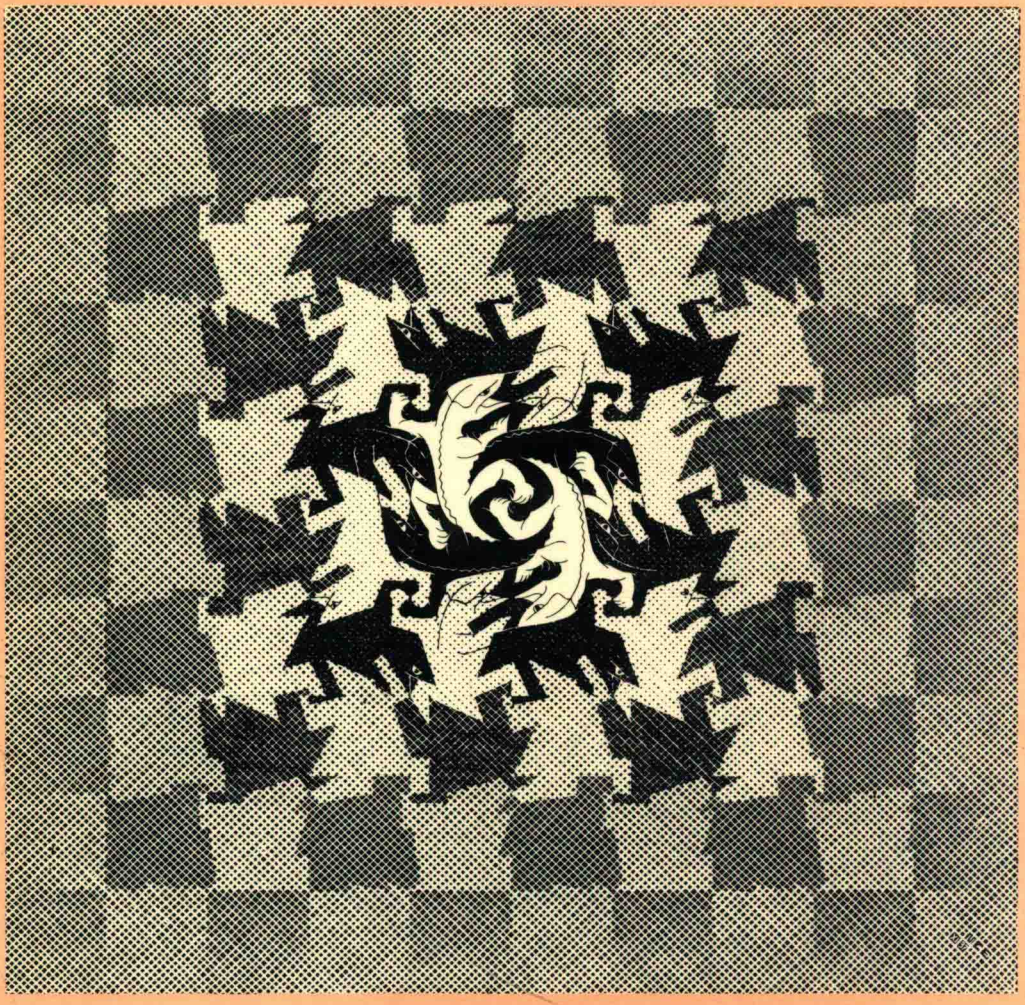
Accuracy in Molecular Processes

ITS CONTROL AND RELEVANCE TO LIVING SYSTEMS

Edited by
T.B.L. KIRKWOOD,
R.F. ROSENBERGER and D.J. GALAS



Accuracy in Molecular Processes

ITS CONTROL AND RELEVANCE TO LIVING SYSTEMS

Edited by

T. B. L. KIRKWOOD AND R. F. ROSENBERGER

National Institute for Medical Research
London, UK

and

D. J. GALAS

University of Southern California
Los Angeles, USA

London New York
CHAPMAN AND HALL

First published in 1986 by
Chapman and Hall Ltd
11 New Fetter Lane, London EC4P 4EE
Published in the USA by
Chapman and Hall
29 West 35th Street, New York, NY 10001

Printed in Great Britain at the University Press, Cambridge

ISBN 0 412 26940 6

British Library Cataloguing in Publication Data

Accuracy in molecular processes : its control and relevance to living systems.
1. Molecular biology
I. Kirkwood, T. B. L. II. Rosenberger, R. F.
III. Galas, D. J.
574.8'8 QH506

ISBN 0-412-26940-6

Library of Congress Cataloging in Publication Data

Accuracy in molecular processes.
Includes bibliographies and index.
1. Molecular biology – Miscellanea. 2. Molecular biology. I. Kirkwood, T. B. L. II. Rosenberger, R. F.
III. Galas, D. J.
QH506.A23 1986 574.87'328 85-25505
ISBN 0-412-26940-6

Accuracy in Molecular Processes

ITS CONTROL AND RELEVANCE TO LIVING SYSTEMS

Contributors

Dr R. P. Anderson
Department of Biology
University of Iowa
Iowa City
Iowa 52242
USA

Dr C. Blomberg
Department of Theoretical Physics
Royal Institute of Technology
S-100 44
Stockholm
Sweden

Dr E. W. Branscomb
Lawrence Livermore National Laboratory
Biomedical Sciences Division
University of California
PO Box 5507 L-452
Livermore
California 94550
USA

Dr R. H. Buckingham
Institut de Biologie Physico-chimique
13, rue Pierre et Marie Curie
75005 Paris
France

Dr M. Ehrenberg
Department of Molecular Biology
Uppsala University
Biomedical Centre, Box 590
S-751 24 Uppsala
Sweden

Professor A. R. Fersht
Department of Chemistry
Imperial College of Science and Technology
London SW7 2AY
UK

Professor D. J. Galas
Molecular Biology, ACBR 126
University of Southern California
Los Angeles
California 90007
USA

Professor J. A. Gallant
Department of Genetics SK-50
University of Washington
Seattle
Washington 98195
USA

Dr M. F. Goodman
University of Southern California
Department of Biological Sciences
Molecular Biology Section
Los Angeles
California 90089–1481
USA

Dr H. Grosjean
Associated Professor
University of Brussels
Co-director
Biological Chemistry Laboratory
Department of Molecular Biology
67, rue des Chevaux
B-1640 Rhode-St-Genese
Belgium

Dr R. Holliday
Genetics Division
National Institute for Medical Research
The Ridgeway
Mill Hill
London NW7 1AA
UK

Dr T. B. L. Kirkwood
Computing Laboratory
National Institute for Medical Research
The Ridgeway
Mill Hill
London NW7 1AA
UK

Professor C. G. Kurland
Department of Molecular Biology
Uppsala University
Biomedical Centre, Box 590
S-751 24 Uppsala
Sweden

Professor J. R. Menninger
Department of Biology
University of Iowa
Iowa City
Iowa 52242
USA

Dr J. Ninio
Institut Jacques Monod
Tour 43, 2 Place Jussieu
75251 Paris cedex 05
France

Professor M. I. Page
Department of Chemical Sciences
The Polytechnic
Huddersfield HD1 3DH
UK

Dr R. F. Rosenberger
Genetics Division
National Institute for Medical Research
The Ridgeway
Mill Hill
London NW7 1AA
UK

Dr S. G. Sedgwick
Genetics Division
National Institute for Medical Research
The Ridgeway
Mill Hill
London NW7 1AA
UK

Preface

Molecular biology proceeds at unremitting pace to unfold new secrets of the living world. Biology, long regarded as an inexact companion to physics and chemistry, has undergone transformation. Now, chemical and physical principles are tools in understanding highly complex biomolecular processes, whose origin lies in a history of chance, constraint and natural selection. The accuracy of these processes, often remarkably high, is crucial to their self-perpetuation, both individually and collectively, as ingredients of the organism as a whole.

In this book are presented thirteen chapters which deal with various facets of the accuracy problem. Subjects covered include: the specificity of enzymes; the fidelity of synthesis of proteins; the replication and repair of DNA: general schemes for the enhancement of biological accuracy; selection for an optimal balance between the costs and benefits of accuracy; and the possible relevance of molecular mistakes to the process of ageing. The viewpoints are distinct, yet complementary, and the book as a whole offers to researchers and students the first comprehensive account of this growing field.

The idea of a book on accuracy in molecular processes was inspired first by a workshop organized in 1978 by Jacques Ninio (a contributor to this volume) with the sponsorship of the European Molecular Biology Organization. So successful was this meeting that two further workshops on similar lines were held in 1981 and 1985. Many of the contributors to this book participated in these workshops, and the book has benefited substantially from the sustained, informal exchange of views which the workshops have helped to bring about. The book is entirely independent, however, of these conferences.

We are grateful to all contributors for the care and patience with which they have written and, where necessary, revised their chapters, and to numerous of our colleagues for helpful comments and suggestions. We thank, in particular, Dr Alan Crowden of Chapman and Hall for his support and encouragement.

T. B. L. Kirkwood
R. F. Rosenberger
D. J. Galas

Acknowledgements

The following acknowledgements are made by authors with reference to their individual chapters

R. H. Buckingham and H. Grosjean
(Chapter 5)

The authors thank Drs C. G. Kurland, E. Murgola and J. Ninio for stimulating discussions during the writing of Chapter 5. We acknowledge Dr M. Grunberg-Manago and Professor H. Chantrenne for their interest and encouragement and the following organizations for financial support: in France, the Centre National de la Recherche Scientifique (Groupe de Recherche no. 18, A.T.P. 'Biologie Moléculaire du Gène', 'Microbiologie 1979' et 'Internationale 1980'), the Délégation Générale à la Recherche Scientifique et Technique (Convention 80.E. 0872), the Ligue Nationale contre le Cancer (Comité de la Seine) and the Commissariat à l'Energie Atomique; in Belgium, 'Actions concertées', Fonds National de la Recherche Scientifique Collective (contrat nî 2.4520.81) and Patrimoine de l'Université de Liège.

M. F. Goodman and E. W. Branscomb
(Chapter 8)

We wish to express appreciation to Dr John Petruska and Dr John J. Hopfield for the generous amounts of time spent instructing us in the subtleties and complexities of nucleotide base-pairings. We are grateful to Dr Christopher K. Mathews, Dr David Korn, Dr Martin Kamen, Dr Robert Ratliff, Dr Bruce Alberts, and Ms Randi Hopkins for their interest and help in numerous aspects of the work. Special thanks are due Dr Maurice J. Bessman and Dr Bernard C. Abbott. We also wish to thank Mrs Leilani Corell for expert and patient preparation of the 'diverging' series of drafts of this manuscript. The work was supported in part by grants GM21422 and CA17358 from the National Institutes of Health and grant 595 from the American Heart Association, Los Angeles Affiliate. The work was also performed in part under the auspices of the United States Department of Energy by the Lawrence Livermore National Laboratory under contract number W-7405-ENG-48 and Interagency Agreement 81-D-X0533 with the Environmental Protection Agency.

J. Ninio
(Chapters 10 and 13)

I am indebted to Leslie Orgel for his teachings in fundamental chemical kinetics. My thanks also to Manolo Gouy for checking all equations, to Pierre Claverie for discussions in chemical physics, to Eduardo Mizraji for contributing to my enzymological culture, to David Galas for suggestions concerning the manuscript and to Charles Kurland for his constant support within and outside the ribosome club.

I also wish to thank David Galas and Vic Norris for style suggestions for Chapter 13.

Contents

1 An introduction to the problem of accuracy

D. J. GALAS, T. B. L. KIRKWOOD
and R. F. ROSENBERGER

1.1 Setting the scene

The primary concern of this book is about how cells copy and maintain the information which is stored as base sequences in their DNA and how they use this information to specify the structure of proteins. It is generally accepted that these processes of information transfer are the most essential and basic functions any living organism has to perform.

