Introduction

This is a book about mathematics. At the same time it is a book about molecular dynamics. Typical books with this combination of topics try to give a mathematical framework for what is done in standard molecular dynamics simulations or consider the temporal discretization schemes used in molecular dynamics. This book will not. Instead it will be concerned with one of the key challenges in molecular dynamics: How can one analyze molecular function if the required simulation timescales are infeasible?

For readers who are not familiar with molecular dynamics a few words explaining this key challenge seem appropriate: Applications in modern biotechnology and molecular medicine require simulation of biomolecular systems in atomic representation with immense length and timescales that are *far* beyond the capacity of computer power currently available. The reason for this is that many of the processes that constitute molecular function are *rare event* processes appearing on timescales that are many orders of magnitude, say 10–15 orders of magnitude, longer than the typical time steps of the numerical simulation. As a consequence, there is an increasing need for reduced models of the dynamical behavior of molecular systems that still allow describing the relevant dynamical properties while at the same time being less complex and thus allowing simulation on longer timescales. However, whether such reduced descriptions reproduce the correct rare event statistics is a question that in principle can only be answered mathematically.

In most molecular systems the biologically interesting and computationally problematic rare events belong to so-called conformation changes. Conformations are *metastable sets* of the dynamical behavior of the molecule, that is, regions of the molecule's state space that are attractive for the dynamics in the sense that typical trajectories remain within such regions for long periods of time before exiting towards other metastable sets. This rough description may already be sufficient for indicating that any mathematical approach to efficient rare event simulation in molecular dynamics should address the issue of metastability of the dynamical behavior of molecular systems. Consequently, this book will be concerned with understanding the origin of metastability and with its mathematical description. This, however, is just the first step. We will see how to exploit the existence of metastable sets for constructing a reduced molecular dynamics model with good approximation properties on the long timescales. In this reduced model, the so-called Markov state model (MSM), the main metastable sets will form the states of a Markov chain or Markov jump process while the transition probabilities or transition rates of the Markov model are given by the transition statistics of the original

molecular dynamics process. Mathematically, the process of reducing the original molecular dynamics process to the MSM process will be a Galerkin discretization of the so-called transfer operator of the molecular dynamics process. Thus, the mathematics of *projected transfer operators* is central for our undertaking.

That is exactly what the book will focus on: metastability, MSMs, and the mathematics of transfer operators that allows the analysis of the relation between one and the other. The rough outline is as follows: We start with a more detailed introduction to the idea behind MSMs in Chapter 1. Next we will outline the different molecular dynamics models and the mathematical framework in which they all can be handled together (Chapter 2). Then we can turn to the mathematical concepts of metastability in Chapter 3 before we introduce transfer operators and the spectral approach to metastability (Chapter 4), and the construction of MSMs via projected transfer operators in Chapter 5. Last, we will investigate how to use an existing MSM to get information on the transition pathways that dominate the transitions from one metastable set into another (Chapter 6).

Recent years have seen an ever increasing publication activity on how to construct MSMs for very different molecular systems ranging from peptides to proteins, from RNA to DNA, and via molecular sensors to molecular aggregation (see the next section for references). Presently, several books are under preparation that will address these practical approaches in all algorithmic detail but not their mathematical background. Therefore, the present book concentrates on the latter.

The material presented in the following has resulted from more than 12 years of research on MSMs and metastability that started with the first publications by the first author and his coworkers in the last millennium; see [**15, 16, 95, 96**]. It has been compiled for mathematicians as well as practical computational scientists interested in modern molecular dynamics.

The book is predominantly addressed to graduate students and researchers wanting to get an overview of the theoretical background of MSMs and associated concepts in molecular dynamics. Therefore the mathematical technicalities have been limited to the level that the authors believe to be accessible for theory-interested nonmathematicians. The text includes many examples and helpful illustrations. Everyone interested in deeper insights into the matter will find comprehensive references to the relevant literature. The authors hope that the present book may help to bridge the still existing gap between mathematical research on molecular dynamics and its practical use for realistic molecular systems by providing readers with tools for performing more in-depth analysis of simulation and data-analysis methods in molecular dynamics.

Many extraordinarily talented people have contributed to the mathematics of metastable systems in relation to the transfer operator approach, its algorithmic realization, and its applications. We wish to acknowledge the efforts and excellent skill of the following mathematicians: Peter Deuflhard (who has contributed from the very beginnings in the last millennium till today), Eric Vanden-Eijnden, Marcus Weber, and Frank Noe. Their contributions have been invaluable. Furthermore, we want to thank the DFG Research Center MATHEON "Mathematics for

key technologie" and the Freie Universität Berlin for making this work possible by providing funding and the creative atmosphere we treasure so highly.