dynamics of a Peptide Fragment: Free Energy, Kinetics, and Long-Time Dynamics Computations* , J. Chem. Phys. 118 (2003) 10762–10773.

[52] F. Noé, S. Fischer, *Transition Networks for Modeling the Kinetics of Conformational Change in Macromolecules* , Curr. Opin. Struct. Biol. 18 (2008) 154–162.

[53] G. R. Bowman, K. A. Beauchamp, G. Boxer, V. S. Pande, *Progress and Challenges in the Automated Construction of Markov State Models for Full Protein Systems* , J. Chem. Phys. 131 (2009) 124101.

[54] N. Singhal, V. S. Pande, *Error Analysis and Efficient Sampling in Markovian State Models for Molecular Dynamics* , J. Chem. Phys. 123 (2005) 204909.

[55] G. R. Bowman, D. L. Ensign, V. S. Pande, *Enhanced Modeling via Network Theory: Adaptive Sampling of Markov State Models* , J. Chem. Theory Comput. 6 (2010) 787–794.

[56] J. K. Weber, V. S. Pande, *Characterization and Rapid Sampling of Protein Folding Markov State Model Topologies* , J. Chem. Theory Comput. 7 (2011) 3405–3411.

[57] N. S. Hinrichs, V. S. Pande, *Calculation of the Distribution of Eigenvalues and Eigenvectors in Markovian State Models for Molecular Dynamics* , J. Chem. Phys. 126 (2007) 244101.

[58] B. Trendelkamp-Schroer, H. Wu, F. Paul, F. Noé, *Estimation and Uncertainty of Reversible Markov Models* , J. Chem. Phys. 143 (2015) 174101.

[59] C. Schütte, W. Huisinga, *Biomolecular Conformations can be Identified as Metastable Sets of Molecular Dynamics*, in: Special Volume, Computational Chemistry, Vol. 10 of *Handbook of Numerical Analysis*, Elsevier, 2003, pp. 699–744.

[60] C. Schütte, M. Sarich, *Metastability and Markov State Models in Molecular Dynamics: Modeling, Analysis, Algorithmic Approaches* , American Mathematical Society, 2013.

[61] G. Caron, J. Kihlberg, G. Goetz, E. Ratkova, V. Poongavanam, G. Ermondi, *Steering New Drug Discovery Campaigns: Permeability, Solubility, and Physicochemical Properties in the bRo5 Chemical Space* , ACS Med. Chem. Lett. 12 (2021) 13–23.

[62] 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 (2015) 2312–2327.

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

[64] H. Nakajima, Y. B. Kim, H. Terano, M. Yoshida, S. Horinouchi, *FR901228, a Potent Antitumor Antibiotic, is a*

Novel Histone Deacetylase Inhibitor , Exp. Cell Res. 241 (1998) 126–133.

[65] I. L. O. Buxton, L. Z. Benet, *Pharmacokinetics: The Dynamics of Drug Absorption, Distribution, Metabolism, and Elimination* , McGraw-Hill Education, New York, NY, 2015.

[66] W. Stein, *Transport and Diffusion Across Cell Membranes* , Academic Press, San Diego, CA, 1986.

[67] N. J. Yang, M. J. Hinner, *Getting Across the Cell Membrane: An Overview for Small Molecules, Peptides, and Proteins* , Methods Mol. Biol. 1266 (2015) 29–53.

[68] L. Di, E. H. Kerns, *Transporters* , Academic Press, San Diego, CA, 2016.

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

[70] C. R. Pye, W. M. Hewitt, J. Schwochert, T. D. Haddad, C. E. Townsend, L. Etienne, Y. Lao, C. Limberakis, A. Furukawa, A. M. Mathiowetz, D. A. Price, S. Liras, R. S. Lokey, *Nonclassical Size Dependence of Permeation Defines Bounds for Passive Adsorption of Large Drug Molecules* , J. Med. Chem. 60 (2017) 1665–1672.

[71] 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 (2001) 3–26.

[72] P. G. Dougherty, A. Sahni, D. Pei, *Understanding Cell Penetration of Cyclic Peptides* , Chem. Rev. 119 (2019) 10241–10287.

[73] D. Gang, D. W. Kim, H. S. Park, *Cyclic Peptides: Promising Scaffolds for Biopharmaceuticals* , Genes 9 (2018) 557.

[74] W. R. Lieb, W. D. Stein, *Implications of Two Different Types of Diffusion for Biological Membranes* , Nat. New Biol. 234 (1971) 220–222.

[75] C. Chipot, J. Comer, *Subdiffusion in Membrane Permeation of Small Molecules* , Sci. Rep. 6 (2016) 35913.

[76] M. Merion, Robert, J. G. White, David, S. Thiru, B. Evans, David, Y. Calne, Roy, *Cyclosporine: Five years' Experience in Cadaveric Renal Transplantation* , N. Engl. J. Med. 310 (2010) 148–154.

[77] D. Tedesco, L. Haragsim, *Cyclosporine: A Review* , J. Transplant. 2012 (2012) 230386.

[78] P. F. Augustijns, T. P. Bradshaw, L. S. Gan, R. W. Hendren, D. R. Thakker, *Evidence for a Polarized Efflux System in Caco-2 Cells Capable of Modulating Cyclosporin A Transport* , Biochem. Biophys. Res. Commun. 197 (1993) 360–365.

[79] J. Lautz, H. Kessler, W. F. van Gunsteren, H. P. Weber, R. M. Wenger, *On the Dependence of Molecular Conformation on the Type of Solvent Environment: A Molecular Dynamics Study of Cyclosporin A* , Biopolymers 29 (1990) 1669–1687.

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

[81] K. M. Corbett, L. Ford, D. B. Warren, C. W. Pouton, D. K. Chalmers, *Cyclosporin Structure and Permeability: From A to Z and Beyond* , J. Med. Chem. 64 (2021) 13131–13151.

[82] S. Ono, M. R. Naylor, C. E. Townsend, C. Okumura, O. Okada, H. W. Lee, R. S. Lokey, *Cyclosporin A: Conformational Complexity and Chameleonicity* , J. Chem. Inf. Model. 61 (2021) 5601–5613.

