

METHODS IN MOLECULAR BIOLOGY

Series Editor

John M. Walker

School of Life and Medical Sciences

University of Hertfordshire

Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:

<http://www.springer.com/series/7651>

Computational Protein Design

Edited by

Ilan Samish

*Department of Plants and Environmental Sciences
Weizmann Institute of Science, Rehovot, Israel*

*Department of Biotechnology Engineering
Braude Academic College of Engineering, Karmiel, Israel*

Amai Proteins Ltd., Ashdod, Israel

Editor

Ilan Samish
Department of Plants and Environmental Sciences
Weizmann Institute of Science, Rehovot, Israel

Department of Biotechnology Engineering
Braude Academic College of Engineering, Karmiel, Israel

Amai Proteins Ltd., Ashdod, Israel

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-6635-6 ISBN 978-1-4939-6637-0 (eBook)
DOI 10.1007/978-1-4939-6637-0

Library of Congress Control Number: 2016959982

© Springer Science+Business Media New York 2017

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.

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.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Humana Press imprint is published by Springer Nature
The registered company is Springer Science+Business Media LLC
The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

Preface

The aim of this first-ever book entitled *Computational Protein Design* (CPD) is to bring the latest know-how on the CPD methods in respect to the process, success, and pitfalls of the field. The book is organized so as to introduce and present the general methodology and main challenges followed by a description of specific software and applications. As seen in the description below, there is more than one way to cluster the different chapters, each highlighting a different aspect of the field.

While there has not been a book dedicated to CPD, books on protein design have often included chapters on CPD. Here, following a chapter on the framework of CPD (Chapter 1) and a summary of past achievements and future challenges (Chapter 2), a chapter on the experimental aspects of production of the designed protein is presented (Chapter 3). Beyond the need to understand the experimental aspects of the computational endeavor, this is to remind us that the final outcome of the computational process is the production of a real protein.

It is widely considered that a global minimum energy conformation (GMEC) reflects the actual native structure of the protein. The protein design process is intrinsically computationally intensive as sequence and structure space should be rigorously sampled in the search for the GMEC of the requested target. Deterministic search methods (Chapter 4) of which dead-end elimination (DEE) is among the first to be used, are guaranteed to find the GMEC while stochastic methods are not guaranteed to find it. Other methods, e.g., the A* search algorithm, were optimized to run in parallel taking advantage of the graphic processing unit (GPU) processor infrastructure (Chapter 13). Complementarily, the CPD effort should consider the solvating milieu, e.g., via a geometric potential (Chapter 5). In addition, the residue-level core building block focus of CPD should be analyzed and predicted in respect to phylogenetic, structural, and energetic properties. These should be treated according to the immediate and possibly changing microenvironment, e.g., as in protein–protein complexes (Chapter 6). The GMEC considers a single minimum conformation and can be applied for the redesign of a given scaffold (Chapter 10), for requested functional motifs (Chapter 11) or for emphasizing specific types of available data, e.g., evolutionary information (Chapter 12). Yet, proteins within their native physiological surrounding are dynamic ensembles intrinsically requiring conformational dynamics. As such, it is important to *a priori* design the protein as a multistate entity (Chapter 7), a characteristic that can be introduced via integrating to the design process methods that analyze dynamics such as molecular dynamics (Chapter 8) or normal mode analysis (Chapter 9).

The computational design scheme can be tailored to specific types of proteins or domains, which in turn should be assessed as to their resemblance to the requested domain or specific designated characteristic. Examples include protein–protein interaction interfaces (Chapter 14), drug-resistance mutations (Chapter 15), symmetric proteins of identical sequence repeats (Chapter 16), self-assemblies exploiting synthetic amino acids (Chapter 17), oligomerized conformations of the defensins (Chapter 18), ligand-binding proteins (Chapter 19), proteins with reduced immunogenicity (Chapter 20), antibodies (Chapter 21), membrane curvature-sensing peptides (Chapter 22), and allosteric drug-binding sites within proteins (Chapter 23). Taken together, these application focus areas

present the breadth of the CPD field along with the intrinsic achievements and challenges upon examining the “devil” in the details of key examples.