Two of the outstanding features of the information transfer processes are the accuracy and the speed with which they operate. All of them involve selecting the correct monomer from a pool of quite similar molecular species, for example the right nucleic acid out of the four available alternatives, or the right amino acid out of the twenty present in the cytoplasm. In the most accurate of the operations, the replication of DNA, the process has the astonishingly low error rate of about one mistake in 10^8. Further, selection occurs at a speed which allows the polymerization of many monomers per second. The mechanisms used to effect polymer synthesis and the problems living cells encounter in doing this are discussed in depth in the following chapters. In the present introductory chapter, we will attempt to bring these systems into a general focus.

Historically, the first serious biological encounter with the concept of random change came with the intellectual ferment that brought forth the Darwinian theory of evolution (see Eisley, 1958). While Darwin discussed the importance of the random variations of characteristics of organisms and commented on the fundamental role that chance must therefore play in evolution, it awaited the rise of genetics and, finally, molecular biology for biologists to come to grips with the mechanistic reality of Darwin's essential variations. One of the earliest attempts to probe the accuracy of replication of the hereditary molecule was made, in fact, by a theoretical physicist (Schrödinger, 1944), who brought to bear the heady optimism of the new

quantum theory of physics on the nature of the genetic mechanism. These bold speculations helped to kindle the intellectual spirit that led in a few short years to the identification of Watson–Crick base-pairing as the chemical rule for replication of genetic information and the elucidation of the genetic code as the set of rules for translating genes into proteins.

At first, the focus of attention on these discoveries was, quite naturally, on the remarkable properties of base-pairing in providing deterministic chemical rules for the replication and translation of genetic information. The possibility of error was recognized, however, since without error there could be no mutation, and without mutation there would be no evolution. In relation to protein synthesis, Pauling (1957) pointed out that the molecular difference between the amino acids valine and isoleucine (see Chapter 3) was so small that it should be very difficult for the protein synthetic apparatus to discriminate sharply between them, as it apparently does. Pauling thereby posed in concrete terms the important question that most of the chapters deal with in some form, namely, what details of the mechanisms for information transfer are responsible for their extraordinary accuracy.

Shortly after Pauling's challenge, Loftfield (1963) showed that discrimination against valine was indeed much stronger than simple chemical differences would predict. This was followed by the discovery that the first steps in the charging of isoleucine tRNA were actually not very strongly discriminatory against valine, but that a subsequent step destroyed the activated valine (Baldwin and Berg, 1966). Subsequently, Yarus (1972a,b) and Eldred and Schimmel (1972) discovered that aminoacyl-tRNA ligases (synthetases) could actually deacylate their own tRNAs that are mischarged with an incorrect amino acid. Thus, the accuracy of tRNA charging was seen to be actively guarded by the charging enzymes (see Chapter 4). An analogous sort of active monitoring or 'proofreading' was proposed by Brutlag and Kornberg (1972) in the replication of DNA (see Chapter 8).

The accuracy of protein synthesis is, of course, not only due to the accuracy with which tRNAs are charged. Over a period of time parallel with the above studies on tRNA charging, firstly the ribosome was characterized and shown to be central to protein synthesis and its accuracy, secondly the phenomenon of 'informational suppression' was discovered (for review see Steege and Söll, 1979), and thirdly the misreading of codons induced by aminoglycoside antibiotics, like streptomycin, was discovered. These phenomena are discussed in Chapters 5 and 6.

Some of the first ideas on the role of the ribosome in determining the accuracy of protein synthesis were provided by the inventive and catalytic work of Luigi Gorini. In a series of highly original studies he examined the effects of aminoglycoside antibiotics and ribosomal mutations on nonsense and missense suppression and discussed the results in terms of the ability of the variously perturbed ribosomes to discriminate among tRNAs and termin-

ation factors. He showed that wild type ribosomes were significantly less accurate than they could become by acquiring certain mutations in ribosomal proteins, and particularly striking was his discovery of mutants with ribosomes more error-prone than the wild type (Rossett and Gorini, 1969). Thus, it was established that the translation error level was readily genetically manipulable and also that the bacterial cell could tolerate an increased level of errors. These systems continue to reveal valuable insights into the mechanisms for control of accuracy (see Chapters 6 and 11).