[83] A. F. Räder, F. Reichart, M. Weinmüller, H. Kessler, *Improving Oral Bioavailability of Cyclic Peptides by N-methylation* , Bioorg. Med. Chem. 26 (2018) 2766–2773.

[84] B. I. Díaz-Eufracio, O. Palomino-Hernández, R. A. Houghten, J. L. Medina-Franco, *Exploring the Chemical Space of Peptides for Drug Discovery: A Focus on Linear and Cyclic Penta-Peptides* , Mol. Divers. 22 (2018) 259–267.

[85] Y. Li, W. Li, Z. Xu, *Improvement on Permeability of Cyclic Ceptide/Ceptidomimetic: Backbone N-methylation as a Useful Tool* , Mar. drugs 19 (2021) 311.

[86] O. Ovadia, S. Greenberg, J. Chatterjee, B. Laufer, F. Opperer, H. Kessler, C. Gilon, A. Hoffman, *The Effect of Multiple N-methylation on Intestinal Permeability of Cyclic Hexapeptides* , Mol. Pharm. 8 (2011) 479–487.

[87] 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. Leung, A. S. Kalgutkar, J. N. Bauman, Y. Zhang, S. Liras,

D. A. Price, A. M. Mathiowetz, M. P. Jacobson, R. S. Lokey, *On-Resin N-methylation of Cyclic Peptides for Discovery of Orally Bioavailable Scaffolds* , Nat. Chem. Biol. 7 (2011) 810–817.

[88] T. A. Hill, R. J. Lohman, H. N. Hoang, D. S. Nielsen, C. C. Scully, W. M. Kok, L. Liu, A. J. Lucke, M. J. Stoermer, C. I. Schroeder, S. Chaousis, B. Colless, P. V. Bernhardt, D. J. Edmonds, D. A. Griffith, C. J. Rotter, R. B. Ruggeri, D. A. Price, S. Liras, D. J. Craik, D. P. Fairlie, *Cyclic Penta- and Hexaleucine Peptides Without N-methylation are Orally Absorbed* , ACS Med. Chem. Lett. 5 (2014) 1148–1151.

[89] J. Witek, S. Wang, B. Schroeder, R. Lingwood, A. Dounas, H.-J. Roth, M. Fouché, M. Blatter, O. Lemke, B. Keller, S. Riniker, *Rationalization of the Membrane Permeability Differences in a Series of Analogue Cyclic Decapeptides* , J. Chem. Inf. Model. 59 (2019) 294–308.

[90] L. K. Buckton, M. N. Rahimi, S. R. McAlpine, *Cyclic Peptides as Drugs for Intracellular Targets: The Next Frontier in Peptide Therapeutic Development* , Chem. Eur. J. 27 (2021) 1487–1513.

[91] M. Kristensen, D. Birch, H. M. Nielsen, *Applications and Challenges for use of Cell-Penetrating Peptides as Delivery Vectors for Peptide and Protein Cargos* , Int. J. Mol. Sci. 17 (2016) 185.

[92] C. E. Overton, *Studien über die Narkose. Zugleich ein Beitrag zur Allgemeinen Pharmakologie* , Gustav Fischer, 1901.

[93] A. K. Ghose, V. N. Viswanadhan, J. J. Wendoloski, *Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods* , J. Phys. Chem. Lett. 102 (1998) 3762–3772.

[94] C. Wang, D. Craik, *Cyclic Peptide Oral Bioavailability: Lessons from the Past* , Biopolymers 106 (2016) 901–909.

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

[96] 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 (2017) 3309–3314.

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

[98] S. Ono, M. R. Naylor, C. E. Townsend, C. Okumura, O. Okada, R. S. Lokey, *Conformation and Permeability: Cyclic Hexapeptide Diastereomers* , J. Chem. Inf. Model. 59 (2019) 2952–2963.

[99] A. S. Kamenik, J. Kraml, F. Hofer, F. Waibl, P. K. Quoika, U. Kahler, M. Schauperl, K. R. Liedl, *Macrocycle Cell Permeability Measured by Solvation Free Energies in Polar and Apolar Environments* , J. Chem. Inf. Model 60 (2020) 3508–3517.

[100] A. Alex, D. S. Millan, M. Perez, F. Wakenhut, G. A. Whitlock, *Intramolecular Hydrogen Bonding to Improve Membrane Permeability and Absorption in Beyond Rule of Five Chemical Space* , MedChemComm 2 (2011) 669–674.

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

[102] T. Rezai, J. Bock, M. Zhou, C. Kalyanaraman, R. Lokey, M. 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 (2006) 14073–14080.

[103] A. Whitty, M. Zhong, L. Viarengo, D. Beglov, D. R. Hall, S. Vajda, *Quantifying the Chameleonic Properties of Macrocycles and Other High-Molecular-Weight Drugs* , Drug Discov. Today 21 (2016) 712–717.

[104] S. Riniker, *Toward the Elucidation of the Mechanism for Passive Membrane Permeability of Cyclic Peptides* , Future Med. Chem. 11 (2019) 637–639.

[105] H. N. Hoang, T. A. Hill, D. P. Fairlie, *Connecting Hydrophobic Surfaces in Cyclic Peptides Increases Membrane Permeability* , Angew. Chem. Int. Ed. 60 (2021) 8385–8390.

[106] C. Comeau, B. Ries, T. Stadelmann, J. Tremblay, S. Poulet, U. Fröhlich, J. Coté, P.-L. Boudreault, R. M. Derbali, P. Sarret, M. Grandbois, G. Leclair, S. Riniker, E. Marsault, *Modulation of the Passive Permeability of Semipeptidic*

Macrocycles: N- and C-Methylations Fine-Tune Conformation and Properties , J. Med. Chem. 64 (2021) 5365–5383.

[107] B. L. Santini, M. Zacharias, *Rapid in Silico Design of Potential Cyclic Peptide Binders Targeting Protein-Protein Interfaces* , Front. Chem. 8 (2020) 573259.

[108] S. Wang, G. König, H. Roth, M. Fouché, S. Rodde, S. Riniker, *Effect of Flexibility, Lipophilicity, and the Location of Polar Residues on the Passive Membrane Permeability of a Series of Cyclic Decapeptides* , J. Med. Chem. 64 (2021) 12761–12773.

