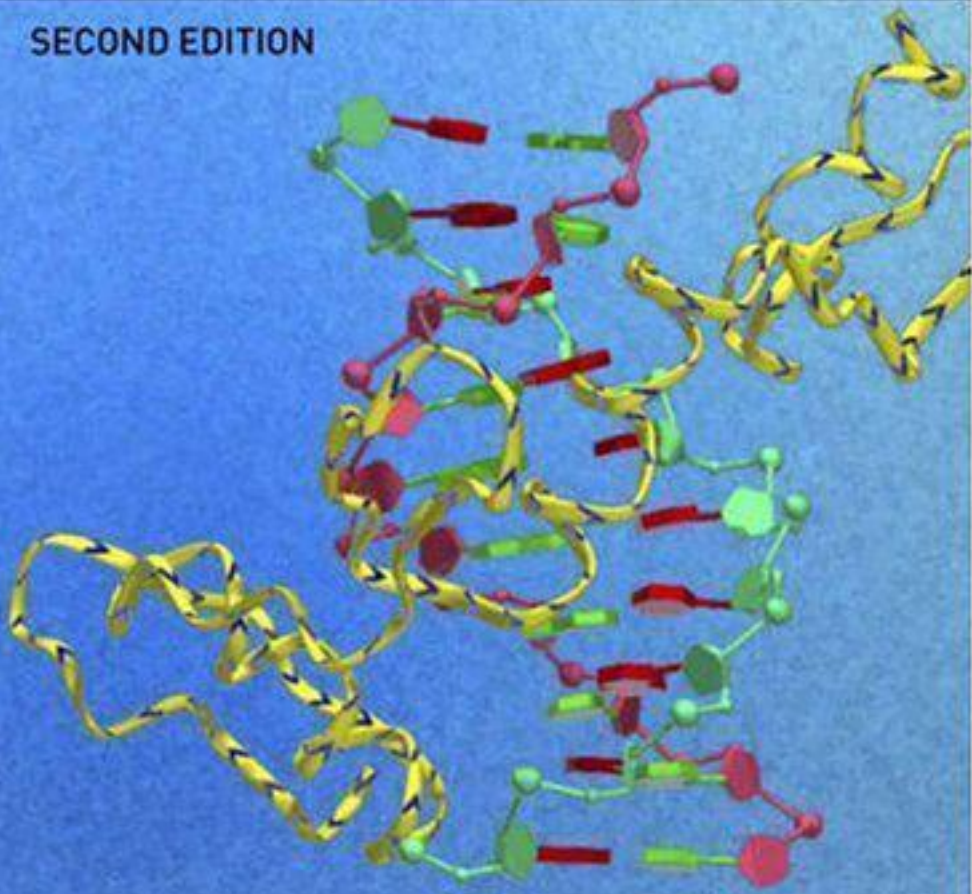
OXFORD

# INTRODUCTION TO BIOINFORMATICS

SECOND EDITION



ARTHUR M. LESK

# Introduction to
# **Bioinformatics**

SECOND EDITION

**Arthur M. Lesk**
*The Pennsylvania State University*

In nature's infinite book of secrecy
A little I can read.
– *Antony and Cleopatra*

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP
Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in
Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto
With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan South Korea Poland Portugal
Singapore Switzerland Thailand Turkey Ukraine Vietnam
Published in the United States
by Oxford University Press Inc., New York

First published 2005

British Library Cataloguing in Publication Data
Data available
Library of Congress Cataloging in Publication Data
Data available
ISBN 0 19 9277877 (hbk)
10 9 8 7 6 5 4 3 2 1
Typeset by Newgen Imaging Systems (P) Ltd., Chennai, India
Printed in Great Britain
on acid-free paper by Ashford Colour Press Limited, Gosport, Hampshire

*Dedicated to Eda, with whom I have merged my genes.*

# Contents

## Plan of the book

- Chapter 1 sets the stage and introduces all of the major players: DNA and protein sequences and structures, genomes and proteomes, databases and information retrieval, the World Wide Web, computer programming. Before developing individual topics in detail it is important to see the framework of their interactions.
- Chapter 2 presents the nature of individual genomes, including the Human Genome, and the relationships among them, from the biological point of view.
- Chapter 3 imparts basic skills in using the Web in bioinformatics. It describes archival databanks, and leads the reader through sample sessions involving information retrieval from some of the major archival databases in molecular biology.
- Chapter 4 treats the analysis of relationships among sequences—alignments and phylogenetic trees. These methods underlie some of the major computational challenges of bioinformatics: detecting distant relatives, understanding relationships among genomes of different organisms, and tracing the course of evolution at the species and molecular levels.
- Chapter 5 moves into three dimensions, treating protein structure and folding. Sequence and structure must be seen as full partners, with bioinformatics developing methods for moving back and forth between them as fluently as possible. Understanding protein structures in detail is essential for determining their mechanisms of action, and for clinical and pharmacological applications.
- Chapter 6 treats proteomics and systems biology, including new high-throughput sources of information about the expression and distribution of proteins in cells, and attempts to synthesize the information to reveal patterns of organization.