The general field of protein design, let alone the computational aspect of it, is expected to present an exponential increase in quality and quantity alike. Such change is fostered by the need to expand protein space for understanding biology, for applying biotechnology, and for expanding pharmaceuticals from the common small molecules to biologics – specific and side-effect-free proteins. Importantly, while scientific research of proteins is often focused towards pharmaceutical applications, CPD presents the possibility to expand the use of proteins in food-tech and white biotechnology, namely, the use of proteins for industrial applications. In addition, the field is nurtured by the exponential increase in raw sequence and structure data, and the increase in cost-effect computational hardware in general and hardware tailored to protein application, in particular. Not less important is the careful feedback loop of quantitative parameterization sequence and fold space followed by software design that will efficiently test our parameterization and produce novel protein design, which in turn can be materialized and characterized experimentally.

Karmiel, Israel

Ilan Samish

Contents

| | |
|---------------------------|-----------|
| <i>Preface</i> | <i>v</i> |
| <i>Contributors</i> | <i>ix</i> |

PART I COMPUTATIONAL PROTEIN DESIGN

| | |
|--|-----|
| 1 The Framework of Computational Protein Design | 3 |
| <i>Ilan Samish</i> | |
| 2 Achievements and Challenges in Computational Protein Design | 21 |
| <i>Ilan Samish</i> | |
| 3 Production of Computationally Designed Small Soluble- and Membrane-Proteins: Cloning, Expression, and Purification | 95 |
| <i>Barsa Tripathy and Rudresh Acharya</i> | |
| 4 Deterministic Search Methods for Computational Protein Design | 107 |
| <i>Seydou Traoré, David Allouche, Isabelle André, Thomas Schiex, and Sophie Barbe</i> | |
| 5 Geometric Potentials for Computational Protein Sequence Design | 125 |
| <i>Jie Li and Patrice Koehl</i> | |
| 6 Modeling Binding Affinity of Pathological Mutations for Computational Protein Design | 139 |
| <i>Miguel Romero-Durana, Chiara Pallara, Fabian Glaser, and Juan Fernández-Recio</i> | |
| 7 Multistate Computational Protein Design with Backbone Ensembles | 161 |
| <i>James A. Davey and Roberto A. Chica</i> | |
| 8 Integration of Molecular Dynamics Based Predictions into the Optimization of De Novo Protein Designs: Limitations and Benefits | 181 |
| <i>Henrique F. Carvalho, Arménio J.M. Barbosa, Ana C.A. Roque, Olga Iranzo, and Ricardo J.F. Branco</i> | |
| 9 Applications of Normal Mode Analysis Methods in Computational Protein Design | 203 |
| <i>Vincent Frappier, Matthieu Chartier, and Rafael Najmanovich</i> | |

PART II SOFTWARE OF COMPUTATIONAL PROTEIN DESIGN APPLICATIONS

| | |
|--|-----|
| 10 Computational Protein Design Under a Given Backbone Structure with the ABACUS Statistical Energy Function | 217 |
| <i>Peng Xiong, Quan Chen, and Haiyan Liu</i> | |
| 11 Computational Protein Design Through Grafting and Stabilization | 227 |
| <i>Cheng Zhu, David D. Mowrey, and Nikolay V. Dokholyan</i> | |
| 12 An Evolution-Based Approach to De Novo Protein Design | 243 |
| <i>Jeffrey R. Brender, David Shultis, Naureen Aslam Khattak, and Yang Zhang</i> | |

