

Calculations for

Molecular Biology and Biotechnology

A Guide to Mathematics in the Laboratory

Second Edition

Frank H. Stephenson Ph.D



Calculations for Molecular Biology and Biotechnology

A Guide to Mathematics in the Laboratory

Second Edition



AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD • PARIS
SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
32 Jamestown Road, London NW1 7BY, UK
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
525 B Street, Suite 1800, San Diego, CA 92101-4495, USA

First edition 2003
Second edition 2010

Copyright © 2010 Elsevier Inc. All rights reserved

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44) (0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions@elsevier.com. Alternatively, visit the Science and Technology Books website at www.elsevierdirect.com/rights for further information

Notice

No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein.

Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made

British Library Cataloguing-in-Publication Data

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

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

ISBN : 978-0-12-375690-9

For information on all Academic Press publications
visit our website at www.elsevierdirect.com

Typeset by MPS Limited, a Macmillan Company, Chennai, India
www.macmillansolutions.com

Printed and bound in the United States of America

10 11 12 13 10 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER

BOOK AID
International

Sabre Foundation

Calculations for Molecular Biology and Biotechnology

*To my parents Mary and Dude and to my wife Laurie
and my beautiful daughter Myla.*

Contents

CHAPTER 1 Scientific Notation and Metric Prefixes	1
Introduction	1
1.1 Significant Digits	1
1.1.1 Rounding Off Significant Digits in Calculations	2
1.2 Exponents and Scientific Notation	3
1.2.1 Expressing Numbers in Scientific Notation	3
1.2.2 Converting Numbers from Scientific Notation to Decimal Notation	5
1.2.3 Adding and Subtracting Numbers Written in Scientific Notation	6
1.2.4 Multiplying and Dividing Numbers Written in Scientific Notation	7
1.3 Metric Prefixes	10
1.3.1 Conversion Factors and Canceling Terms	10
Chapter Summary	14
 CHAPTER 2 Solutions, Mixtures, and Media	 15
Introduction	15
2.1 Calculating Dilutions – A General Approach	15
2.2 Concentrations by a Factor of X	17
2.3 Preparing Percent Solutions	19
2.4 Diluting Percent Solutions	20
2.5 Moles and Molecular Weight – Definitions	24
2.5.1 Molarity	25
2.5.2 Preparing Molar Solutions in Water with Hydrated Compounds	28
2.5.3 Diluting Molar Solutions	30
2.5.4 Converting Molarity to Percent	32
2.5.5 Converting Percent to Molarity	33
2.6 Normality	34
2.7 pH	35
2.8 pK_a and the Henderson–Hasselbalch Equation	40
Chapter Summary	43

CHAPTER 3 Cell Growth	45
3.1 The Bacterial Growth Curve	45
3.1.1 Sample Data	49
3.2 Manipulating Cell Concentration	50
3.3 Plotting OD₅₅₀ vs. Time on a Linear Graph.....	53
3.4 Plotting the Logarithm of OD₅₅₀ vs. Time on a	
Linear Graph	54
3.4.1 Logarithms	54
3.4.2 Sample OD ₅₅₀ Data Converted to Logarithm Values	54
3.4.3 Plotting Logarithm OD ₅₅₀ vs. Time	54
3.5 Plotting the Logarithm of Cell Concentration vs. Time	56
3.5.1 Determining Logarithm Values.....	56
3.6 Calculating Generation Time.....	57
3.6.1 Slope and the Growth Constant	57
3.6.2 Generation Time	58
3.7 Plotting Cell Growth Data on a Semilog Graph	60
3.7.1 Plotting OD ₅₅₀ vs. Time on a Semilog Graph.....	60
3.7.2 Estimating Generation Time from a Semilog Plot of	
OD ₅₅₀ vs. Time	61
3.8 Plotting Cell Concentration vs. Time on a Semilog Graph	62
3.9 Determining Generation Time Directly from a Semilog Plot	
of Cell Concentration vs. Time	63
3.10 Plotting Cell Density vs. OD₅₅₀ on a Semilog Graph	64
3.11 The Fluctuation Test	66
3.11.1 Fluctuation Test Example	67
3.11.2 Variance	69
3.12 Measuring Mutation Rate	71
3.12.1 The Poisson Distribution	71
3.12.2 Calculating Mutation Rate Using the Poisson	
Distribution.....	72
3.12.3 Using a Graphical Approach to Calculate Mutation	
Rate from Fluctuation Test Data.....	73
3.12.4 Mutation Rate Determined by Plate	
Spreading	78
3.13 Measuring Cell Concentration on a Hemocytometer	79
Chapter Summary	80
References	81

CHAPTER 4 Working with Bacteriophages.....	83
Introduction	83
4.1 Multiplicity of Infection (moi).	83
4.2 Probabilities and Multiplicity of Infection (moi).	85
4.3 Measuring Phage Titer.	91
4.4 Diluting Bacteriophage	93
4.5 Measuring Burst Size.	95
Chapter Summary	98
 CHAPTER 5 Nucleic Acid Quantification	 99
5.1 Quantification of Nucleic Acids by Ultraviolet (UV) Spectroscopy	99
5.2 Determining the Concentration of Double-Stranded DNA (dsDNA).	100
5.2.1 Using Absorbance and an Extinction Coefficient to Calculate Double-Stranded DNA (dsDNA) Concentration	102
5.2.2 Calculating DNA Concentration as a Millimolar (mM) Amount.	104
5.2.3 Using PicoGreen® to Determine DNA Concentration	105
5.3 Determining the Concentration of Single-Stranded DNA (ssDNA) Molecules	108
5.3.1 Single-Stranded DNA (ssDNA) Concentration Expressed in $\mu\text{g/mL}$	108
5.3.2 Determining the Concentration of High-Molecular- Weight Single-Stranded DNA (ssDNA) in $\text{pmol}/\mu\text{L}$	109
5.3.3 Expressing Single-Stranded DNA (ssDNA) Concentration as a Millimolar (mM) Amount	110
5.4 Oligonucleotide Quantification	111
5.4.1 Optical Density (OD) Units	111
5.4.2 Expressing an Oligonucleotide's Concentration in $\mu\text{g/mL}$	111
5.4.3 Oligonucleotide Concentration Expressed in $\text{pmol}/\mu\text{L}$	112
5.5 Measuring RNA Concentration	115
5.6 Molecular Weight, Molarity, and Nucleic Acid Length	115
5.7 Estimating DNA Concentration on an Ethidium Bromide- Stained Gel.	120
Chapter Summary	121

CHAPTER 6 Labeling Nucleic Acids with Radioisotopes	123
Introduction	123
6.1 Units of Radioactivity – The Curie (Ci)	123
6.2 Estimating Plasmid Copy Number	124
6.3 Labeling DNA by Nick Translation.	126
6.3.1 Determining Percent Incorporation of Radioactive Label from Nick Translation.	127
6.3.2 Calculating Specific Radioactivity of a Nick Translation Product.	128
6.4 Random Primer Labeling of DNA.	128
6.4.1 Random Primer Labeling – Percent Incorporation.	129
6.4.2 Random Primer Labeling – Calculating Theoretical Yield	130
6.4.3 Random Primer Labeling – Calculating Actual Yield	131
6.4.4 Random Primer Labeling – Calculating Specific Activity of the Product	132
6.5 Labeling 3' Termini with Terminal Transferase	133
6.5.1 3'-end Labeling with Terminal Transferase – Percent Incorporation.	133
6.5.2 3'-end Labeling with Terminal Transferase – Specific Activity of the Product	134
6.6 Complementary DNA (cDNA) Synthesis	135
6.6.1 First Strand cDNA Synthesis	135
6.6.2 Second Strand cDNA Synthesis	139
6.7 Homopolymeric Tailing.	141
6.8 <i>In Vitro</i> Transcription.	147
Chapter Summary	149
 CHAPTER 7 Oligonucleotide Synthesis	 155
Introduction	155
7.1 Synthesis Yield	156
7.2 Measuring Stepwise and Overall Yield by the Dimethoxytrityl (DMT) Cation Assay	158
7.2.1 Overall Yield.	159
7.2.2 Stepwise Yield	160
7.3 Calculating Micromoles of Nucleoside Added at Each Base Addition Step	161
Chapter Summary	162