[109] S. Wang, J. Witek, G. A. Landrum, S. Riniker, *Improving Conformer Generation for Small Rings and Macrocycles Based on Distance Geometry and Experimental Torsional-Angle Preferences* , J. Chem. Inf. Model. 60 (2020) 2044–2058.

[110] C. Schütte, A. Fischer, W. Huisinga, P. Deuflhard, *A Direct Approach to Conformational Dynamics Based on Hybrid Monte Carlo* , J. Comput. Phys. 51 (1999) 146–168.

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

[112] T. Stadelmann, G. Subramanian, S. Menon, C. E. Townsend, R. S. Lokey, M. O. Ebert, S. Riniker, *Connecting the Conformational Behavior of Cyclic Octadepsipeptides with Their Ionophoric Property and Membrane Permeability* , Org. Biomol. Chem. 18 (2020) 7110–7126.

[113] R. G. Weiß, B. Ries, S. Wang, S. Riniker, *Volume-Scaled*

Common Nearest Neighbor Clustering Algorithm with Free-Energy Hierarchy , J. Chem. Phys. 154 (2021) 084106.

[114] M. Sugita, S. Sugiyama, T. Fujie, Y. Yoshikawa, K. Yanagisawa, M. Ohue, Y. Akiyama, *Large-Scale Membrane Permeability Prediction of Cyclic Peptides Crossing a Lipid Bilayer Based on Enhanced Sampling Molecular Dynamics Simulations* , J. Chem. Inf. Model. 61 (2021) 3681–3695.

[115] S. M. Linker, C. Schellhaas, B. Ries, H.-J. Roth, M. Fouché, S. Rodde, S. Riniker, *Polar/Apolar Interfaces Modulate the Conformational Behavior of Cyclic Peptides with Impact on Their Passive Membrane Permeability* , RSC Adv. 12 (2022) 5782–5796.

[116] A. P. Russ, S. Lampel, *The Druggable Genome: An Update* , Drug Discov. Today 10 (2005) 1607–1610.

[117] M. Pertea, A. Shumate, G. Pertea, A. Varabyou, F. P. Breitwieser, Y.-C. Chang, A. K. Madugundu, A. Pandey, S. L. Salzberg, *CHESS: A New Human Gene Catalog Curated from Thousands of Large-Scale RNA Sequencing Experiments Reveals Extensive Transcriptional Noise* , Genome Biol. 19 (2018) 208.

[118] X.-P. Chen, G.-H. Du, *Target Validation: A Door to Drug Discovery* , Drug. Discov. Ther. 1 (2007) 23–29.

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

[120] E. Marsault, M. L. Peterson, *Macrocycles Are Great Cycles: Applications, Opportunities, and Challenges of Syn-*

thetic Macrocycles in Drug Discovery , J. Med. Chem. 54 (2011) 1961–2004.

[121] E. A. Villar, D. Beglov, S. Chennamadhavuni, J. A. Porco, D. Kozakov, S. Vajda, A. Whitty, *How Proteins Bind Macrocycles* , Nature Chem. Biol. 10 (2014) 723–731.

[122] F. J. Duffy, D. O'Donovan, M. Devocelle, N. Moran, D. J. O'Connell, D. C. Shields, *Virtual Screening Using Combinatorial Cyclic Peptide Libraries Reveals Protein Interfaces Readily Targetable by Cyclic Peptides* , J. Chem. Inf. Model. 55 (2015) 600–613.

[123] P. Dougherty, Z. Qian, D. Pei, *Macrocycles as Protein-Protein Interaction Inhibitors* , Biochem. J. 474 (2017) 1109–1125.

[124] M. F. Sanner, K. Zoghebi, S. Hanna, S. Mozaffari, S. Rahighi, R. K. Tiwari, K. Parang, *Cyclic Peptides as Protein Kinase Inhibitors: Structure–Activity Relationship and Molecular Modeling* , J. Chem. Inf. Model. 61 (2021) 3015–3026.

[125] T. Morioka, N. D. Loik, C. J. Hipolito, Y. Goto, H. Suga, *Selection-Based Discovery of Macrocyclic Peptides for the Next Generation Therapeutics* , Curr. Opin. Chem. Biol. 26 (2015) 34–41.

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

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

[128] A. F. Raeder, F. Reichart, M. Weinmüller, H. Kessler, *Improving Oral Bioavailability of Cyclic Peptides by N-Methylation* , Bioorg. Med. Chem. 26 (2018) 2766–2773.

[129] S. Linker, S. Wang, B. Ries, T. Stadelmann, S. Riniker, *Passing the Barrier – How Computer Simulations Can Help to Understand and Improve the Passive Membrane Permeability of Cyclic Peptides* , CHIMIA 75 (2021) 518–521.

[130] K. Moehle, H. J. Hofmann, *Secondary Structure Formation in N-Substituted Peptides* , J. Pep. Res. 51 (1998) 19–28.

[131] F. Cipcigan, P. Smith, J. Crain, A. Hogner, L. D. Maria, A. Llinas, E. Ratkova, *Membrane Permeability in Cyclic Peptides is Modulated by Core Conformations* , J. Chem. Inf. Model. 61 (2020) 263–269.

[132] K. Moritsugu, K. Takeuchi, N. Kamiya, J. Higo, I. Yasumatsu, Y. Fukunishi, I. Fukuda, *Flexibility and Cell Permeability of Cyclic Ras-Inhibitor Peptides Revealed by the Coupled Nosé–Hoover Equation* , J. Chem. Inf. Model. 61 (2021) 1921–1930.

[133] S. Park, F. Khalili-Araghi, E. Tajkhorshid, K. Schulten, *Free Energy Calculation from Steered Molecular Dynamics Simulations Using Jarzynski's Equality* , J. Chem. Phys. 119 (2003) 3559–3566.

[134] Y. Sugita, A. Kitao, Y. Okamoto, *Multidimensional Replica-Exchange Method for Free-Energy Calculations* , J. Chem. Phys. 113 (2000) 6042–6051.

