SOFTWARE DIRECTORY FOR MOLECULAR BIOLOGISTS

SOFTWARE DIRECTORY FOR MOLECULAR BIOLOGISTS

**A complete guide
to the selection of
computer software
for the management
and analysis of
molecular sequences**

Christopher J Rawlings

M
STOCKTON
PRESS

© Macmillan Publishers Limited, 1986

All rights reserved. No part of this publication
may be reproduced or transmitted, in any form
or by any means, without permission.

Published in the United States and Canada by
Stockton Press, 1986

Library of Congress Cataloging-in-Publication Data

Rawlings, Christopher J., 1954–
Software directory for molecular biologists

Includes index.

1. Molecular biology — Computer programs — Catalogs.
2. Nucleotide sequence — Computer programs — Catalogs.
3. Amino acid sequence — Computer programs — Catalogs.

I. Title

QH506.R38 1986 574.8'8'02855369 86-14521

ISBN 978-0-943818-37-5

First published in the United Kingdom by
MACMILLAN PUBLISHERS LTD (Journals Division), 1986

British Library Cataloguing in Publication Data

Rawlings, Chris

Software directory for molecular biologists.

1. Molecular biology — Computer programs —
Catalogs

I. Title

574.8'8'0285425 QH506

ISBN 978-1-349-08236-0 ISBN 978-1-349-08234-6 (eBook)

DOI 10.1007/978-1-349-08234-6

Important Note

Whilst every care has been taken in compiling
the information contained in this directory,
the Author and Publishers accept no
responsibility for any errors or omissions.

CONTENTS

Contents	i
Introduction	iv
How to use the Directory	v
GENERAL PURPOSE SOFTWARE	1
INTRODUCTION	1
TEXT PROCESSING	2
Manuscript Preparation	2
Choosing a Word Processor	3
Mailing Lists	4
Additional Features	6
<i>Spelling Correction</i>	6
<i>Sorting</i>	6
Scientific Word Processing	7
Hardware Considerations	8
ELECTRONIC SPREADSHEETS	8
Example: Quench Correction	9
Scientific Spreadsheets	10
Hardware Considerations	11
INFORMATION MANAGEMENT	11
Data Management Packages	11
<i>File Management Systems</i>	12
<i>Database Management Systems</i>	12
Building and Using A Database	12
<i>An example: DNA Sequence Database Summary Files</i>	13
<i>Data Definition</i>	13
<i>Query Language</i>	14
<i>Sorting and Arithmetic</i>	15
Inventory Management	17
Reference Management Systems	17
Hardware Considerations	18
TERMINAL EMULATION AND FILE TRANSFER	19
Display Terminal Emulation	19
File Transfer	20
<i>KERMIT</i>	20

Hardware Considerations	21
<i>Modem</i>	21
Summary	22
GRAPHICAL PRESENTATION OF DATA	22
Graphing	22
Drawing	23
Hardware Considerations	23
<i>Permanent Graphics Output</i>	24
<i>Graphical Input</i>	24
INTEGRATED SOFTWARE SYSTEMS	25
Levels of Integration	26
Integrating Molecular Biological Software	27
Hardware Considerations	27
PROGRAM PERFORMANCE AND HARDWARE ENHANCEMENTS	27
Selecting Add-in Boards	28
Limits to PC Performance	29
Remedies for Input/Output Bottlenecks	30
<i>Winchester Disks</i>	31
<i>RAM Disk</i>	32
Remedies for Computational Bottlenecks	33
<i>The Maths Coprocessor</i>	33
Graphics Boards	34
SOFTWARE FOR MOLECULAR BIOLOGY	36
INTRODUCTION	36
NOMENCLATURE	37
ENTRIES IN THE DIRECTORY	37
Authors	37
Programs	37
Program Functions (FUNCTIONS)	37
Comments	38
Computer Style (STYLE)	38
<i>Some Comments on Program Portability</i>	40
Computer System (COMPUTERS)	41
Operating System (OP/SYS)	42
Implementation Languages (LANGUAGES)	42
Memory Requirements (MEMORY)	42
Input Hardware (I. H/WARE)	42
Output Hardware (O. H/WARE)	42
Communications (COMMS)	42
Cost to Non Profit-Making Organisations (COST: NP)	42