| | | |
|---|--|-----|
| 13 | Parallel Computational Protein Design | 265 |
| | <i>Yichao Zhou, Bruce R. Donald, and Jianyang Zeng</i> | |
| 14 | BindML/BindML+: Detecting Protein-Protein Interaction Interface Propensity from Amino Acid Substitution Patterns | 279 |
| | <i>Qing Wei, David La, and Daisuke Kihara</i> | |
| 15 | OSPREY Predicts Resistance Mutations Using Positive and Negative Computational Protein Design..... | 291 |
| | <i>Adegoke Ojewole, Anna Lowegard, Pablo Gainza, Stephanie M. Reeve, Ivelin Georgiev, Amy C. Anderson, and Bruce R. Donald</i> | |
| PART III COMPUTATIONAL PROTEIN DESIGN OF SPECIFIC TARGETS | | |
| 16 | Evolution-Inspired Computational Design of Symmetric Proteins | 309 |
| | <i>Arnout R.D. Voet, David Simoncini, Jeremy R.H. Tame, and Kam Y.J. Zhang</i> | |
| 17 | A Protocol for the Design of Protein and Peptide Nanostructure Self-Assemblies Exploiting Synthetic Amino Acids | 323 |
| | <i>Nurit Haspel, Jie Zheng, Carlos Aleman, David Zanuy, and Ruth Nussinov</i> | |
| 18 | Probing Oligomerized Conformations of Defensin in the Membrane | 353 |
| | <i>Wenxun Gan, Dina Schneidman, Ning Zhang, Buyong Ma, and Ruth Nussinov</i> | |
| 19 | Computational Design of Ligand Binding Proteins | 363 |
| | <i>Christine E. Tinberg and Sagar D. Khare</i> | |
| 20 | EpiSweep: Computationally Driven Reengineering of Therapeutic Proteins to Reduce Immunogenicity While Maintaining Function | 375 |
| | <i>Yoonjoo Choi, Deeptak Verma, Karl E. Griswold, and Chris Bailey-Kellogg</i> | |
| 21 | Computational Tools for Aiding Rational Antibody Design | 399 |
| | <i>Konrad Krawczyk, James Dunbar, and Charlotte M. Deane</i> | |
| 22 | Computational Design of Membrane Curvature-Sensing Peptides | 417 |
| | <i>Armando Jerome de Jesus and Hang Yin</i> | |
| 23 | Computational Tools for Allosteric Drug Discovery: Site Identification and Focus Library Design | 439 |
| | <i>Wenkang Huang, Ruth Nussinov, and Jian Zhang</i> | |
| | <i>Index</i> | 447 |

Contributors

- RUDRESH ACHARYA • *School of Biological Sciences, National Institute of Science Education and Research, Odisha, India; Homi Bhabha National Institute, Training School Complex, Anushakti Nagar, Mumbai, India*
- CARLOS ALEMAN • *Departament d'Enginyeria Química, E. T. S. d'Enginyeria Industrial de Barcelona, Universitat Politècnica de Catalunya, Barcelona, Spain; Center for Research in Nano-Engineering, Universitat Politècnica de Catalunya, Barcelona, Spain*
- DAVID ALLOUCHE • *Unité de Mathématiques et Informatique Appliquées de Toulouse, Castanet Tolosan, France*
- AMY C. ANDERSON • *Department of Pharmaceutical Sciences, University of Connecticut, Storrs, CT, USA*
- ISABELLE ANDRÉ • *INSA, UPS, INP, Université de Toulouse, Toulouse, France; Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, INRA, UMR792, Toulouse, France; CNRS, UMR5504, Toulouse, France*
- NAUREEN ASLAM • *Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA*
- CHRIS BAILEY-KELLOGG • *Department of Computer Science, Dartmouth College, Dartmouth, USA*
- SOPHIE BARBE • *INSA, UPS, INP, Université de Toulouse, Toulouse, France; Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, INRA, UMR792, Toulouse, France; CNRS, UMR5504, Toulouse, France*
- ARMÉNIO J. M. BARBOSA • *UCIBIO, REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal*
- RICARDO J. F. BRANCO • *UCIBIO, REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal*
- JEFFREY R. BRENDER • *Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA*
- HENRIQUE F. CARVALHO • *UCIBIO, REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal; Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Av. da República, Oeiras, Portugal*
- MATTHIEU CHARTIER • *Department of Biochemistry, Faculty of Medicine and Health Sciences, University of Sherbrooke, Sherbrooke, QC, Canada*
- QUAN CHEN • *School of Life Sciences, Hefei National Laboratory for Physical Sciences at the Microscales, and Collaborative Innovation Center of Chemistry for Life Sciences, University of Science and Technology of China, Hefei, Anhui, China*
- ROBERTO A. CHICA • *Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, ON, Canada*
- YOONJOO CHOI • *Department of Computer Science, Dartmouth College, Dartmouth, USA*
- JAMES A. DAVEY • *Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, ON, Canada*
- CHARLOTTE M DEANE • *Department of Statistics, University of Oxford, Oxford, UK*