[135] 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 (2016) 1048–1059.

[136] 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 (2016) 1060–1068.

[137] 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 (2012) 890–903.

[138] M. M. Reif, P. H. Hünenberger, C. Oostenbrink, *New Interaction Parameters for Charged Amino Acid Side Chains in the GROMOS Force Field* , J. Chem. Theory. Comput. 8 (2012) 3705–3723.

[139] H. J. Berendsen, J. P. Postma, W. F. van Gunsteren, J. Hermans, *Interaction Models for Water in Relation to Protein Hydration* , in: Intermolecular Forces, Springer, 1981, pp. 331–342.

[140] 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 (1995) 5451–5459.

[141] 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 (2001) 1125–1136.

[142] J.-P. Ryckaert, G. Ciccotti, H. J. Berendsen, *Numerical Integration of the Cartesian Equations of Motion of a System*

with Constraints: Molecular Dynamics of n-Alkanes , J. Comput. Phys. 23 (1977) 327–341.

[143] E. Lindahl, B. Hess, D. Van Der Spoel, *GROMACS 3.0: A Package for Molecular Simulation and Trajectory Analysis* , Mol. Model. Annu. 7 (2001) 306–317.

[144] I. Marzuoli, C. Margreitter, F. Fraternali, *Lipid Head Group Parameterization for GROMOS 54A8: A Consistent Approach with Protein Force Field Description* , J. Chem. Theory Comput. 15 (2019) 5175–5193.

[145] M. Parrinello, A. Rahman, *Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method* , J. Applied Phys. 52 (1981) 7182–7190.

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

[147] B. Hess, H. Bekker, H. J. Berendsen, J. G. Fraaije, *LINCS: A Linear Constraint Solver for Molecular Simulations* , J. Comput. Chem. 18 (1997) 1463–1472.

[148] M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz, F. Noé, *PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models* , J. Chem. Theory Comput. 11 (2015) 5525–5542.

[149] R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane, V. S. Pande, *MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories* , Biophys. J. 109 (2015) 1528–1532.

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

[151] F. Sittel, G. Stock, *Robust Density-Based Clustering to Identify Metastable Conformational States of Proteins* , J. Chem. Theory Comput. 12 (2016) 2426–2435.

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

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

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

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

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

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

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

[159] F. Giordanetto, J. Kihlberg, *Macrocyclic Drugs and Clinical Candidates: What Can Medicinal Chemists Learn from Their Properties?* , J. Med. Chem. 57 (2014) 278–295.

[160] P. Ermert, *Design, Properties and Recent Application of Macrocycles in Medicinal Chemistry* , CHIMIA 71 (2017) 678–702.

[161] A. A. Rzepiela, L. A. Viarengo-Baker, V. Tatarskii, R. Kombarov, A. Whitty, *Conformational Effects on the Passive Membrane Permeability of Synthetic Macrocycles* , J. Med. Chem. 65 (2022) 10300–10317.

[162] H. Wang, R. S. Dawber, P. Zhang, M. Walko, A. J. Wilson, X. Wang, *Peptide-Based Inhibitors of Protein–Protein Interactions: Biophysical, Structural and Cellular Consequences of Introducing a Constraint* , Chem. Sci. 12 (2021) 5977–5993.

[163] J. P. Gray, M. N. Uddin, R. Chaudhari, M. N. Sutton, H. Yang, P. Rask, H. Locke, B. J. Engel, N. Batistatou, J. Wang, B. J. Grindel, P. Bhattacharya, S. T. Gammon, S. Zhang, D. Piwnica-Worms, J. A. Kritzer, Z. Lu, R. C. Bast, S. W. Millward, *Directed Evolution of Cyclic Peptides for Inhibition of Autophagy* , Chem. Sci. 12 (2021) 3526–3543.

[164] S. D. Krämer, H. E. Aschmann, M. Hatibovic, K. F. Hermann, C. S. Neuhaus, C. Brunner, S. Belli, *When Barriers Ignore the "Rule-of-Five"* , Adv. Drug Deliv. Rev. 101 (2016) 62–74.

[165] K. Sugano, M. Kansy, P. Artursson, A. Avdeef, S. Bendels, L. Di, G. F. Ecker, B. Faller, H. Fischer, G. Gerebtzoff, H. Lennernaes, F. Senner, *Coexistence of Passive and*

Carrier-Mediated Processes in Drug Transport , Nat. Rev. Drug Discov. 9 (2010) 597–614.

[166] A. T. Bockus, K. W. Lexa, C. R. Pye, A. S. Kalgutkar, J. W. Gardner, K. C. Hund, W. M. Hewitt, J. A. Schwochert, E. Glassey, D. A. Price, A. M. Mathiowetz, S. Liras, M. P. Jacobson, S. Lokey, *Probing the Physicochemical Boundaries of Cell Permeability and Oral Bioavailability in Lipophilic Macrocycles Inspired by Natural Products* , J. Med. Chem. 58 (2015) 4581–4589.

[167] T.-X. Xiang, B. Anderson, *The Relationship Between Permeant Size and Permeability in Lipid Bilayer Membranes* , J. Membr. Biol. 140 (1994) 111–122.

[168] D. F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward, K. D. Kopple, *Molecular Properties That Influence the Oral Bioavailability of Drug Candidates* , J. Med. Chem. 45 (2002) 2615–2623.

[169] A. A. Golosov, A. N. Flyer, J. Amin, C. Babu, C. Gampe, J. Li, E. Liu, K. Nakajima, D. Nettleton, T. J. Patel, P. C. Reid, L. Yang, L. G. Monovich, *Design of Thioether Cyclic Peptide Scaffolds with Passive Permeability and Oral Exposure* , J. Med. Chem. 64 (2021) 2622–2633.

[170] D. Lee, S. Lee, J. Choi, Y.-K. Song, M. J. Kim, D.-S. Shin, M. A. Bae, Y.-C. Kim, C.-J. Park, K.-R. Lee, J.-H. Choi, J. Seo, *Interplay Among Conformation, Intramolecular Hydrogen Bonds, and Chameleonicity in the Membrane Permeability and Cyclophilin A Binding of Macrocyclic Peptide Cyclosporin O Derivatives* , J. Med. Chem. 64 (2021) 8272–8286.

[171] J. Johansen-Leete, S. Ullrich, S. E. Fry, R. Frkic, M. J. Bedding, A. Aggarwal, A. S. Ashhurst, K. B. Ekanayake, M. C. Mahawaththa, V. M. Sasi, S. Luedtke, D. J. Ford, A. J. O'Donoghue, T. Passioura, M. Larance, G. Otting, S. Turville, C. J. Jackson, C. Nitsche, R. J. Payne, *Antiviral Cyclic Peptides Targeting the Main Protease of SARS-CoV-2* , Chem. Sci. 13 (2022) 3826–3836.

[172] B. Bhushan, D. Granata, C. S. Kaas, M. A. Kasimova, Q. Ren, C. N. Cramer, M. D. White, A. M. K. Hansen, C. Fledelius, G. Invernizzi, K. Deibler, O. D. Coleman, X. Zhao, X. Qu, H. Liu, S. S. Zurmühl, J. T. Kodra, A. Kawamura, M. Münzel, *An Integrated Platform Approach Enables Discovery of Potent, Selective and Ligand-Competitive Cyclic Peptides Targeting the GIP Receptor* , Chem. Sci. 13 (2022) 3256–3262.

[173] S. B. Rafi, B. R. Hearn, P. Vedantham, M. P. Jacobson, A. R. Renslo, *Predicting and Improving the Membrane Permeability of Peptidic Small Molecules* , J. Med. Chem. 55 (2012) 3163–3169.

[174] W. M. Hewitt, S. S. 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 (2015) 715–721.

[175] A. T. Bockus, J. A. Schwochert, C. R. Pye, C. E. Townsend, V. Sok, M. A. Bednarek, R. S. Lokey, *Going Out on a Limb: Delineating The Effects of β-Branching, N-Methylation, and Side Chain Size on the Passive Permeability, Solubility, and Flexibility of Sanguinamide A Analogues* , J. Med. Chem. 58 (2015) 7409–7418.

[176] P. Thansandote, R. M. Harris, H. L. Dexter, G. L. Simpson, S. Pal, R. J. Upton, K. Valko, *Improving the Passive Permeability of Macrocyclic Peptides: Balancing Permeability with other Physicochemical Properties* , Bioorg. Med. Chem. 23 (2015) 322–327.

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

[178] 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. Leung, A. S. Kalgutkar, J. N. Bauman, Y. Zhang, S. Liras, D. A. Price, A. M. Mathiowetz, M. P. Jacobson, R. S. Lokey, *On-Resin N-Methylation of Cyclic Peptides for Discovery of Orally Bioavailable Scaffolds* , Nat. Chem. Biol. 7 (2011) 810–817.

[179] J. Schwochert, Y. Lao, C. R. Pye, M. R. Naylor, P. V. Desai, I. C. Gonzalez Valcarcel, J. A. Barrett, G. Sawada, M.-J. Blanco, R. S. Lokey, *Stereochemistry Balances Cell Permeability and Solubility in the Naturally Derived Phepropeptin Cyclic Peptides* , ACS Med. Chem. Lett. 7 (2016) 757–761.

[180] S. Kannan, P. G. A. Aronica, S. Ng, D. T. Gek Lian, Y. Frosi, S. Chee, J. Shimin, T. Y. Yuen, A. Sadruddin, H. Y. K. Kaan, A. Chandramohan, J. H. Wong, Y. S. Tan, Z. W. Chang, F. J. Ferrer-Gago, P. Arumugam, Y. Han, S. Chen, L. Rénia, C. J. Brown, C. W. Johannes, B. Henry, D. P. Lane, T. K. Sawyer, C. S. Verma, A. W. Partridge, *Macrocyclization of an All-D Linear α-Helical Peptide Im-*

parts Cellular Permeability , Chem. Sci. 11 (2020) 5577–5591.

[181] Y. Tian, X. Zeng, J. Li, Y. Jiang, H. Zhao, D. Wang, X. Huang, Z. Li, *Achieving Enhanced Cell Penetration of Short Conformationally Constrained Peptides Through Amphiphilicity Tuning* , Chem. Sci. 8 (2017) 7576–7581.

[182] A. C. Rand, S. S. Leung, H. Eng, C. J. Rotter, R. Sharma, A. S. Kalgutkar, Y. Zhang, M. V. Varma, K. A. Farley, B. Khunte, C. Limberakis, D. A. Price, S. Liras, A. M. Mathiowetz, M. P. Jacobson, R. S. Lokey, *Optimizing PK Properties of Cyclic Peptides: The Effect of Side Chain Substitutions on Permeability and Clearance* , MedChemComm 3 (2012) 1282–1289.

[183] B. C. Doak, B. Over, F. Giordanetto, J. Kihlberg, *Oral Druggable Space beyond the Rule of 5: Insights from Drugs and Clinical Candidates* , Chem. Biol. 21 (2014) 1115–1142.

[184] L. A. Viarengo-Baker, L. E. Brown, A. A. Rzepiela, A. Whitty, *Defining and Navigating Macrocycle Chemical Space* , Chem. Sci. 12 (2021) 4309–4328.

[185] A. S. Kamenik, S. M. Linker, S. Riniker, *Matching Simulations and Experiments of Conformationally Flexible Cyclic Peptides: Steps toward a Holistic View of Passive Membrane Permeability* , in: *Approaching the Next Inflection in Peptide Therapeutics: Attaining Cell Permeability and Oral Bioavailability* , ACS Symp. Ser., 2022, Ch. 5, pp. 137–154.

[186] B. J. Williams-Noonan, M. N. Speer, T. C. Le, M. M. Sadek, P. E. Thompson, R. S. Norton, E. Yuriev, N. Barlow, D. K. Chalmers, I. Yarovsky, *Membrane Permeating Macrocycles:*

Design Guidelines from Machine Learning , J. Chem. Inf. Model. 62 (2022) 4605–4619.