Cost to Profit-Making Organisations (COST: P)	43
Contact Address (ADDRESS)	43
References	43
 PROGRAM FUNCTIONS	 43
Molecular Sequence Data	43
<i>Sequence Data Entry</i>	43
<i>Sequence Data Editing</i>	44
<i>Sequence Data Management and Manipulation</i>	45
<i>Sequencing Project Management</i>	46
<i>Sequence Data Collections</i>	48
<i>Sequence Data Analysis</i>	51
<i>Structure Prediction and Analysis</i>	59
Restriction Enzyme, Genetic and Functional Site Maps	61
<i>Restriction Mapping Software</i>	61
<i>Map Data Entry and Editing</i>	62
<i>Map Data Management and Manipulation</i>	64
<i>Restriction Mapping Project Management</i>	64
<i>Map Data Analysis</i>	65
Molecular Genetics and Engineering	65
<i>Bacterial Strain and Recombinant Clone Databases</i>	65
<i>Oligonucleotide Probes and Hybridisation</i>	65
<i>Engineering Restriction Sites</i>	66
<i>Simulation Systems</i>	67
Miscellaneous	67
<i>Data Collections</i>	67
<i>Computer Communications</i>	67
 SOFTWARE DIRECTORY	 68
References	294
Glossary	312
Appendix I - Trademark Declarations	323
Appendix II - Miscellaneous Addresses	324
Author Index	327
Program Name Index	333
Program Function Index	337
Computer System Index	369
Programming Language Index	375
General Index	378

INTRODUCTION

The main purpose of this *Directory* is to provide a comprehensive guide to computer programs available to molecular biologists, from commercial and academic sources, for the analysis and management of molecular sequences and other data generated during molecular sequencing and genetic engineering projects (e.g. restriction maps). The *Directory* has been written for molecular biologists, biotechnologists, laboratory managers and practitioners in the field of biocomputing, who need information about the availability and range of molecular sequencing software. The *Directory* has been compiled from reports in scientific journals since 1981, from demonstration programs and manuals provided by commercial software distributors and on occasion by personal communication with software authors.

Molecular sequencing software is presented in the *Directory* so that it may be used to select programs for an existing computer system, or to help ensure that a laboratory computer is purchased which has an adequate range of software available. Cross-reference indexes are provided by *Programming Language*, *Computer System* and *Program Function* to help guide the reader to the appropriate software and supplier. As a reference text for scientific and technical progress in molecular sequence analysis methods, there are also cross-reference indexes by *Author* and *Program Name*. A *General* subject index provides rapid access to all parts of the *Directory*.

On the assumption that many readers will be newcomers to computerised methods in the molecular biology laboratory, the first part of the *Directory* reviews how software developed on microcomputers for business users (e.g. spreadsheet calculators) might usefully be exploited in a scientific environment. A glossary of computing terms is provided to explain computer terminology when it cannot be avoided.

HOW TO USE THIS DIRECTORY

The *Directory* is divided into four main sections. The first section on *General Purpose Software* has been written for scientists new to the use of a microcomputer in the laboratory. It is intended to help the reader make good use of the considerable power and flexibility of personal computers, by indicating potential scientific uses of software developed for the business community. This type of software is often available *bundled* together with a personal computer system as a marketing ploy, and it makes sense if such "free" software can be put to good use. This section also describes important general purpose software for communicating with other computers: display terminal emulation and file transfer programs. It concludes with descriptions of how the performance and utility of a personal computer can be enhanced using add-in boards and specialised coprocessors.

The second section, *Software for Molecular Biology*, is the introduction to the *Directory* in section three. The first three parts are important for all readers because they describe the nomenclature, how entries in the *Directory* are structured and what information will be found there. The fourth part of section two (*Program Functions*) is a detailed account of over two hundred and thirty program functions that can be found listed in the *Directory*. This subsection is essential reading for anyone who is not aware of the range of computer-assisted functions available for the molecular biologist. The *Program Functions* subsection should also be consulted when a particular program or package has been selected, in order to confirm that functions listed in the *Directory* meet with the reader's actual requirements. The index of program functions in section four provides the cross reference between the description of a function and those programs in the *Directory* that support it.

The third section is the *Directory* itself. Entries are listed in alphabetical order, using the surname of a principal author or the name of the company that developed the programs. Each program or suite of programs is described by sixteen features: the *Authors*, the names of the *Programs*, a list of *Program Functions*, general *Comments*, the *Computer Styles* (e.g. minicomputer, main-frame etc.) and *Operating Systems* on which the programs run, the *Implementation Languages* used, the *Input Hardware* and *Output Hardware* requirements, the computer *Communications* features supported, the *Cost to non-Profit-Making Organisations* and the *Cost to Profit-Making Organisations*, the *Address* of the contact persons and the *bibliographic References*.

The fourth section contains the bibliography, the *Glossary* of computing terms and the *Author*, *Program Name*, *Program Function*, *Computer System*, *Programming Language* and *General* indexes.