



Chapman & Hall/CRC  
Mathematical and Computational Biology Series

# Biological Computation

Ehud Lamm and Ron Unger



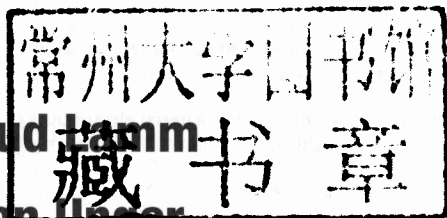
CRC Press  
Taylor & Francis Group

A CHAPMAN & HALL BOOK

Chapman & Hall/CRC Mathematical and Computational Biology Series

# Biological Computation

Ehud Lamm  
Ron Unger



**CRC Press**  
Taylor & Francis Group  
Boca Raton London New York

CRC Press is an imprint of the  
Taylor & Francis Group an **informa** business  
A CHAPMAN & HALL BOOK

Chapman & Hall/CRC  
Taylor & Francis Group  
6000 Broken Sound Parkway NW, Suite 300  
Boca Raton, FL 33487-2742

© 2011 by Taylor and Francis Group, LLC  
Chapman & Hall/CRC is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed in the United States of America on acid-free paper  
10 9 8 7 6 5 4 3 2 1

International Standard Book Number: 978-1-4200-8795-6 (Hardback)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access [www.copyright.com](http://www.copyright.com) (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

**Trademark Notice:** Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

---

Library of Congress Cataloging-in-Publication Data

---

Lamm, Ehud.  
Biological computation / authors, Ehud Lamm, Ron Unger.  
p. cm.  
Includes bibliographical references and index.  
ISBN 978-1-4200-8795-6 (hardcover : alk. paper)  
1. Biocomputers. I. Unger, Ron. II. Title.

QA76.884.L36 2010  
006.3'2--dc22

2010028310

---

Visit the Taylor & Francis Web site at  
<http://www.taylorandfrancis.com>

and the CRC Press Web site at  
<http://www.crcpress.com>

# **Biological Computation**

# CHAPMAN & HALL/CRC

## Mathematical and Computational Biology Series

### Aims and scope:

This series aims to capture new developments and summarize what is known over the entire spectrum of mathematical and computational biology and medicine. It seeks to encourage the integration of mathematical, statistical, and computational methods into biology by publishing a broad range of textbooks, reference works, and handbooks. The titles included in the series are meant to appeal to students, researchers, and professionals in the mathematical, statistical and computational sciences, fundamental biology and bioengineering, as well as interdisciplinary researchers involved in the field. The inclusion of concrete examples and applications, and programming techniques and examples, is highly encouraged.

### Series Editors

N. F. Britton

*Department of Mathematical Sciences  
University of Bath*

Xihong Lin

*Department of Biostatistics  
Harvard University*

Hershel M. Safer

Maria Victoria Schneider

*European Bioinformatics Institute*

Mona Singh

*Department of Computer Science  
Princeton University*

Anna Tramontano

*Department of Biochemical Sciences  
University of Rome La Sapienza*

Proposals for the series should be submitted to one of the series editors above or directly to:

