

Contents

Foreword	1
Chapter 1 The Logic of Genetic Code	1
§ 1.1 The Degeneracy Rule of the Genetic Code	2
§ 1.2 The Distribution of Stop Codons in the Genetic Code	14
§ 1.3 The Hydrophobic-hydrophilic Domain in the Code Table	18
§ 1.4 The Global Minimum of Mutational Deterioration and the Deduction of Standard Code Table	27
§ 1.5 The Multiplicity of Amino Acid and the Abnormal Genetic Code	36
§ 1.6 The Symmetry and Duality of the Genetic Code	42
§ 1.7 A Statistical Theory of Amino Acid Mutation	52
References	60
Chapter 2 Gene Informatics and Sequence Evolution (I)	63
§ 2.1 The Evolution of Protein Sequence and Rate of Molecular Evolution	65
§ 2.2 Statistical Correlation of Nucleotides and Its Short- range-dominance	68
§ 2.3 The Evolutionary Relation of Gene Information Parameters	83
§ 2.4 Preferential Modes in DNA Sequence	89
§ 2.5 Molecular Evolutionary Tree	104
§ 2.6 The Spectral Analysis of DNA Sequence and Long-range Correlation of Bases	124
§ 2.7 DNA Walk and the Diagrammatic Representation for DNA Sequence	140
§ 2.8 The Markovian Chain and Probability Matrix of Nucleotide Sequence	150

Chapter 3 Gene Informatics and Sequence Evolution (II)	159
§ 3.1 Gene Expression, Codon Usage and Base Correlation	159
§ 3.2 Statistical Investigation on Nucleotide Distribution in Different Regions of Genes	177
§ 3.3 Open-reading-frame Organization and Gene Recognition in Several Model Genomes of Lower Organisms	188
§ 3.4 Methods for Identification of Protein-Coding Genes	207
§ 3.5 Evolutionary Mechanism of Gene Sequence — Maximum Information Principle	227
§ 3.6 Evolutionary Equation of DNA Sequence	235
§ 3.7 Minimal Model for Genome Evolution and Growth	256
§ 3.8 The Complexity of Genetic Language — A Piece of Dark Cloud in the Blue Sky of the 21st Century Science	265
References	279
Chapter 4 From Sequence to Structure	286
§ 4.1 Statistical Mechanical Approach to the Empirical Prediction of Protein Secondary Structure	288
§ 4.2 Information-theoretic Approach to Structure Prediction and Sequence-structure Relation Viewed from Cryptography	296
§ 4.3 Relationship between Protein Secondary Structure and Synonymous Codon Usage	313
§ 4.4 Relations of Protein Secondary Structure with Codon tRNA Abundance and mRNA Structure	324
§ 4.5 Lattice Simulation of Nascent Peptide Folding	354
§ 4.6 Regular Secondary Structure Clustering and the Architecture of Framework Structure of Protein	366
§ 4.7 Recognition and Prediction of Protein Structural Class and Topological Structure	380
§ 4.8 DNA Curvature and Flexibility	408
References	428
Chapter 5 Approaches to Biological Function	435
§ 5.1 Conformation Dynamics of Biological Macromolecules	437
§ 5.2 The Rate of Protein Folding	457
§ 5.3 Protein Denaturation and Cooperativity	464

determining the parabolic motion on the Earth belong to the same force — Universal Gravitational Force? It is the integration of experimentalism and rationalism that constitutes the recent four hundred years scientific tradition and paradigm. This integration is represented most fully and fruitfully in physics, especially in the 20th century's modern physics. With invincible power, it is conquering every aspect of natural science. The permeation of this paradigm into life science is thus inevitable. As the result of this diffusion, life science must be not only experimental but also theoretical and comprehensive.

This book is the summary of our group's work in theoretical biology in the past 15 years (1987—2002). By choosing *Theoretic-Physical Approach* as its title instead of using a more common one *Theoretical approach*, the author intends to emphasize the significance of exploring the essence of the phenomena and finding the 'physics' behind them rather than only resorting to mathematical methods. Richard Feynman once said, "Physicists always have the habit of taking the simplest example of any phenomenon and calling it 'physics'." Although the life phenomena are extremely complex, we expect that this method of constructing the simplified model, neglecting all non-essential factors and extracting the 'coarse-grained' laws for the model system, can work effectively on the main topics of molecular biology. The characteristic feature of life which differentiates from an inanimate piece of matter is the large amount of information contained in it. Different from "matter" and "energy", "information" constitutes the third category in natural sciences. The central task of a mature theoretical approach to molecular biology thus lies in exploring the law on the formation, storage, expression and transmission of life information in each step, from DNA to mRNA, to amino acid sequence and finally to protein structure and function. Of course, to attain the goal, there is still a long way to go. The choice of the subject matter in this book only reflects on the author's personal point of view and his present research experience. Therefore it cannot be expected to cover every aspect of this field. In fact, in such a broad and rapidly developing field, even a relative comprehensiveness is hard to achieve. So it is probably more proper to consider this book a research report on the progress of theoretical biology.

Unexceptionally, every mature discipline has its own relatively stable scientific community and a rather consistent value system. Being a new discipline, theoretical biology is forming its community. I consider myself