[187] A. Furukawa, J. Schwochert, C. R. Pye, D. Asano, Q. D. Edmondson, A. C. Turmon, V. G. Klein, S. Ono, O. Okada, R. S. Lokey, *Drug-Like Properties in Macrocycles above MW 1000: Backbone Rigidity versus Side-Chain Lipophilicity* , Angew. Chem. Int. Ed. 59 (2020) 21571–21577.

[188] M. Bermejo, A. Avdeef, A. Ruiz, R. Nalda, J. A. Ruell, O. Tsinman, I. González, C. Fernández, G. Sánchez, T. M. Garrigues, V. Merino, *PAMPA—a Drug Absorption In Vitro Model: 7. Comparing Rat In Situ, Caco-2, and PAMPA Permeability of Fluoroquinolones* , Eur. J. Pharm. Sci. 21 (2004) 429–441.

[189] M. Sugita, T. Fujie, K. Yanagisawa, M. Ohue, Y. Akiyama, *Lipid Composition Is Critical for Accurate Membrane Permeability Prediction of Cyclic Peptides by Molecular Dynamics Simulations* , J. Chem. Inf. Model. 62 (2022) 4549–4560.

[190] R. Gautier, A. Bacle, M. L. Tiberti, P. F. Fuchs, S. Vanni, B. Antonny, PackMem: A Versatile Tool to Compute and Visualize Interfacial Packing Defects in Lipid Bilayers, Biophys. J. 115 (2018) 436–444.

[191] S. Chakraborty, M. Doktorova, T. R. Molugu, F. A. Heberle, H. L. Scott, B. Dzikovski, M. Nagao, L.-R. Stingaciu, R. F. Standaert, F. N. Barrera, J. Katsaras, G. Khelashvili, M. F. Brown, R. Ashkar, How Cholesterol Stiffens Unsaturated Lipid Membranes, Proc. Natl. Acad. Sci. U. S. A. 117 (2020) 21896–21905.

[192] E. Corvera, O. G. Mouritsen, M. A. Singer, M. J. Zuckermann, The Permeability and the Effect of Acyl-Chain Length for Phospholipid Bilayers Containing Cholesterol: Theory and Experiment, Biochim. Biophys. Acta 1107 (1992) 261–270.

[193] A. S. Kamenik, S. M. Linker, S. Riniker, *Enhanced Sampling Without Borders: On Global Biasing Functions and How to Reweight Them* , Phys. Chem. Chem. Phys. 24 (2022) 1225–1236.

[194] J. Morstein, A. Capecchi, K. Hinnah, B. Park, J. Petit-Jacques, R. C. Van Lehn, J.-L. Reymond, D. Trauner, *Medium-Chain Lipid Conjugation Facilitates Cell-Permeability and Bioactivity* , J. Am. Chem. Soc. 144 (2022) 18532–18544.

[195] T. E. Cheatham III, P. A. Kollman, *Molecular Dynamics Simulation of Nucleic Acids* , Annu. Rev. Phys. Chem. 51 (2000) 435–471.

[196] T. Hansson, C. Oostenbrink, W. van Gunsteren, *Molecular Dynamics Simulations* , Curr. Opin. Struct. Biol. 12 (2002) 190–196.

[197] M. Karplus, J. Kuriyan, *Molecular Dynamics and Protein Function* , Proceed. Nat. Acad. Sci. USA 102 (2005) 6679.

[198] Y. Dong, Q. Li, A. Martini, *Molecular Dynamics Simulation of Atomic Friction: A Review and Guide* , J. Vac. Sci. Technol. A 31 (2013) 030801.

[199] K. Lindorff-Larsen, P. Maragakis, S. Piana, D. E. Shaw, *Picosecond to Millisecond Structural Dynamics in Human Ubiquitin* , J. Phys. Chem. B 120 (2016) 8313–8320.

[200] G. M. Torrie, J. P. Valleau, *Nonphysical Sampling Distributions in Monte Carlo Free-energy Estimation: Umbrella Sampling* , J. Comput. Phys. 23 (1977) 187–199.

[201] P. Virnau, M. Müller, *Calculation of Free Energy Through Successive Umbrella Sampling* , J. Chem. Phys. 120 (2004) 10925.

[202] U. H. E. Hansmann, *Parallel Tempering Algorithm for Conformational Studies of Biological Molecules* , Chem. Phys. Lett. 281 (1997) 140–150.

[203] Y. Sugita, Y. Okamoto, *Replica-exchange Multicanonical Algorithm and Multicanonical Replica-exchange Method for Simulating Systems with Rough Energy Landscape* , Chem. Phys. Lett. 329 (2000) 261–270.

[204] A. F. Voter, *Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events* , Phys. Rev. Lett. 78 (1997) 3908–3911.

[205] D. Hamelberg, J. Mongan, J. A. McCammon, *Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules* , J. Chem. Phys. 120 (2004) 11919.

[206] H. Grubmüller, *Predicting Slow Structural Transitions in Macromolecular Systems: Conformational Flooding* , Phys. Rev. E 52 (1995) 2893–2906.

[207] R. W. Zwanzig, *High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases* , J. Chem. Phys. 22 (1954) 1420–1426.

[208] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. K. Kollman, *The Weighted Histogram Analysis*

Method for Free-Energy Calculations on Biomolecules. I. The Method , J. Comput. Chem. 13 (1992) 1011–1021.

[209] M. Leitgeb, C. Schröder, S. Boresch, *Alchemical Free Energy Calculations and Multiple Conformational Substates* , J. Chem. Phys. 122 (2005) 084109.

[210] M. R. Shirts, J. D. Chodera, *Statistically Optimal Analysis of Samples from Multiple Equilibrium States* , J. Chem. Phys. 129 (2008) 124105.

[211] G. König, S. Boresch, *Non-Boltzmann Sampling and Bennett's Acceptance Ratio Method: How to Profit from Bending the Rules* , J. Comput. Chem. 32 (2011) 1082–1090.

[212] J. D. Chodera, W. C. Swope, F. Noé, J. H. Prinz, M. R. Shirts, V. S. Pande, *Dynamical Reweighting: Improved Estimates of Dynamical Properties From Simulations at Multiple Temperatures* , J. Chem. Phys. 134 (2011) 244107.