The recent history of research into questions of molecular accuracy has sounded two principal themes: (1) an ever more detailed analysis of the molecular structures and the kinetics involved in determining accuracy (see Chapters 3, 10 and 11) and (2) a convergence and cross-fertilization of ideas found useful in the various realms of molecular accuracy in biology – transcription (Chapter 7), translation (Chapters 5 and 6), charging of tRNAs (Chapter 4), DNA replication and repair (Chapters 8 and 9), and the substrate specificity of various enzymes (Chapter 3). Attention has also been paid to the integrity of the genetic information transfer system as a whole (Chapter 2 and see Fig. 1.1). As early as 1963, Orgel posed the question: how can the translation process, which itself is mediated by highly specific proteins, be stable against the feedback of mistakes in protein synthesis (Orgel, 1963). The notion of potential instability in the cellular translation process is important to the question of how life, with a stable translation system, emerged in the first place (see below and Chapter 13) and has also been suggested to have relevance to the process of ageing (Orgel, 1963). Since

Replication	Transcription	Translation
DNA polymerase	RNA polymerase	Ribosomes, aminoacyl-tRNA synthetases, etc.

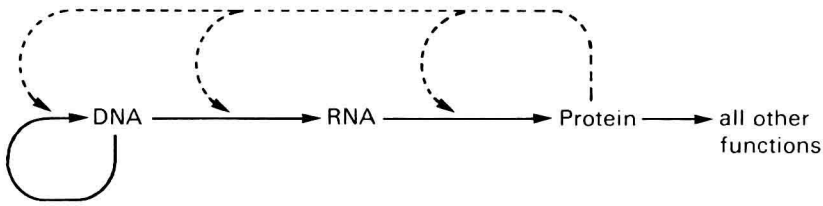


Figure 1.1 Schematic representation of the main pathways of genetic information transfer. The continuous arrows indicate the replication of DNA, the transcription of DNA into RNA, and the translation of RNA into proteins. The dashed arrows indicate the participation of *products* of genetic translation in these processes.

accuracy of information transfer and its associated costs are so fundamental in evolution, the possible role of accuracy breakdown in ageing has stimulated a substantial amount of research (Chapter 12).

1.2 Some preliminary concepts

1.2.1 DEFINITIONS OF ACCURACY

There are several possible ways to define the accuracy of a molecular process. First, the kind of error must be specified. In most of the instances considered in this book, the basic error will be insertion of an incorrect, or noncognate, monomer into a growing polymeric chain in place of the correct, or cognate, monomer. Second, one must be clear whether it is the insertion of a *particular* noncognate monomer which is of interest, or whether all possible misinsertions are to be considered. If the latter, one should be aware that different misinsertions will not usually be made at equal rates. Third, there is the possibility, supported by some experimental evidence (see Chapters 5 and 6), that the error rate for insertion of any given monomer will be influenced by the neighbouring sequence.

Let us assume that the basic error rate for insertion of monomers into a growing polymeric chain is e (misinsertions/insertion), ignoring possible differences in the error frequencies for different monomers. The proportion of correctly synthesized polymers will depend on their lengths. For a polymer made up of N monomeric units, the proportion of correctly synthesized polymers will be $(1-e)^N$, and the proportion of polymers containing one or more incorrect monomers will be $E_N = 1-(1-e)^N$. Table 1.1 shows values of E_N for various values of e and N. Generally, it may be seen that for any given value of e, the proportion of error-containing polymers rises sharply with increasing N. This is likely to be why the larger proteins, for example, are usually made up of smaller subunits. At a basic error rate of $e = 10^{-4}$–10^{-3} (Loftfield, 1963), synthesis of a very long protein molecule without error

Table 1.1 Frequencies of error-containing polymers (E_N) as a function of the frequency of inserting incorrect monomers (e) and the length of the polymer (N)

	e			
N	10^{-2}	10^{-3}	10^{-4}	10^{-5}
10	0.096	0.010	0.001	0.000
30	0.260	0.030	0.003	0.000
100	0.634	0.095	0.010	0.001
300	0.951	0.260	0.030	0.003
1000	1.000	0.632	0.095	0.010