Data-Handling in Biomedical Science

Packed with worked examples and problems for you to try, this book will help to improve your confidence and skill in data-handling. The mathematical methods needed for problem-solving are described in the first part of the book, with chapters covering topics such as indices, graphs and logarithms. The following eight chapters explore data-handling using different areas of microbiology and biochemistry including microbial growth, enzymes and radioactivity as examples. Each chapter is fully illustrated with worked examples that provide a step-by-step guide to the solution of the most common problems. Over 30 exercises, ranging in difficulty and length, allow you to practise your skills and are accompanied by a full set of hints and solutions.

PETER WHITE taught practical classes and tutorials in microbiology for over 30 years during his time as a Lecturer and Senior Lecturer at Sheffield University. His research interests include microbial drug resistance, biochemistry of bacterial walls and metabolism of members of the genus *Bacillus*. He is a member of the Society for General Microbiology and a past member of the Biochemical Society and the American Society for Microbiology.

Data-Handling in Biomedical Science

PETER WHITE
University of Sheffield

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment,
a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of
education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9780521143868

First published 2010

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging-in-Publication data
White, Peter, 1931–
Data-handling in biomedical science / Peter White.
p. cm.
Includes index.
ISBN 978-0-521-19455-6 – ISBN 978-0-521-14386-8 (pbk.) 1. Microbiology – Data processing. I. Title.
QR69.D35W45 2010
616.9´041–dc22
2009053993

ISBN 978-0-521-19455-6 Hardback
ISBN 978-0-521-14386-8 Paperback

Dedication

To all my teachers, but especially to remember
Miss Dollan
(Leigh Church of England Junior School 1941–1942)

Contents

Introduction

Data-handling means the interpreting and refining of experimental results. This book is aimed at helping to improve confidence and skill in data-handling. It is intended for undergraduate students, and for graduate students who may still have a little to learn.

Although microbiology began with simple observations (the organisms are small, they have various shapes, and some are motile) the subject has become a quantitative, experimental science. As an example consider the following statement:

> Poly β-hydroxybutyrate may make up 70% of the dry weight of *Azotobacter*.

To reach this conclusion one must grow the organisms in a certain way, make weighings, extract the polymer, do a chemical assay, and then put all the information together in its proper order. None of these practical steps is difficult, but to achieve the final result, clarity of thought, rather than great mathematical ability, is definitely needed.

In recent times, the words 'data-handling' have taken on a second meaning, that is, the manipulation of very large quantities of data (such as DNA sequences) by using computer programs for analysis and comparison. This new big area of database management is not covered here. Everything in this book can be done with pencil and paper and a pocket calculator. That is by no means to decry computers; the whole of this text was written on a word processor, and all the figures have been drawn with Excel® or Corel Draw®.

How is data-handling to be learned? Best of all, **by personal experience in the laboratory**. If you design an experiment yourself, then you will have thought about the form of your results and the way in which they will need to be manipulated. If you do an experiment that someone else has designed (as in a practical class) it may be harder, but is still very instructive, to work out how to do the data-handling. When you are told how to do it, then make sure that you really understand the steps; this is where many students fall short.

The other way to learn is by solving written problems. Reading through problems, without seriously attempting to attack them, is a waste of time. Listening to someone explaining how a problem is done may be instructive, but it can make things seem too easy (like seeing the solution to a crossword puzzle) or it can make things appear too difficult (when really you could have done it if you had tried).

How is data-handling to be taught? The answer is simple: by encouraging **practice, practice and practice** with written problems and at the bench with real situations. People who are not alarmed by simple mathematics or who like puzzles might have an advantage in data-handling, yet practice will improve anyone's ability. Becoming familiar with types of calculation that occur again and again builds confidence to tackle new situations. Building this confidence is extremely important, and is best done by starting with easier problems and moving gradually towards the harder.

This book starts with four chapters about simple mathematics and one about graphs. All of this material can be skipped, but it has to be well understood before attempting the later chapters. The next two chapters, about logarithms and statistics, are more difficult. However, these should not be skipped unless you are very assured in both topics.

After this come eight chapters in which data-handling in different areas of microbiology and biochemistry are discussed and are illustrated by fully worked-out problems. Finally, a miscellany of problems is given, with the answers separated to a following chapter.

These problems were designed by various people; their names are given at the starts of the problems; no name means by me. All the answers and any errors are my sole responsibility. I hope the quotations may be seen to have some relevance, and not to be just show-offs. Blame Edgar Allan Poe or Colin Dexter.

How important is maths in data-handling?

The answer is vitally, but do not despair. There are few areas of microbiology or biochemistry that require any kind of advanced mathematical ability. What is necessary is to know how to crank out answers from standard methods. Most students, and most of their teachers too, have not been trained to a high level in mathematics. In the author's case this means trained only as far as School Certificate, the forerunner of 'O' level. An expert is usually consulted in those rare cases where advanced knowledge is needed.

Students are expected to be numerate, and modest skills are certainly necessary. Appreciation of simple proportion, knowledge of some easy algebra (such as solving simultaneous equations), an understanding of logarithms and basic statistics, and the ability to draw and interpret graphs are all needed. These topics are revised in the first chapters of this book. However, most important of all is to develop a confidence that nearly all data-handling problems can be tackled without having to be a talented mathematician.

Clear thinking and simple mathematics will solve most problems (at least in this branch of science)!

> *The fact is that there are few more 'popular' subjects than mathematics. Most people have some appreciation of mathematics, just as most people can enjoy a pleasant tune, and there are probably more people really interested in mathematics than in music. Appearances may suggest the contrary, but there are easy explanations. Music can be used to stimulate mass emotion, while mathematics cannot; and musical incapacity is recognized (no doubt rightly) as mildly discreditable, whereas most people are so frightened of the name of mathematics that they are ready, quite unaffectedly, to exaggerate their own mathematical stupidity.*
>
> G. H. Hardy (1940)

Abbreviations and the Système International

When the same word or phrase appears many times in a piece of writing, common practice is to abbreviate. For instance, in this book adenosine triphosphate is written as ATP; reduced nicotinamide adenine dinucleotide is NADH. Some abbreviations are understood by every reader of English, like Mr, USA, Prof., i.e. (= id est = that is). Other abbreviations, for example μg, are only understood by specialists, and others may be too new, such as cu (= see you), for everyone to understand.

In science there is a need for abbreviations that are intelligible to all readers in any language. In order to express ten grams one abbreviates as 10 g; to write twenty seconds one puts 20 s (note that numerals are abbreviations too); and one hundred degrees Centigrade becomes 100 °C. There may be times, of course, when you don't want to use an abbreviation, and then it might be appropriate to write (say) 100 millilitres rather than 100 ml.

The Système International (SI) is an agreement that in all science there shall be certain basic units for measuring distance (the metre, m), time (the second, s), weight (the kilogram, kg), temperature (degree Kelvin, K), luminous intensity (the candela, cd), electric current (the ampere, A), and quantity (the mole, mol). There are multiples or fractions of these units, some of which have names of their own (e.g. 1 min = 60 s; 1 h = 3600 s). Derived units are formed by combination of two or more of these basic units; the joule (J) is m^2 .kg. s^{-2} and the watt (W) is J.s^{-1} (= m^2 .kg. s^{-3}).

SI units are now used in all scientific writing. **Their abbreviations are fixed and are never pluralised and never have a full stop following (unless at the end of a sentence or in place of a multiplication symbol** [×]**)**. Thus, in an article for a journal you can write 50 g but cannot have 50 g. or 50 gs or 50 gm or 50 gms. Extensive lists of SI units are available in libraries and on the Internet with their definitions and proper abbreviations, and with quite elaborate rules too. It is important to know the abbreviations for units that you may frequently use.

The litre (1000 ml) is a special problem. The SI abbreviates litre as l (hence ml, μl etc.) and trouble comes when you want to express a number of whole litres, like 10 l, or 100 l, or to abbreviate 'per litre' (l^{-1}). In these cases, where l looks too like a 1 or an I, then I think it is better to write 10 L and 100 L, or L^{-1}. This usage is common, and I follow it. In the USA ml and μl will often be written as mL and μL, but I am not so thoroughly Americanised (or consistent) as to do this.

Any abbreviation that may not be understood by one's entire readership must be explained. If in any doubt then explain. The medical profession coins abbreviations that all too often go without explanation. What about MRSA? It is improbable that every reader knows that this contraction means 'methicillin-resistant *Staphylococcus aureus*' and not something like 'most repulsive slimy animal' that you meet in a hospital.

Abbreviations used in this book

SI units and abbreviations are used wherever possible. Note the litre anomaly (above).

The erg is the full name of a unit, not an abbreviation, and because it was defined in terms of $cm^2.g.\ s^{-2}$, the erg is equal to 1×10^{-7} J.

Prefixes

M mega- ($\times 10^6$)
k kilo- ($\times 10^3$)
m milli- ($\times 10^{-3}$)
μ micro- ($\times 10^{-6}$)
n nano- ($\times 10^{-9}$)

Other abbreviations

A adenine
ADP adenosine diphosphate
Ala alanine
ATP adenosine triphosphate

C	cytosine
cal	calorie
CoA	coenzyme A
CoQ	coenzyme Q (ubiquinone)
Da	dalton = unified atomic mass unit [= 1 / (Avogadro's number) gram]
Dap	2,6-diaminopimelic acid
DEAE-	diethylaminoethyl-[cellulose]
DNA	deoxyribonucleic acid
e.g.	for example
F	the Faraday; the total electrical charge on 1 mole of electrons
FAD	flavin adenine dinucleotide (oxidised form)
$FADH_2$	flavin adenine dinucleotide (reduced form)
G	guanine
Glu	glutamate
i.e.	that is
Lys	lysine
mol. wt	molecular weight (= relative molecular mass, RMM)
M	molar (note use of smaller typeface compared with the prefix M = mega)
mur	muramic acid
NAD^+	nicotinamide adenine dinucleotide (oxidised form)
NADH	nicotinamide adenine dinucleotide (reduced form)
$NADP^+$	nicotinamide adenine dinucleotide phosphate (oxidised form)
NADPH	nicotinamide adenine dinucleotide phosphate (reduced form)
OD	optical density
P_i	inorganic phosphate
PP_i	inorganic pyrophosphate
PQQ	pyrroloquinoline quinone
R	the universal gas constant (1.986 calories per degree Kelvin per mole)
STP	standard temperature (273 degrees Kelvin) and pressure (1 atmosphere, 760 mm mercury)
T	thymine
T	temperature in degrees Kelvin
UDP	uridine diphosphate
t_d	doubling time during exponential growth

UV	ultra violet
V	volt
v.	versus (= against)
v / v	volume per volume
wt	weight
w / v	weight per volume

Some other abbreviations may be defined where they occur in the text.

Acknowledgements

I am grateful to all those students and staff, academic, technical and secretarial in the Department of Molecular Biology at Sheffield University who have helped so much in the making of this book. Particularly I want to thank Dr D. J. Gilmour and Professor Anne Moir for their interest and advice.

Martin Griffiths and Lynette Talbot, of Cambridge University Press, have been most friendly and helpful throughout production. It has also been a pleasure to work with my copy-editor and cricket-lover, Anna Hodson.

Most of all, of course, I thank my wife, Ruth, for just about everything.