CHAPTER 8 The Polymerase Chain Reaction (PCR)	165
Introduction	165
8.1 Template and Amplification	165
8.2 Exponential Amplification	167
8.3 Polymerase Chain Reaction (PCR) Efficiency	170
8.4 Calculating the T_m of the Target Sequence	173
8.5 Primers	176
8.6 Primer T_m	181
8.6.1 Calculating T_m Based on Salt Concentration, G/C	
Content, and DNA Length	182
8.6.2 Calculating T_m Based on Nearest-Neighbor Interactions	183
8.7 Deoxynucleoside Triphosphates (dNTPs)	189
8.8 DNA Polymerase	191
8.8.1 Calculating DNA Polymerase's Error Rate	192
8.9 Quantitative Polymerase Chain Reaction (PCR)	195
Chapter Summary	207
References	209
Further Reading	209
CHAPTER 9 The Real-time Polymerase Chain Reaction (RT-PCR)	211
Introduction	211
9.1 The Phases of Real-time PCR	212
9.2 Controls	215
9.3 Absolute Quantification by the TaqMan Assay	216
9.3.1 Preparing the Standards	216
9.3.2 Preparing a Standard Curve for Quantitative Polymerase	
Chain Reaction (qPCR) Based on Gene Copy Number	220
9.3.3 The Standard Curve	224
9.3.4 Standard Deviation	227
9.3.5 Linear Regression and the Standard Curve	230
9.4 Amplification Efficiency	232
9.5 Measuring Gene Expression	236
9.6 Relative Quantification – The $\Delta\Delta C_T$ Method	238
9.6.1 The $2^{-\Delta\Delta C_T}$ Method – Deciding on an Endogenous	
Reference	239
9.6.2 The $2^{-\Delta\Delta C_T}$ Method – Amplification Efficiency	250
9.6.3 The $2^{-\Delta\Delta C_T}$ Method – is the Reference Gene	
Affected by the Experimental Treatment?	259

9.7	The Relative Standard Curve Method	276
9.7.1	Standard Curve Method for Relative Quantitation	276
9.8	Relative Quantification by Reaction Kinetics	294
9.9	The R_0 Method of Relative Quantification	299
9.10	The Pfaffl Model	303
	Chapter Summary	306
	References	310
	Further Reading	310
CHAPTER 10	Recombinant DNA	313
	Introduction	313
10.1	Restriction Endonucleases	313
10.1.1	The Frequency of Restriction Endonuclease Cut Sites	315
10.2	Calculating the Amount of Fragment Ends	316
10.2.1	The Amount of Ends Generated by Multiple Cuts	317
10.3	Ligation	319
10.3.1	Ligation Using λ -Derived Vectors	322
10.3.2	Packaging of Recombinant λ Genomes	327
10.3.3	Ligation Using Plasmid Vectors	330
10.3.4	Transformation Efficiency	335
10.4	Genomic Libraries – How Many Clones Do You Need?	336
10.5	cDNA Libraries – How Many Clones are Enough?	337
10.6	Expression Libraries	339
10.7	Screening Recombinant Libraries by Hybridization to DNA Probes	340
10.7.1	Oligonucleotide Probes	342
10.7.2	Hybridization Conditions	344
10.7.3	Hybridization Using Double-Stranded DNA (dsDNA) Probes	350
10.8	Sizing DNA Fragments by Gel Electrophoresis	351
10.9	Generating Nested Deletions Using Nuclease BAL 31	359
	Chapter Summary	363
	References	367

CHAPTER 11 Protein	369
Introduction	369
11.1 Calculating a Protein's Molecular Weight from Its Sequence	369
11.2 Protein Quantification by Measuring Absorbance at 280 nm	373
11.3 Using Absorbance Coefficients and Extinction Coefficients to Estimate Protein Concentration	374
11.3.1 Relating Absorbance Coefficient to Molar Extinction Coefficient	377
11.3.2 Determining a Protein's Extinction Coefficient	378
11.4 Relating Concentration in Milligrams Per Milliliter to Molarity	380
11.5 Protein Quantitation Using A_{280} When Contaminating Nucleic Acids are Present	382
11.6 Protein Quantification at 205 nm	383
11.7 Protein Quantitation at 205 nm When Contaminating Nucleic Acids are Present	383
11.8 Measuring Protein Concentration by Colorimetric Assay – The Bradford Assay	385
11.9 Using β -Galactosidase to Monitor Promoter Activity and Gene Expression	387
11.9.1 Assaying β -Galactosidase in Cell Culture	388
11.9.2 Specific Activity	390
11.9.3 Assaying β -Galactosidase from Purified Cell Extracts	390
11.10 Thin Layer Chromatography (TLC) and the Retention Factor (R_f)	392
11.11 Estimating a Protein's Molecular Weight by Gel Filtration	394
11.12 The Chloramphenicol Acetyltransferase (CAT) Assay	399
11.12.1 Calculating Molecules of Chloramphenicol Acetyltransferase (CAT)	401
11.13 Use of Luciferase in a Reporter Assay	403
11.14 <i>In Vitro</i> Translation – Determining Amino Acid Incorporation	404
11.15 The Isoelectric Point (pI) of a Protein	405
Chapter Summary	408

References	411
Further Reading	412
CHAPTER 12 Centrifugation	413
Introduction	413
12.1 Relative Centrifugal Force (RCF) (<i>g</i> Force)	413
12.1.1 Converting <i>g</i> Force to Revolutions Per Minute (rpm)	415
12.1.2 Determining <i>g</i> Force and Revolutions Per Minute (rpm) by Use of a Nomogram	416
12.2 Calculating Sedimentation Times	418
Chapter Summary	420
References	421
Further Reading	421
CHAPTER 13 Forensics and Paternity	423
Introduction	423
13.1 Alleles and Genotypes	424
13.1.1 Calculating Genotype Frequencies	425
13.1.2 Calculating Allele Frequencies	426
13.2 The Hardy–Weinberg Equation and Calculating Expected Genotype Frequencies	427
13.3 The Chi-Square Test – Comparing Observed to Expected Values	430
13.3.1 Sample Variance	434
13.3.2 Sample Standard Deviation	435
13.4 The Power of Inclusion (P_i)	435
13.5 The Power of Discrimination (P_d)	436
13.6 DNA Typing and Weighted Average	437
13.7 The Multiplication Rule	438
13.8 The Paternity Index (PI)	439
13.8.1 Calculating the Paternity Index (PI) When the Mother’s Genotype is not Available	441
13.8.2 The Combined Paternity Index (CPI)	443

Chapter Summary	444
References	445
Further Reading	445
Appendix A	447
Index	455

Scientific notation and metric prefixes

■ INTRODUCTION

There are some 3 000 000 000 base pairs (bp) making up human genomic DNA within a haploid cell. If that DNA is isolated from such a cell, it will weigh approximately 0.000 000 000 003 5 grams (g). To amplify a specific segment of that purified DNA using the polymerase chain reaction (PCR), 0.000 000 000 01 moles (*M*) of each of two primers can be added to a reaction that can produce, following some 30 cycles of the PCR, over 1 000 000 000 copies of the target gene.

On a day-to-day basis, molecular biologists work with extremes of numbers far outside the experience of conventional life. To allow them to more easily cope with calculations involving extraordinary values, two shorthand methods have been adopted that bring both enormous and infinitesimal quantities back into the realm of manageability. These methods use scientific notation and metric prefixes. They require the use of exponents and an understanding of significant digits.