[213] L. Donati, C. Hartmann, B. G. Keller, *Girsanov Reweighting for Path Ensembles and Markov State Models* , J. Chem. Phys. 146 (2017) 244112.

[214] S. Kieninger, L. Donati, B. G. Keller, *Dynamical Reweighting Methods for Markov Models* , Curr. Opin. Struct. Biol. 61 (2020) 124–131.

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

[216] W. Wang, S. Cao, L. Zhu, X. Huang, *Constructing Markov State Models to Elucidate the Functional Conformational Changes of Complex Biomolecules* , WIREs Comput. Mol. Sci. 8 (2018) e1343.

[217] A. Kells, A. Annibale, E. Rosta, *Limiting Relaxation Times from Markov State Models* , J. Chem. Phys. 149 (2018) 072324.

[218] S. M. Linker, A. Magarkar, J. Köfinger, G. Hummer, D. Seeliger, *Fragment Binding Pose Predictions Using Unbiased Simulations and Markov-State Models* , J. Chem. Theory Comput. 15 (2019) 4974–4981.

[219] S. Marco, J.-H. Prinz, S. Christof, *Markov Model Theory* , in: B. G. R., V. S. Pande, N. Frank (Eds.), *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation* , Springer Netherlands, Dordrecht, 2014, p. 23.

[220] S. Winkelmann, C. Schütte, *The Spatiotemporal Master Equation: Approximation of Reaction-Diffusion Dynamics via Markov State Modeling* , J. Chem. Phys. 145 (2016) 214107.

[221] D. J. Bicout, A. Szabo, *Electron Transfer Reaction Dynamics in Non-Debye Solvents* , J. Chem. Phys. 109 (1998) 2325–2338.

[222] M. Badaoui, A. Kells, C. Molteni, C. J. Dickson, V. Hornak, E. Rosta, *Calculating Kinetic Rates and Membrane Permeability from Biased Simulations* , J. Phys. Chem. B 122 (2018) 11571.

[223] I. V. Girsanov, *On Transforming a Certain Class of Stochastic Processes by Absolutely Continuous Substitution of Measures* , Theory Probab. Its Appl. 5 (1960) 285–301.

[224] B. Øksendal, *Stochastic Differential Equations* , 6th Edition, Springer Verlag, Berlin, 2003.

[225] A. Brünger, C. L. Brooks, M. Karplus, *Stochastic Boundary Conditions for Molecular Dynamics Simulations of ST2 Water* , Chem. Phys. Lett. 105 (1984) 495–500.

[226] G. Zhang, T. Schlick, *Implicit Discretization Schemes for Langevin Dynamics* , Mol. Phys. 84 (1995) 1077–1098.

[227] A. Barducci, G. Bussi, M. Parrinello, *Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method* , Phys. Rev. Lett. 100 (2008) 020603.

[228] W. F. V. Gunsteren, H. J. C. Berendsen, *A Leap-frog Algorithm for Stochastic Dynamics* , Mol. Simul. 1 (1988) 173–185.

[229] N. Goga, A. J. Rzepiela, A. H. de Vries, S. J. Marrink, H. J. C. Berendsen, *Efficient Algorithms for Langevin and DPD Dynamics* , J. Chem. Theory Comput. 8 (2012) 3637–3649.

[230] A. Rahman, *Correlations in the Motion of Atoms in Liquid Argon* , Phys. Rev. 136 (1964) 405–411.

[231] G. Hummer, *Position-Dependent Diffusion Coefficients and Free Energies from Bayesian Analysis of Equilibrium and Replica Molecular Dynamics Simulations* , New J. Phys. 7 (2005) 34.

[232] M. Hinczewski, Y. von Hansen, J. Dzubiella, R. R. Netz, *How the Diffusivity Profile Reduces the Arbitrariness of Protein Folding Free Energies* , J. Chem. Phys. 132 (2010) 245103.

[233] E. Marsault, M. L. Peterson, *Practical Medicinal Chemistry with Macrocycles: Design, Synthesis, and Case Studies*, 1st Edition, John Wiley & Sons, Inc., 2017.

[234] M. A. Abdalla, L. J. McGaw, *Natural Cyclic Peptides as an Attractive Modality for Therapeutics: A Mini Review*, Molecules. 23 (2018) 2080.

[235] P. Chène, *Drugs Targeting Protein-Protein Interactions*, ChemMedChem. 1 (2006) 400–411.

[236] D. E. Scott, A. R. Bayly, C. Abell, J. Skidmore, *Small Molecules, Big Targets: Drug Discovery Faces the Protein-Protein Interaction Challenge*, Nat. Rev. Drug. Discov. 15 (2016) 533–550.

[237] A. E. Modell, S. L. Blosser, P. S. Arora, *Systematic Targeting of Protein-Protein Interactions Approaches to Targeting Protein-Protein Interactions*, Trends. Pharmacol. Sci. 37 (2016) 702–713.

[238] G. Kolata, *FDA Speeds Approval of Cyclosporin* , Science 221 (1983) 1273.

[239] S. Hyung, X. Feng, Y. Che, J. Stroh, M. Shapiro, *Detection of Conformation Types of Cyclosporin Retaining Intramolecular Hydrogen Bonds by Mass Spectrometry* , Anal. Bioanal. Chem. 406 (2014) 5785—5794.

[240] P. Artursson, J. Karlsson, *Correlation Between Oral-Drug Absorption in Humans and Apparent Drug Permeability Coefficients in Human Intestinal Epithelial (Caco-2) Cells* , Biochem. Biophys. Res. Commun. 175 (1991) 880–885.

[241] J. D. Irvine, L. Takahashi, K. Lockhart, J. Cheong, J. W. Tolan, H. E. Selick, J. R. Grove, *Mdck (Madin-Darby Canine Kidney) Cells: A Tool for Membrane Permeability Screening* , J. Pharm. Sci. 88 (1999) 28–33.

[242] S. M. Linker, C. Schellhaas, A. s. Kamenik, H.-J. Roth, M. Fouché, S. Rodde, S. Riniker, *Lessons for Oral Bioavailability: How Conformationally Flexible Cyclic Peptides Enter and Cross Lipid Membranes* , ChemRxiv (2022). doi:10.26434/chemrxiv-2022-9b47d.

