
Advances in Experimental Medicine and Biology

Volume 797

Editorial Board:

IRUN R. COHEN, *The Weizmann Institute of Science, Rehovot, Israel*

ABEL LAJTHA, *N.S. Kline Institute for Psychiatric Research, Orangeburg,
NY, USA*

JOHN D. LAMBRIS, *University of Pennsylvania, Philadelphia, PA, USA*

RODOLFO PAOLETTI, *University of Milan, Milan, Italy*

For further volumes:
www.springer.com/series/5584

Gregory R. Bowman • Vijay S. Pande •
Frank Noé
Editors

An Introduction
to Markov State
Models and
Their Application
to Long Timescale
Molecular Simulation

 Springer

Editors

Gregory R. Bowman
University of California
Berkeley, CA, USA

Frank Noé
Freie Universität Berlin
Berlin, Germany

Vijay S. Pande
Department of Chemistry
Stanford University
Stanford, CA, USA

ISSN 0065-2598

ISSN 2214-8019 (electronic)

Advances in Experimental Medicine and Biology

ISBN 978-94-007-7605-0

ISBN 978-94-007-7606-7 (eBook)

DOI 10.1007/978-94-007-7606-7

Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2013956358

© Springer Science+Business Media Dordrecht 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

1	Introduction and Overview of This Book	1
	Gregory R. Bowman, Vijay S. Pande, and Frank Noé	
2	An Overview and Practical Guide to Building Markov State Models	7
	Gregory R. Bowman	
3	Markov Model Theory	23
	Marco Sarich, Jan-Hendrik Prinz, and Christof Schütte	
4	Estimation and Validation of Markov Models	45
	Jan-Hendrik Prinz, John D. Chodera, and Frank Noé	
5	Uncertainty Estimation	61
	Frank Noé and John D. Chodera	
6	Analysis of Markov Models	75
	Frank Noé and Jan-Hendrik Prinz	
7	Transition Path Theory	91
	Eric Vanden-Eijnden	
8	Understanding Protein Folding Using Markov State Models .	101
	Vijay S. Pande	
9	Understanding Molecular Recognition by Kinetic Network Models Constructed from Molecular Dynamics Simulations .	107
	Xuhui Huang and Gianni De Fabritiis	
10	Markov State and Diffusive Stochastic Models in Electron Spin Resonance	115
	Deniz Sezer and Benoît Roux	
11	Software for Building Markov State Models	139
	Gregory R. Bowman and Frank Noé	

Contributors

Gregory R. Bowman Departments of Molecular & Cell Biology and Chemistry, University of California, Berkeley, CA, USA; University of California, Berkeley, USA

John D. Chodera Memorial Sloan-Kettering Cancer Center, New York, NY, USA

Gianni De Fabritiis Computational Biophysics Laboratory (GRIB-IMIM), Universitat Pompeu Fabra, Barcelona Biomedical Research Park (PRBB), Barcelona, Spain

Xuhui Huang Department of Chemistry, Division of Biomedical Engineering, Center of Systems Biology and Human Health, Institute for Advance Study, The Hong Kong University of Science and Technology, Kowloon, Hong Kong

Frank Noé Freie Universität Berlin, Berlin, Germany

Vijay S. Pande Department of Chemistry, Stanford University, Stanford, CA, USA; Stanford University, Stanford, CA, USA

Jan-Hendrik Prinz Freie Universität Berlin, Berlin, Germany

Benoît Roux Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, USA

Marco Sarich Freie Universität Berlin, Berlin, Germany

Christof Schütte Freie Universität Berlin, Berlin, Germany

Deniz Sezer Faculty of Engineering and Natural Sciences, Sabancı University, Istanbul, Turkey

Eric Vanden-Eijnden Courant Institute, New York University, New York, NY, USA

Acronyms

ESR	Electron spin resonance
MD	Molecular dynamics (simulation)
MSM	Markov state model
PCCA	Perron cluster cluster analysis
TPT	Transition path theory
TPS	Transition path sampling

Mathematical Symbols

$\mathbf{T}(\tau)$	A transition probability matrix (row-stochastic) in $\mathbb{R}^{n \times n}$ describing the probabilities of hopping amongst a discrete set of states. The elements $T_{ij}(\tau)$ give the probability of an $i \rightarrow j$ transition during a time interval τ .
$\hat{\mathbf{T}}(\tau)$	An estimate of $\mathbf{T}(\tau)$ from trajectory data.
$\mathbf{C}(\tau)$	A transition count matrix (row-dominant) in $\mathbb{R}^{n \times n}$ describing the number of transitions observed amongst a discrete set of states. The elements $c_{ij}(\tau)$ count the number of $i \rightarrow j$ transitions observed, each of which occurred during a time interval τ .
τ	The time resolution (or lag time) of a model.
n	The number of discrete states.
$\mathbf{p}(t)$	A (column) vector in \mathbb{R}^n where the entry $\mathbf{p}_i(t)$ specifies the probability of being in state i at time t .
$\boldsymbol{\pi}$	A (column) vector in \mathbb{R}^n where the entry π_i specifies the equilibrium probability of being in state i .
λ_i	The i 'th largest eigenvalue of a transition probability matrix T . The largest eigenvalue is λ_1 and eigenvalues are ordered such that $1 = \lambda_1 > \lambda_2 > \lambda_3$.
$\boldsymbol{\psi}_i$	The i 'th right eigenvector of a transition probability matrix T in \mathbb{R}^n . The first right eigenvector is $\boldsymbol{\psi}_1$.
$\boldsymbol{\phi}_i$	The i 'th left eigenvector of a transition probability matrix T in \mathbb{R}^n . The first left eigenvector is $\boldsymbol{\phi}_1$.
χ_i	An indicator function for state i that is 1 within state i and 0 elsewhere. It may also refer to the degree of membership in state i .
θ_i	An experimental observable characteristic of state i .
q_i	The commitor probability for state i . That is, the probability of reaching some predefined set of final states from state i before reaching some predefined set of initial states.
Ω	A continuous state space (including positions and momenta).
$\mathbf{x}(t)$	A state in Ω (including positions and momenta) at time t .
$\mu(\mathbf{x})$	The stationary density of \mathbf{x} .
$p(\mathbf{x}, \mathbf{y}; \tau)$	The transition probability density to $\mathbf{y} \in \Omega$ after time τ given the system is in $\mathbf{x} \in \Omega$.

$\mathcal{T}(\tau)$	A transfer operator that propagates the continuous dynamics for a time τ .
m	The number of dominant eigenfunctions/eigenvalues considered.
S_1, \dots, S_n	Discrete sets which partition the state space Ω .
$\mu_i(\mathbf{x})$	The local stationary density restricted to a discrete state i .
$\langle f, g \rangle$	The scalar product $\langle f, g \rangle = \int f(\mathbf{x})g(\mathbf{x}) d\mathbf{x}$.
$\langle f, g \rangle_\mu$	The weighted scalar product $\langle f, g \rangle_\mu = \int \mu(\mathbf{x})f(\mathbf{x})g(\mathbf{x}) d\mathbf{x}$.