# Contents

## Part I Molecules: Proteins and RNA

1 Modeling Conformational Flexibility and Evolution of Structure: RNA as an Example

*P. Schuster and P.F. Stadler* ........................................ 3

1.1 Definition and Computation of RNA Structures .......... 3

1.1.1 RNA Secondary Structures ......................... 4

1.1.2 Compatibility of Sequences and Structures ........ 8

1.1.3 Sequence Space, Shape Space, and Conformation Space ........ 11

1.1.4 Computation of RNA Secondary Structures .......... 14

1.1.5 Mapping Sequences into Structures ................. 15

1.1.6 Suboptimal Structures and Partition Functions .... 18

1.2 Design of RNA Structures .................................. 19

1.2.1 Inverse Folding .................................. 19

1.2.2 Multiconformational RNAs ......................... 20

1.2.3 Riboswitches .................................. 22

1.3 Processes in Conformation, Sequence, and Shape Space .... 23

1.3.1 Kinetic Folding ................................ 23

1.3.2 Evolutionary Optimization ......................... 25

1.3.3 Evolution of Noncoding RNAs ....................... 30

References .................................................. 32

2 Gene3D and Understanding Proteome Evolution

*J.G. Ranea, C. Yeats, R. Marsden, and C. Orengo* ................... 37

2.1 Protein Family Clustering .................................. 42

2.1.1 SYSTERS ...................................... 42

2.1.2 ProtoNet ...................................... 42

2.1.3 ADDA ....................................... 42

2.1.4 ProDom ...................................... 43
4.2 Structural Aspects of Molecular Evolution
   4.2.1 Neutral Theory and Protein Folding Thermodynamics
   4.2.2 Structural Conservation and Functional Changes in Protein Evolution
   4.2.3 Models of Molecular Evolution with Structural Conservation

4.3 The SCN Model of Evolution
   4.3.1 Representation of Protein Structures
   4.3.2 Stability Against Unfolding
   4.3.3 Stability Against Misfolding
   4.3.4 Calculation of $\alpha(A)$
   4.3.5 Sampling the Neutral Networks
   4.3.6 Fluctuations and Correlations in the Evolutionary Process
   4.3.7 Substitution Process

4.4 Site-Specific Amino Acid Distributions
   4.4.1 Vectorial Representation of Protein Sequences
   4.4.2 Vectorial Representation of Protein Folds
   4.4.3 Relation Between Sequence and Structure
   4.4.4 The PE as a Structural Determinant of Evolutionary Conservation
   4.4.5 Site-Dependent Amino Acid Distributions
   4.4.6 Sequence Conservation and Structure Designability
   4.4.7 Site-Specific Amino Acid Distributions in the PDB
   4.4.8 Mean-Field Model of Mutation plus Selection

4.5 Conclusions

References

5 Towards Unifying Protein Evolution Theory

N.V. Dokholyan and E.I. Shakhnovich

5.1 Two Views on Protein Evolution

5.2 Challenges in Functionally Annotating Structures

5.3 The Importance of the Tree of Life

5.4 Building the PDUG

5.5 Properties of the PDUG:
   Power Laws on Very Different Evolutionary Scales

5.6 Functional Flexibility Score:
   Calculating Entropy in Function Space

5.7 Lattice Proteins and Its Random Subspaces: Structure Graphs

5.8 Divergence and Convergence Explored:
   What Power Laws Tell Us about Evolution

5.9 Context Is Important

5.10 Not All Functions Are Created Equal and Neither Are Structures
Contents

5.11 Concluding Remarks .............................. 124
References .................................................. 124

Part II Molecules: Genomes

6 A Twenty-First Century View of Evolution: Genome System Architecture, Repetitive DNA, and Natural Genetic Engineering
J.A. Shapiro ................................................ 129
6.1 Introduction: Cellular Computation and DNA as an Interactive Data Storage Medium ............. 129
6.2 Genome System Architecture and Repetitive DNA ................................................... 130
6.3 Genomes and Cellular Computation: E. coli lac Operon ........................................ 132
6.4 New Principles of Evolution: The Lessons of Sequenced Genomes ............................... 135
6.5 Natural Genetic Engineering ............................................................... 136
6.6 Conclusions: A Twenty-First Century View of Evolution ........................................ 141
6.7 Twenty-First Century Directions in Evolution Research ........................................... 143
References ...................................................... 144

7 Genomic Changes in Bacteria: From Free-Living to Endosymbiotic Life
F.J. Silva, A. Latorre, L. Gómez-Valero, and A. Moya .................. 149
7.1 Introduction .............................................. 149
7.2 Genetic and Genomic Features of Endosymbiotic Bacteria .......................................... 153
7.2.1 Sequence Evolution in Endosymbionts ........................................ 153
7.2.2 Reductive Evolution: DNA Loss and Genome Reduction in Obligate Bacterial Mutualists ........................................ 158
7.2.3 Chromosomal Rearrangements Throughout Endosymbiont Evolution ......................... 160
7.3 Conclusions and Prospects .............................................. 162
References ...................................................... 163