[243] K. Kajitani, M. Fujihashi, Y. Kobayashi, S. Shimizu, Y. Tsujimoto, K. Miki, *Crystal Structure of Human Cyclophilin D in Complex With its Inhibitor, Cyclosporin A at 0.96-Å Resolution* , Proteins 70 (2008) 1635–1639.

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

[245] Schrödinger, *The PyMOL Molecular Graphics System, Version 1.8* (November 2015).

[246] S. Yang, S. S. Cho, Y. Levy, M. S. Cheung, H. Levine, P. G. Wolynes, J. N. Onuchic, *Domain Swapping is a Consequence of Minimal Frustration* , Proc. Natl. Acad. Sci. 101 (2004) 13786–13791.

[247] H. S. Hansen, P. H. Hünenberger, *Using the Local Elevation Method to Construct Optimized Umbrella Sampling Potentials: Calculation of the Relative Free Energies and Interconversion Barriers of Glucopyranose Ring Conformers in Water* , J. Comput. Chem. 31 (2010) 1–23.

Stephanie M. Linker

PhD Student in Computational Chemistry

17.07.1993

German

Stephanie Linker

Publication List

Languages

- German: Native Speaker
- English: CEFR C1-C2
- Chinese: HSK 3-4
- Spanish: CEFR A1
- Latin: Latinum

Education

today – 11/2018 — **Eidgenössische Technische Hochschule Zurich - PhD Studies - Computational Chemistry** — CH
Focus: In-silico Drug Discovery, Macrocyclic Drugs, Molecular Dynamics, Drug Bioavailability, Kinetic Models
Coursework: Computer-Assisted Drug Design, Human Resource Management: Leading Teams, Microeconomics

04/2017 – 09/2014 — **Universität Frankfurt - MSc. Biochemistry (⌀ 1.0)** — DE
Focus: Computer Simulations, Structural Biology, Biophysical Chemistry

Max-Planck Institute for Biophysics - Master Thesis (1.0) — DE

11/2017 – 10/2012 — **Universität Frankfurt - BSc. (Bio-)Physics (⌀ 1.0)** — DE
Focus: Machine Learning, Biophysics, Single Cell Omics

European Bioinformatics Institute Cambridge - Bachelor Thesis (1.0) — UK

07/2014 – 10/2011 — **Universität Frankfurt - BSc. Biochemistry (⌀ 1.0)** — DE
Focus: Membrane Proteins, Kinetic Models

Max-Planck Institute for Biophysics - Bachelor Thesis (1.0) — DE

06/2011 – 09/2003 — **St. Lioba Gymnasium Bad Nauheim - Abitur (⌀ 1.0)** — DE
Focus: Biology and Physics

Working Experience

today – 11/2018 — **Eidgenössische Technische Hochschule Zurich Institute of Physical Chemistry - Research Assistant** — CH
Full responsibility for two research projects, Managing collaborations with industrial and academic stakeholders, Design and supervision of student projects, Authorship of multiple publications, Communicating science through presentations at international conferences and managing of the group's Twitter account, Maternity leave from November 2021 – March 2022

10/2018 – 03/2018 — **Eidgenössische Technische Hochschule Zurich Institute of Molecular Systems Biology - Research Assistant** — CH
Studying bacterial metabolomics using mass spectrometry, Simulating protein networks

02/2018 – 05/2017 — **European Bioinformatics Institute Cambridge - Research Assistant** — UK
Multiomics analysis of single cell sequencing data, Machine Learning

09/2016 – 05/2016 — **Boehringer Ingelheim - Research Intern, Computational Chemistry** — DE
Computational workflow design for drug discovery

Stephanie M. Linker

PhD Student in Computational Chemistry

Data Sciences

Methods

- Molecular dynamics simulations
- Convolutional neural networks
- Regression models
- Object oriented programming
- Network modeling
- Stochastic simulations
- Statistics
- Monte Carlo simulations

Languages and Programs

- Python
  - Pandas
  - MDTraj
  - Numpy
  - Keras
  - Scikit-learn
- R
- SQL
- Git version control
- Microsoft Office
- Pymol
- Gromacs
- openMM

Interests

- Piano & Harp
- Rowing
- Chinese Studies
- Escape- and Boardgames

Awards & Scholarships

today – 02/2020	PhD Scholarship of the German National Academic Foundation (GNAF)
today – 01/2020	Scholarship Fund of the Swiss Chemical Industry
08/2020	Swiss Chemical Society Fall Meeting Best oral presentation award - runner up
11/2018	Roche Data Science Hackathon Basel, 1st Place
02/2018	Universität Frankfurt Master prize Biochemistry
11/2017	German Chemical Society CIC-Prize for Computational Chemistry
11/2017 – 07/2013	Student Scholarship of the German National Academic Foundation
06/2013 – 11/2011	Deutschlandstipendium

Trainings

04/2022	**McKinsey Next Generation Women Leader** Leadership and networking training
03/2022 – 10/2021	**GNAF China Competence Center** Training and lectures to strengthen China competences of future leaders, Focus: Chinese health care and economic system
10/2021	**Pharmaakademie: Intensive GxP Training** GLP, GCP, GVP, GMP basics including an examination
04/2021 – 10/2020	**Feminno: Female Innovation and Career Development Program** Start-up training, Pitch training, Patenting laws, Business plan design
06/2019	**Merck Biopharma Innovation Cup, Finalist** Topic: Digitalization in the pharmaceutical industry, Invited to one-week entrepreneurial training at Merck

Committee work

today – 2019	**young Swiss Chemical Society - International delegate (2020–today), Secretary (2019)** Representing Switzerland at the European and global level, Growing the board from 8 members to over 40 active members, Organizing conferences, meetings and talks for the >1,000 members
today – 2018	**Society for Women in Natural Sciences - President (2020–2022), Board member** Leading a team of 12 inspirational women, Representing females of two ETH departments, Organize over 20 events for female career development and networking, Raise awareness for gender gaps, 3rd prize ETH diversity award 2021
2019 – 2018	**ETH PhD Lecture Series committee - Board member** Planing the visits of international renowned researchers to ETH, Managing a budget of 10,000 CHF