**CRC Press, Taylor & Francis Group**

4th, Floor, Albert House

1-4 Singer Street

London EC2A 4BQ

UK

# Published Titles

## **Algorithms in Bioinformatics: A Practical Introduction**

*Wing-Kin Sung*

## **Bioinformatics: A Practical Approach**

*Shui Qing Ye*

## **Biological Computation**

*Ehud Lamm and Ron Unger*

## **Biological Sequence Analysis Using the SeqAn C++ Library**

*Andreas Gogol-Döring and Knut Reinert*

## **Cancer Modelling and Simulation**

*Luigi Preziosi*

## **Cancer Systems Biology**

*Edwin Wang*

## **Cell Mechanics: From Single Scale-Based Models to Multiscale Modeling**

*Arnaud Chauvière, Luigi Preziosi, and Claude Verdier*

## **Clustering in Bioinformatics and Drug Discovery**

*John D. MacCuish and Norah E. MacCuish*

## **Combinatorial Pattern Matching Algorithms in Computational Biology Using Perl and R**

*Gabriel Valiente*

## **Computational Biology: A Statistical Mechanics Perspective**

*Ralf Blossey*

## **Computational Hydrodynamics of Capsules and Biological Cells**

*C. Pozrikidis*

## **Computational Neuroscience: A Comprehensive Approach**

*Jianfeng Feng*

## **Data Analysis Tools for DNA Microarrays**

*Sorin Draghici*

## **Differential Equations and Mathematical Biology, Second Edition**

*D.S. Jones, M.J. Plank, and B.D. Sleeman*

## **Engineering Genetic Circuits**

*Chris J. Myers*

## **Exactly Solvable Models of Biological Invasion**

*Sergei V. Petrovskii and Bai-Lian Li*

## **Gene Expression Studies Using Affymetrix Microarrays**

*Hinrich Göhlmann and Willem Talloen*

## **Glycome Informatics: Methods and Applications**

*Kiyoko F. Aoki-Kinoshita*

## **Handbook of Hidden Markov Models in Bioinformatics**

*Martin Gollery*

## **Introduction to Bioinformatics**

*Anna Tramontano*

## **Introduction to Computational Proteomics**

*Golan Yona*

## **An Introduction to Systems Biology: Design Principles of Biological Circuits**

*Uri Alon*

## **Kinetic Modelling in Systems Biology**

*Oleg Demin and Igor Goryanin*

## **Knowledge Discovery in Proteomics**

*Igor Jurisica and Dennis Wigle*

## **Meta-analysis and Combining Information in Genetics and Genomics**

*Rudy Guerra and Darlene R. Goldstein*

## **Methods in Medical Informatics: Fundamentals of Healthcare Programming in Perl, Python, and Ruby**

*Jules J. Berman*

## **Modeling and Simulation of Capsules and Biological Cells**

*C. Pozrikidis*

## **Niche Modeling: Predictions from Statistical Distributions**

*David Stockwell*

## **Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems**

*Qiang Cui and Ivet Bahar*

## **Optimal Control Applied to Biological Models**

*Suzanne Lenhart and John T. Workman*

## **Pattern Discovery in Bioinformatics: Theory & Algorithms**

*Laxmi Parida*

## **Python for Bioinformatics**

*Sebastian Bassi*

## **Spatial Ecology**

*Stephen Cantrell, Chris Cosner, and Shigui Ruan*

## **Spatiotemporal Patterns in Ecology and Epidemiology: Theory, Models, and Simulation**

*Horst Malchow, Sergei V. Petrovskii, and Ezio Venturino*

## **Stochastic Modelling for Systems Biology**

*Darren J. Wilkinson*

## **Structural Bioinformatics: An Algorithmic Approach**

*Forbes J. Burkowski*

## **The Ten Most Wanted Solutions in Protein Bioinformatics**

*Anna Tramontano*

---

# Preface

---

## THE SPIRIT OF THIS BOOK

---

It is often said that biology is going to be the science of the 21<sup>st</sup> century as physics was the science of the 20<sup>th</sup>. Fascinating discoveries about the living world around us, as well as about our own bodies, are brought about daily by molecular biology, neuroscience, and other biological disciplines. In addition, biological understanding, whether on the molecular scale or on the ecological level, is fast becoming the foundation of new engineering disciplines, such as nanotechnology and bioengineering, which have the potential to fundamentally change the way we live.

Computers, and computer science ideas and techniques, are of course an important part of all these scientific and engineering activities. Computer science and its concepts and methods are not only a servant of biological research but also provide mental models used by a new generation of biologists, who often refer to themselves as systems biologists, in thinking about the living world. Ideas and approaches, however, travel in both directions: reflecting on biological ideas has inspired a wide range of computer science questions and has led to the development of important new techniques for solving hard computational problems. The result might be called **biological computation** (or biologically inspired computing) and is the subject of this book.