- NIKOLAY V. DOKHOLYAN • *Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA*
- BRUCE R. DONALD • *Department of Computer Science, Duke University, Durham, NC, USA; Department of Biochemistry, Duke University, Durham, NC, USA; Department of Chemistry, Duke University, Durham, NC, USA*
- JAMES DUNBAR • *Department of Statistics, University of Oxford, Oxford, UK*
- JUAN FERNÁNDEZ-RECIO • *Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Barcelona, Spain*
- VINCENT FRAPPIER • *Department of Biochemistry, Massachusetts Institute of Technology, Cambridge, MA; Faculty of Medicine and Health Sciences, University of Sherbrooke, Sherbrooke, QC, Canada*
- PABLO GAINZA • *Department of Computer Science, Duke University, Durham, NC, USA*
- WENXUN GAN • *Research Center of Basic Medical Sciences and Cancer Institute and Hospital, Tianjin Medical University, Tianjin, China*
- IVELIN GEORGIEV • *Department of Computer Science, Duke University, Durham, NC, USA; Vaccine Research Center, National Institute of Allergy and Infectious Diseases, Bethesda, MD, USA*
- FABIAN GLASER • *Bioinformatics Knowledge Unit, The Lorry I. Lokey Interdisciplinary Center for Life Sciences and Engineering, Technion, Israel*
- KARL E. GRISWOLD • *Thayer School of Engineering, Dartmouth, USA*
- NURIT HASPEL • *Department of Computer Science, The University of Massachusetts Boston, Boston, MA, USA*
- WENKANG HUANG • *Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Shanghai JiaoTong University School of Medicine (SJTU-SM), Shanghai, China*
- OLGA IRANZO • *Aix Marseille Université, Centrale Marseille, CNRS, iSm2 UMR 7313, Marseille, France*
- ARMANDO J. DE JESUS • *Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO, USA; The BioFrontiers Institute, University of Colorado, Boulder, CO, USA*
- SAGAR D. KHARE • *Department of Chemistry and Chemical Biology, Center for Integrative Proteomics Research, Rutgers, The State University of New Jersey, Piscataway, NJ, USA*
- DAISUKE KIHARA • *Department of Biological Sciences, Purdue University, West Lafayette, IN, USA; Department of Computer Science, Purdue University, West Lafayette, IN, USA*
- PATRICE KOEHL • *Department of Computer Science and Genome Center, University of California, Davis, CA, USA*
- KONRAD KRAWCZYK • *Department of Statistics, University of Oxford, Oxford, UK*
- DAVID LA • *Department of Biochemistry, University of Washington, Seattle, WA, USA*
- JIE LI • *Computational and Systems Biology Group, Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore*
- HAIYAN LIU • *School of Life Sciences, Hefei National Laboratory for Physical Sciences at the Microscales, University of Science and Technology of China, Hefei, Anhui, China; Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei, Anhui, China*
- ANNA LOWEGARD • *Program in Computational Biology and Bioinformatics, Duke University, Durham, NC, USA*
- BUYONG MA • *Basic Science Program, Leidos Biomedical Research, Inc, Frederick, MD, USA; Cancer and Inflammation Program, National Cancer Institute, Frederick, MD, USA*
- DAVID D. MOWREY • *Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA*