1.1 SIGNIFICANT DIGITS

Certain techniques in molecular biology, as in other disciplines of science, rely on types of instrumentation capable of providing precise measurements. An indication of the level of precision is given by the number of digits expressed in the instrument's readout. The numerals of a measurement representing actual limits of precision are referred to as **significant digits**.

Although a zero can be as legitimate a value as the integers one through nine, significant digits are usually nonzero numerals. Without information on how a measurement was made or on the precision of the instrument used to make it, zeros to the left of the decimal point trailing one or more nonzero numerals are assumed not to be significant. For example, in stating that the human genome is 3 000 000 000 bp in length, the only significant digit in the number is the 3. The nine zeros are not significant. Likewise, zeros to the right of the decimal point preceding a set of nonzero numerals are assumed not to be significant. If we determine that the DNA within a

sperm cell weighs 0.000 000 000 003 5 g, only the 3 and the 5 are significant digits. The 11 zeros preceding these numerals are not significant.

Problem 1.1 How many significant digits are there in each of the following measurements?

- a) 3 001 000 000 bp
- b) 0.003 04 g
- c) 0.000 210 liters (L) (volume delivered with a calibrated micropipettor).

Solution 1.1

- a) Number of significant digits: 4; they are: 3001
- b) Number of significant digits: 3; they are: 304
- c) Number of significant digits: 3; they are: 210

1.1.1 Rounding off significant digits in calculations

When two or more measurements are used in a calculation, the result can only be as accurate as the least precise value. To accommodate this necessity, the number obtained as solution to a computation should be rounded off to reflect the weakest level of precision. The guidelines in the following box will help determine the extent to which a numerical result should be rounded off.

Guidelines for rounding off significant digits

1. When adding or subtracting numbers, the result should be rounded off so that it has the same number of significant digits to the right of the decimal as the number used in the computation with the fewest significant digits to the right of the decimal.
2. When multiplying or dividing numbers, the result should be rounded off so that it contains only as many significant digits as the number in the calculation with the fewest significant digits.

Problem 1.2 Perform the following calculations, and express the answer using the guidelines for rounding off significant digits described in the preceding box

- a) $0.2884 \text{ g} + 28.3 \text{ g}$
- b) $3.4 \text{ cm} \times 8.115 \text{ cm}$
- c) $1.2 \text{ L} \times 0.155 \text{ L}$

Solution 1.2

a) $0.2884 \text{ g} + 28.3 \text{ g} = 28.5884 \text{ g}$

The sum is rounded off to show the same number of significant digits to the right of the decimal point as the number in the equation with the fewest significant digits to the right of the decimal point. (In this case, the value 28.3 has one significant digit to the right of the decimal point.)

28.5884 g is rounded off to 28.6 g

b) $3.4 \text{ cm} \times 8.115 \text{ cm} = 27.591 \text{ cm}^2$

The answer is rounded off to two significant digits since there are as few as two significant digits in one of the multiplied numbers (3.4 cm).

27.591 cm² is rounded off to 28 cm²

c) $1.2 \text{ L} \div 0.155 \text{ L} = 7.742 \text{ L}$

The quotient is rounded off to two significant digits since there are as few as two significant digits in one of the values (1.2 L) used in the equation.

7.742 L is rounded off to 7.7 L

1.2 EXPONENTS AND SCIENTIFIC NOTATION

An **exponent** is a number written above and to the right of (and smaller than) another number (called the **base**) to indicate the power to which the base is to be raised. Exponents of base 10 are used in scientific notation to express very large or very small numbers in a shorthand form. For example, for the value 10^3 , 10 is the base and 3 is the exponent. This means that 10 is multiplied by itself three times ($10^3 = 10 \times 10 \times 10 = 1000$). For numbers less than 1.0, a negative exponent is used to express values as a reciprocal of base 10. For example,

$$10^{-3} = \frac{1}{10^3} = \frac{1}{10 \times 10 \times 10} = \frac{1}{1000} = 0.001$$

1.2.1 Expressing numbers in scientific notation

To express a number in scientific notation:

1. Move the decimal point to the right of the leftmost nonzero digit. Count the number of places the decimal has been moved from its original position.