This book is written from the perspective of computer scientists who are fascinated with biology. A large part of the excitement and fun of bio-inspired computing, at least for us, is learning the amazing and quirky details discovered by biologists. Among the biological stories that have informed computer science you will find discoveries about how bacteria communicate, how ants organize their nests, and the way the immune system learns to recognize pathogens before actually encountering them. All these, and more, are discussed in the chapters to come, along with the

computational techniques they led to. We hope the book manages to convey the sense of wonder and fun that we feel about the field. It goes without saying that it is impossible to go into all details of such varied phenomena, and we concentrate on the aspects of the biological phenomena most closely related to the computational approaches we discuss.

## THE CONTENT OF THE BOOK

---

The term *biological computation* encompasses quite a few approaches. In this book we focus on the most fundamental and important ideas, and on the classic works in each of the subjects we discuss, in an attempt to give a unified overview of computer science ideas inspired by biology. The four major topics we focus on are **cellular automata**, **evolutionary computation**, **neural networks**, and **molecular computation**. Each of these topics is the subject of a chapter that begins by exploring the biological background and then moves on to describe the computational techniques, followed by examples of applications and a discussion of possible variants of the basic techniques introduced in the chapter. Each chapter also includes exercises and solutions. Exercises with solutions are marked with bold numbers. Important ideas and techniques are presented through the example applications and exercises. In addition to the chapters discussing these techniques, Chapter 1 provides a general biological background, and Chapter 6 concludes the book by introducing, more briefly, some of the new topics that are emerging within the field.

We made a special effort to make our explanation of molecular computation accessible to readers who lack a background in molecular biology, without sacrificing the details. In contrast to the other techniques we discuss in the book that can be immediately used by programmers to attack computational challenges, molecular computation is still mostly in its infancy and requires equipment that can be found only in professional laboratories. We feel, however, that thinking about the computational power of molecular events is enlightening, and we predict that computer scientists will enjoy the puzzle-like challenge of trying to represent computational problems as sets of interacting molecules. With the possibility of biological hacking and “Do It Yourself Biology” just around the corner, the use of these techniques may become more widespread than can presently be imagined.

The topics we focus on, with the exception of molecular computing, are already the subject of several good textbooks, which can be found listed in the Recommendations for Additional Reading section of Chapter 6. Most of these books, however, are dedicated to only one of the subjects we



discuss or are extremely detailed reference books. Our goal was to present to you, the reader, an overview of the terrain, allowing you to then focus your attention on the techniques that are most relevant for you. Each of the approaches we cover exists in a multitude of variants and is covered by a large amount of theoretical work—it is very easy to get buried in the details. This book attempts to convey in an easily digestible form the gist of each of the major approaches in the field and to bring you to the point where you can produce a working implementation of each of the basic techniques or to effectively use one of the many existing implementations that can be found online. All the details can be easily found in the literature or by searching online once the basic ideas introduced here are understood.

The techniques we discuss reflect fundamental principles whose applicability goes beyond bio-inspired computing—for example, self-organization, redundancy, using noise, asynchronicity, nondeterminism, and other methods of parallelism and distributed computing. These ideas manifest themselves in other areas of computer science and software engineering, specifically in the development of very large-scale distributed systems, of the sort underlying cloud and grid computing. While these fields are not discussed here, we feel that getting acquainted with these fundamental ideas and playing with simple computational models that exhibit them, such as the ones presented throughout this book, can be rewarding.

## FOR WHOM IS THIS BOOK INTENDED?

We wrote this book thinking primarily of readers with a computer science background and we assume no previous background in biology. For readers who feel they would benefit from a deeper understanding of the biological context we provide references to several recommended books in the Further Reading list in Chapter 1. This book is intended to be a gentle introduction to the field and should be suitable for self-study as well as for use in university courses. We assume the reader is familiar with basic computer science terminology and basic algebra and probability theory but provide detailed explanations of all derivations. There are programming exercises at the end of each chapter, but it is possible to follow the explanations and discussions without programming knowledge. We did not include many formal proofs, but throughout the chapters and exercises we give easy-to-follow examples of several important proof techniques. This should make the book accessible to readers with biological or medical

backgrounds—those coming to the field of bio-inspired computing from biology rather than from computer science.

## USING THIS BOOK AS A TEXTBOOK

---

While the book can be used for self-study, its main purpose is to serve as a textbook for a course on biological computation. Such a course can be given to advanced undergraduate or early graduate students in programs that combine biology and computer science (a double major in computer science and biology or special bioinformatics tracks). For such students a course on biological computation can complement a suite of courses in bioinformatics, algorithms for computational biology, and systems biology.

A course based on this book can also be given to students who major in computer science and for whom a course in biological computation can enrich the perspectives about alternative models of computation. The book contains in the first chapter and in each one of Chapters 2–5 an accessible biological introduction. Nevertheless, it is a good idea for these students to take a basic course in biology prior to a course based on this book or, as was done in Bar-Ilan University, to add lectures and teaching assistant (TA) sessions giving a “crash course” in biology.

The material covered in this book can be delivered in a semester (13–14 weeks) with weekly two-hour lectures and weekly TA sessions. Thus, we devote about three weeks to each one of the four main subjects we cover. From our experience, students gain a lot from homework and especially from the programming exercises, so we provide a good number of those. As the book was written as a textbook, we tried not to overwhelm the readers with footnotes and references.

When we designed and delivered this course in the last several years, no suitable textbook was available, and we felt its absence. Our course was well received by students, and we hope that this book will encourage and enable many teachers and universities to offer similar courses.

## ACKNOWLEDGMENTS

---

Writing a book is a long and complicated process, and we could not have done it without the help and support of many individuals and institutions.

The idea to collaborate on a book came to us while we worked on the development of a course on bio-inspired computing for the Open University of Israel. The structure of this book is based on courses taught by Ron Unger at the Weizmann Institute of Science and for many years at

Bar-Ilan University, and the course developed by the two authors for the Open University of Israel. Our thanks go to these institutions.

Special thanks go to Edna Wigderson, who helped us transform a first draft written in Hebrew into the book before you. Not only did she translate our original material, but she also edited the material, pointed out inconsistencies and mistakes, and helped us improve the presentation immeasurably. Without her this book would never be.

We thank Assaf Massoud for the artwork resulting in the illuminating illustrations that accompany the text. Working with Assaf was a real pleasure. Both Edna and Assaf had to endure the difficulties of dealing with two authors who often disagreed; not only did both endure this with grace, but their prodding also helped the two authors converge.

We also wish to thank all those who read the manuscripts or parts of it, pointed out our mistakes, and made valuable suggestions. First and foremost we thank Yair Horesh, who was involved in the courses in Bar-Ilan and in the Open University and made important contributions to the manuscript. We would also want to recognize the assistance we received from Tania Gottlieb in the biological aspects of the book and Orly Noivirt-Brik, Yochai Gat, Nurit Zer-Kavod, Ariel Azia, Tirza Doniger, Inbal Yomtovian, Ari Yakir, and Ilana Lebenthal for their valuable comments on the manuscript.

As is always the case, the responsibility for whatever errors remain is ours alone.

On a personal note, Ron wishes to thank David Harel, Joel L. Sussman, and John Moulton, mentors and friends who helped him in his transition from computer scientist to computational biologist, and his colleagues at Bar-Ilan University and especially Elisha Haas and Shula Michaeli. Special thanks go to my family—my parents Zipora and Ozer; my sister Vered, who was so helpful in preparing the book; and my wonderful children Amir, Ayelet, Hilla, and Inbal. I want to express special thanks to my vibrant wife Tamar. Tami, without your support and love, this book project could not have been undertaken, let alone completed.

Ehud wishes to thank Eva Jablonka for intellectual stimulation and biological insight, in addition to friendship and moral support; his many colleagues in the computer science department at the Open University of Israel for encouragement and support; and Dror Bar-Nir and Sara Schwartz for invaluable discussions. I wish to thank my family for their love and support. I am especially grateful to my wife Ayelet whose love and friendship are the greatest gift of all.

---

# Table of Contents

---

Preface, xv

CHAPTER 1 ■ Introduction and Biological Background	1
1.1 BIOLOGICAL COMPUTATION	1
1.2 THE INFLUENCE OF BIOLOGY ON MATHEMATICS— HISTORICAL EXAMPLES	4
1.3 BIOLOGICAL INTRODUCTION	7
1.3.1 The Cell and Its Activities	12
1.3.2 The Structure of DNA	14
1.3.3 The Genetic Code	16
1.3.4 Protein Synthesis and Gene Regulation	18
1.3.5 Reproduction and Heredity	23
1.4 MODELS AND SIMULATIONS	26
1.5 SUMMARY	33
1.6 FURTHER READING	34
1.7 EXERCISES	34
1.7.1 Biological Computation	34
1.7.2 History	35
1.7.3 Biological Introduction	35
1.7.4 Models and Simulations	37
1.8 ANSWERS TO SELECTED EXERCISES	37
CHAPTER 2 ■ Cellular Automata	39
2.1 BIOLOGICAL BACKGROUND	39

2.1.1	Bacteria Basics	39
2.1.2	Genetic Inheritance—Downward and Sideways	40
2.1.3	Diversity and the Species Question	41
2.1.4	Bacteria and Humans	42
2.1.5	The Sociobiology of Bacteria	42
2.2	THE “GAME OF LIFE”	44
2.3	GENERAL DEFINITION OF CELLULAR AUTOMATA	48
2.4	1-DIMENSIONAL AUTOMATA	50
2.5	EXAMPLES OF CELLULAR AUTOMATA	54
2.5.1	Fur Color	54
2.5.2	Ecological Models	57
2.5.3	Food Chain	58
2.6	COMPARISON WITH A CONTINUOUS MATHEMATICAL MODEL	59
2.7	COMPUTATIONAL UNIVERSALITY	61
2.7.1	What Is Universality?	61
2.7.2	Cellular Automata as a Computational Model	65
2.7.3	How to Prove That a CA Is Universal	67
2.7.4	Universality of a Two-Dimensional Cellular Automaton—Proof Sketch	68
2.7.5	Universality of the “Game of Life”—Proof Sketch	71
2.8	SELF-REPLICATION	73
2.9	SUMMARY	77
2.10	PSEUDO-CODE	78
2.11	FURTHER READING	79
2.12	EXERCISES	79
2.12.1	“Game of Life”	79
2.12.2	Cellular Automata	80
2.12.3	Computing Using Cellular Automata	82
2.12.4	Self-Replication	82
2.12.5	Programming Exercises	83

2.13	ANSWERS TO SELECTED EXERCISES	84
CHAPTER 3	■ Evolutionary Computation	87
3.1	EVOLUTIONARY BIOLOGY AND EVOLUTIONARY COMPUTATION	87
3.1.1	Natural Selection	87
3.1.2	Evolutionary Computation	93
3.2	GENETIC ALGORITHMS	94
3.2.1	Selection and Fitness	98
3.2.2	Variations on Fitness Functions	102
3.2.3	Genetic Operators and the Representation of Solutions	104
3.3	EXAMPLE APPLICATIONS	108
3.3.1	Scheduling	108
3.3.2	Engineering Optimization	109
3.3.3	Pattern Recognition and Classification	109
3.3.4	Designing Cellular Automata	110
3.3.5	Designing Neural Networks	110
3.3.6	Bioinformatics	110
3.4	ANALYSIS OF THE BEHAVIOR OF GENETIC ALGORITHMS	111
3.4.1	Holland's Building Blocks Hypothesis	115
3.4.2	The Schema Theorem	116
3.4.3	Corollaries of the Schema Theorem	118
3.5	LAMARCKIAN EVOLUTION	119
3.6	GENETIC PROGRAMMING	121
3.7	A SECOND LOOK AT THE EVOLUTIONARY PROCESS	126
3.7.1	Mechanisms for the Generation and Inheritance of Variations	126
3.7.2	Selection	129
3.8	SUMMARY	130
3.9	PSEUDO-CODE	131
3.10	FURTHER READING	132

3.11	EXERCISES	132
3.11.1	Evolutionary Computation	132
3.11.2	Genetic Algorithms	133
3.11.3	Selection and Fitness	133
3.11.4	Genetic Operators and the Representation of Solutions	134
3.11.5	Analysis of the Behavior of Genetic Algorithms	135
3.11.6	Genetic Programming	136
3.11.7	Programming Exercises	136
3.12	ANSWERS TO SELECTED EXERCISES	140
CHAPTER 4	■ Artificial Neural Networks	143
4.1	BIOLOGICAL BACKGROUND	143
4.1.1	Neural Networks as Computational Model	146
4.2	LEARNING	146
4.3	ARTIFICIAL NEURAL NETWORKS	148
4.3.1	General Structure of Artificial Neural Networks	148
4.3.2	Training an Artificial Neural Network	151
4.4	THE PERCEPTRON	152
4.4.1	Definition of a Perceptron	152
4.4.2	Formal Description of the Behavior of a Perceptron	156
4.4.3	The Perceptron Learning Rule	158
4.4.4	Proving the Convergence of the Perceptron Learning Algorithm	159
4.5	LEARNING IN A MULTILAYERED NETWORK	162
4.5.1	The Backpropagation Algorithm	162
4.5.2	Analysis of Learning Algorithms	170
4.5.3	Network Design	172
4.5.4	Examples of Applications	174
4.6	ASSOCIATIVE MEMORY	180
4.6.1	Biological Memory	180
4.6.2	Hopfield Networks	181

4.6.3	Memorization in a Hopfield Network	181
4.6.4	Data Retrieval in a Hopfield Network	183
4.6.5	The Convergence of the Process of Updating the Neurons	185
4.6.6	Analyzing the Capacity of a Hopfield Network	186
4.6.7	Application of a Hopfield Network	189
4.6.8	Further Uses of the Hopfield Network	191
4.7	UNSUPERVISED LEARNING	194
4.7.1	Self-Organizing Maps	195
4.7.2	WEBSOM: Example of Using SOMs for Document Text Mining	198
4.8	SUMMARY	200
4.9	FURTHER READING	201
4.10	EXERCISES	202
4.10.1	Single-Layer Perceptrons	202
4.10.2	Multilayer Networks	203
4.10.3	Hopfield Networks	205
4.10.4	Self-Organizing Maps	208
4.10.5	Summary	208
4.11	ANSWERS TO SELECTED EXERCISES	210
CHAPTER 5 ■ Molecular Computation		215
5.1	BIOLOGICAL BACKGROUND	217
5.1.1	PCR: Polymerase Chain Reaction	217
5.1.2	Gel Electrophoresis	219
5.1.3	Restriction Enzymes	219
5.1.4	Ligation	220
5.2	COMPUTATION USING DNA	220
5.2.1	Hamiltonian Paths	220
5.2.2	Solving SAT	230
5.2.3	DNA Tiling	233
5.2.4	DNA Computing—Summary	236



5.3	ENZYMATIC COMPUTATION	237
5.3.1	Finite Automata	238
5.3.2	Enzymatic Implementation of Finite Automata	242
5.4	SUMMARY	248
5.5	FURTHER READING	250
5.6	EXERCISES	250
5.6.1	Biological Background	250
5.6.2	Computing with DNA	250
5.6.3	Enzymatic Computation	253
5.7	ANSWERS TO SELECTED EXERCISES	254
CHAPTER 6 ■ The Never-Ending Story: Additional Topics at the Interface between Biology and Computation		259
6.1	SWARM INTELLIGENCE	261
6.1.1	Ant Colony Optimization Algorithms	262
6.1.2	Cemetery Organization, Larval Sorting, and Clustering	264
6.1.3	Particle Swarm Optimization	267
6.2	ARTIFICIAL IMMUNE SYSTEMS	270
6.2.1	Identifying Intrusions in a Computer Network	271
6.3	ARTIFICIAL LIFE	273
6.3.1	Avida	276
6.3.2	Evolvable Virtual Creatures	281
6.4	SYSTEMS BIOLOGY	284
6.4.1	Evolution of Modularity	287
6.4.2	Robustness of Biological Systems	289
6.4.3	Formal Languages for Describing Biological Systems	290
6.5	SUMMARY	294
6.6	RECOMMENDATIONS FOR ADDITIONAL READING	297
6.6.1	Biological Introduction	297