- RAFAEL NAJMANOVICH • *Department of Pharmacology and Physiology, Faculty of Medicine, University of Montreal, Montreal, QC, Canada*
- RUTH NUSSINOV • *Department of Human Genetics, Sackler School of Medicine, Sackler Inst. of Molecular Medicine and Molecular Medicine, Tel Aviv University, Tel Aviv, Israel; Basic Science Program, Cancer and Inflammation Program, Leidos Biomedical Research, Inc., National Cancer Institute, Frederick, MD, USA*
- ADEGOKE OJEWOLE • *Program in Computational Biology and Bioinformatics, Duke University, Durham, NC, USA*
- CHIARA PALLARA • *Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Barcelona, Spain*
- STEPHANIE M. REEVE • *Department of Pharmaceutical Sciences, University of Connecticut, Storrs, CT, USA*
- MIGUEL ROMERO-DURANA • *Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center, Barcelona, Spain*
- ANA C. A. ROQUE • *UCIBIO, REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal*
- ILAN SAMISH • *Department of Plants and Environmental Sciences, Weizmann Institute of Science, Rehovot, Israel; Department of Biotechnology Engineering, Braude Academic College of Engineering, Karmiel, Israel; Amai Proteins Ltd., Ashdod, Israel; Dept of Biotechnology Engineering, Braude Academic College of Engineering, Karmiel, Israel; Amai Proteins Ltd., Ashdod, Israel*
- THOMAS SCHIEX • *Unité de Mathématiques et Informatique Appliquées de Toulouse, UR 875, INRA, Castanet Tolosan, France*
- DINA SCHNEIDMAN • *Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, USA*
- DAVID SHULTIS • *Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA*
- DAVID SIMONCINI • *Structural Bioinformatics Team, Division of Structural and Synthetic Biology, Center for Life Science Technologies, RIKEN, Yokohama, Kanagawa, Japan; MIAT, UR-875, INRA, Castanet Tolosan, France*
- JEREMY R. H. TAME • *Drug Design Laboratory, Graduate School of Medical Life Science, Yokohama City University, Yokohama, Kanagawa, Japan*
- CHRISTINE E. TINBERG • *Department of Biochemistry, University of Washington, Seattle, WA, USA*
- SEYDOU TRAORÉ • *INSA, UPS, INP, Université de Toulouse, Toulouse, France; INRA, UMR792, Université de Toulouse, Toulouse, France; Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés, CNRS, UMR5504, Toulouse, France*
- BARSA TRIPATHY • *School of Biological Sciences, National Institute of Science Education and Research, Odisha, India; Homi Bhabha National Institute, Training School Complex, Anushakti Nagar, Mumbai, India*
- DEEPTAK VERMA • *Department of Computer Science, Dartmouth College, Dartmouth, USA*
- ARNOUT R. D. VOET • *Structural Bioinformatics Team, Division of Structural and Synthetic Biology, Center for Life Science Technologies, RIKEN, Yokohama, Kanagawa, Japan*
- QING WEI • *Department of Computer Science, Purdue University, West Lafayette, IN, USA*
- PENG XIONG • *School of Life Sciences, Hefei National Laboratory for Physical Sciences at the Microscales, and Collaborative Innovation Center of Chemistry for Life Sciences, University of Science and Technology of China, Hefei, Anhui, China*

- HANG YIN • *Department of Chemistry and Biochemistry, The BioFrontiers Institute, University of Colorado, Boulder, CO, USA*
- DAVID ZANUY • *Departament d'Enginyeria Química, E. T. S. d'Enginyeria Industrial de Barcelona, Universitat Politècnica de Catalunya, Barcelona, Spain*
- JIANYANG ZENG • *Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, P. R. China*
- JIAN ZHANG • *Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, School of Medicine (SJTU-SM), Shanghai JiaoTong University, Shanghai, P. R. China*
- KAM Y. J. ZHANG • *Structural Bioinformatics Team, Division of Structural and Synthetic Biology, Center for Life Science Technologies, RIKEN, Yokohama, Kanagawa, Japan*
- NING ZHANG • *Research Center of Basic Medical Sciences and Cancer Institute and Hospital, Tianjin Medical University, Tianjin, China*
- YANG ZHANG • *Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA; Department of Biological Chemistry, University of Michigan, Ann Arbor, MI, USA*
- JIE ZHENG • *Department of Chemical and Biomolecular Engineering, The University of Akron, Akron, OH, USA*
- YICHAO ZHOU • *Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, P. R. China*
- CHENG ZHU • *Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA*