

Contents

Preface to the first edition, ix
Preface to the second edition, xi
Acknowledgements, xiii

Introduction, xv

1 Protein-Coding Genes, 1

1.1 Structure of protein-coding genes, 2
1.2 Transcription, 3
1.3 Translation, 10
References, 14
Useful internet resources, 15

2 Protein Structure, 18

2.1 The polypeptide backbone, 18
2.2 The amino acids, 18
2.3 Covalent modifications of amino acid side chains, 23
2.3.1 Enzymatic modifications, 23
2.3.2 Nonenzymatic chemical modifications, 27
2.4 Interactions that govern protein folding and stability, 28
2.4.1 Noncovalent interactions, 28
2.4.2 The hydrophobic interaction, 29
2.5 Secondary structural elements, 30
2.5.1 The α -helix, 30
2.5.2 β -sheets, 31
2.5.3 Reverse turns, 32
2.6 Supersecondary structures, 32
2.7 Tertiary structures of proteins, 33
2.7.1 Globular proteins, 33
2.7.2 Fibrous proteins, 35
2.7.3 Unusual structures of internally repeated proteins, 36
2.7.4 Secreted proteins and membrane proteins, 37
2.7.5 Intrinsically disordered proteins, 40
2.8 Multidomain proteins, 40
2.9 Multisubunit proteins, 40
References, 41
Useful internet resources, 41

3 Mutations, 45

3.1 Types of mutations, 45
3.1.1 Substitutions, 45
3.1.2 Deletion, duplication, insertion and fusion, 48
3.2 Factors affecting rates of mutation, 50
3.3 The fate of mutations, 52
3.4 The molecular clock, 59

References, 61

Useful internet resources, 62

4 Evolution of Protein-Coding Genes, 63

4.1 Alignment of nucleotide and amino acid sequences, 63

4.2 Estimating the number of nucleotide substitutions, 65

4.2.1 Substitutions in translated regions, 67

4.2.2 Substitutions in untranslated regions, introns and 5' and 3' flanking regions of protein-coding genes, 68

4.3 Rates and patterns of nucleotide substitution, 68

4.3.1 Rates of nucleotide substitution, 69

4.4 Variation in substitution rates, 71

4.4.1 Variation among different sites of the translated region, 72

4.4.2 Variation among genes, 74

4.4.3 Constancy and variation in substitution rates of orthologous genes, 75

4.4.4 Nonrandom substitutions at synonymous positions, 76

4.5 Molecular phylogeny, 78

4.5.1 Phylogenetic trees, 78

4.5.2 Tree reconstruction, 79

4.5.3 Tree-making methods, 81

4.5.4 Estimation of species-divergence times, 85

References, 85

Useful internet resources, 87

5 Evolution of Orthologous Proteins, 89

5.1 Orthologous proteins with the same function in different species, 91

5.2 Orthologous proteins with modified function in different species, 94

5.3 Orthologous proteins with major modification of function, 98

5.4 Orthologous proteins that have lost their function, 98

5.5 Orthologous proteins that have gained additional functions, 99

5.6 Prediction of the function of orthologous proteins, 99

5.7 The three-dimensional structure of orthologous proteins, 100

5.7.1 Prediction of secondary structure of proteins, 100

5.7.2 Prediction of the three-dimensional structure of proteins, 102

5.8 Detecting sequence homology of protein-coding genes, 103

References, 104

Useful internet resources, 105

6 Formation of Novel Protein-Coding Genes, 108

6.1 *De novo* formation of novel protein-coding genes, 108

6.2 Gene duplications, 110

6.2.1 Mechanisms of gene duplication, 111

6.2.2 Fate of duplicated genes, 117

6.2.3 Fate of genes acquired by lateral gene transfer, 121

6.2.4 Dating gene duplications, 121

References, 124

Useful internet resources, 125

7 Evolution of Paralogous Proteins, 126

7.1 Advantageous duplications, 127

7.1.1 Unprocessed genes, 127

7.1.2 Processed genes, 127

- 7.2 Neutral duplications, 130
 - 7.2.1 Modification of function by point mutations, 132
 - 7.2.2 Major change of function by point mutations, 140
 - 7.2.3 Major change of function by domain acquisitions, 143
- 7.3 Similarities and differences in the evolution of paralogous and orthologous proteins, 147
- 7.4 Predicting the function of proteins by homology, 150
- 7.5 Nonhomology-based methods for the prediction of the function of proteins, 151
- 7.6 Detecting distant homology of protein-coding genes, 152
 - 7.6.1 Detecting distant homology by consensus approaches, 152
 - 7.6.2 Detecting distant homology by comparing three-dimensional structures, 161
 - 7.6.3 Detecting distant homology by comparing exon-intron structures, 162
- References, 163
- Useful internet resources, 166

8 Protein Evolution by Assembly from Modules, 170

- 8.1 Modular assembly by intronic recombination, 171
 - 8.1.1 Introns, 173
 - 8.1.2 Internal gene duplications/deletions via recombination in introns, 182
 - 8.1.3 Fusion of genes via recombination in introns, 183
 - 8.1.4 Exon shuffling via recombination in introns, 183
 - 8.1.5 Factors affecting acceptance of mutants created by intronic recombination, 192
 - 8.1.6 Classification of modules and mosaic proteins produced by exon shuffling, 199
 - 8.1.7 Genome evolution and the evolution of exon shuffling, 207
 - 8.1.8 Evolutionary significance of exon shuffling, 209
 - 8.1.9 Genome evolution and the evolution of alternative splicing, 211
- 8.2 Modular assembly by exonic recombination, 211
- References, 213
- Useful internet resources, 216

9 Genome Evolution and Protein Evolution, 218

- 9.1 Evolution of genome size, 218
- 9.2 The role and survival of nongenic DNA, 221
- 9.3 Repetitiveness of genomic DNA, 221
- 9.4 Mechanisms responsible for increases in genome size, 223
- 9.5 Compositional organization of eukaryotic genomes, 224
- 9.6 Genomes of model organisms, 225
 - 9.6.1 Viral genomes, 226
 - 9.6.2 Cellular genomes, 230
 - 9.6.2.1 Eubacterial genomes, 231
 - 9.6.2.2 Archaeal genomes, 238
 - 9.6.2.3 Organelle genomes, 241
 - 9.6.2.4 Eukaryotic genomes, 244
 - 9.6.2.5 Genome duplications in the evolution of early vertebrates, 271
 - 9.6.3 Value of comparative genomics for the identification of functional elements, 276
 - 9.6.4 Finding protein-coding genes in genome sequences, 277
- 9.7 The genome of the cenacestator, 280
- 9.8 Changes in gene number and gene density in different evolutionary lineages, 280

9.9 Proteome evolution, 282

9.9.1 Proteome evolution – classification of proteins by structural features, 282

9.9.2 Proteome evolution – classification of proteins by homology, 283

9.9.3 Proteome evolution – classification of proteins by function, 283

9.9.4 Proteome evolution – evolution of proteome complexity, 287

9.9.5 Proteome evolution and organismic complexity, 291

References, 293

Useful internet resources, 302

Glossary, 309**Index, 367**