Part III Phylogenetic Analysis

8 Molecular Phylogenetics: Mathematical Framework and Unsolved Problems
X. Xia ......................................................... 169
8.1 Introduction .............................................. 169
8.2 Substitution Models ........................................ 170
8.2.1 Nucleotide-Based Substitution Models and Genetic Distances .............................. 171
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.2.2</td>
<td>Amino Acid-Based and Codon-Based Substitution Models</td>
<td>176</td>
</tr>
<tr>
<td>8.3</td>
<td>Tree-Building Methods</td>
<td>178</td>
</tr>
<tr>
<td>8.3.1</td>
<td>Distance-Based Methods</td>
<td>178</td>
</tr>
<tr>
<td>8.3.2</td>
<td>Maximum Parsimony Methods</td>
<td>181</td>
</tr>
<tr>
<td>8.3.3</td>
<td>Maximum Likelihood Methods</td>
<td>182</td>
</tr>
<tr>
<td>8.3.4</td>
<td>Bayesian Inference</td>
<td>185</td>
</tr>
<tr>
<td>8.4</td>
<td>Final Words</td>
<td>187</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>187</td>
</tr>
<tr>
<td>9</td>
<td>Phylogenetics and Computational Biology of Multigene Families</td>
<td>191</td>
</tr>
<tr>
<td>9.1</td>
<td>Introduction</td>
<td>191</td>
</tr>
<tr>
<td>9.2</td>
<td>How Do Large Gene Families Arise?</td>
<td>193</td>
</tr>
<tr>
<td>9.3</td>
<td>The Classical Model of Gene Duplication</td>
<td>193</td>
</tr>
<tr>
<td>9.4</td>
<td>Subfunctionalization Model</td>
<td>194</td>
</tr>
<tr>
<td>9.5</td>
<td>Subneofunctionalization</td>
<td>195</td>
</tr>
<tr>
<td>9.6</td>
<td>Tests for Subfunctionalization</td>
<td>196</td>
</tr>
<tr>
<td>9.7</td>
<td>Tests for Functional Divergence After Duplication</td>
<td>196</td>
</tr>
<tr>
<td>9.7.1</td>
<td>Case Study 1: Chemokine Receptors</td>
<td>197</td>
</tr>
<tr>
<td>9.7.2</td>
<td>Case Study 2: The Evolution of TIM Barrel Coding Genes</td>
<td>199</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>204</td>
</tr>
<tr>
<td>10</td>
<td>SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis</td>
<td>207</td>
</tr>
<tr>
<td>10.1</td>
<td>Introduction</td>
<td>207</td>
</tr>
<tr>
<td>10.1.1</td>
<td>About R and CRAN</td>
<td>207</td>
</tr>
<tr>
<td>10.1.2</td>
<td>About this Document</td>
<td>208</td>
</tr>
<tr>
<td>10.1.3</td>
<td>About Sequin and seqinR</td>
<td>208</td>
</tr>
<tr>
<td>10.1.4</td>
<td>About Getting Started</td>
<td>208</td>
</tr>
<tr>
<td>10.1.5</td>
<td>About Running R in Batch Mode</td>
<td>208</td>
</tr>
<tr>
<td>10.1.6</td>
<td>About the Learning Curve</td>
<td>209</td>
</tr>
<tr>
<td>10.2</td>
<td>How to Get Sequence Data</td>
<td>213</td>
</tr>
<tr>
<td>10.2.1</td>
<td>Importing Raw Sequence Data from Fasta Files</td>
<td>213</td>
</tr>
<tr>
<td>10.2.2</td>
<td>Importing Aligned Sequence Data</td>
<td>214</td>
</tr>
<tr>
<td>10.2.3</td>
<td>Complex Queries in ACNUC Databases</td>
<td>218</td>
</tr>
<tr>
<td>10.3</td>
<td>How to Deal with Sequence</td>
<td>220</td>
</tr>
<tr>
<td>10.3.1</td>
<td>Sequence Classes</td>
<td>220</td>
</tr>
<tr>
<td>10.3.2</td>
<td>Generic Methods for Sequences</td>
<td>220</td>
</tr>
<tr>
<td>10.3.3</td>
<td>Internal Representation of Sequences</td>
<td>221</td>
</tr>
</tbody>
</table>
11 Evolutionary Genomics of Gene Expression
I.K. Jordan and L. Marín-Ramírez .............................................. 235
11.1 Sequence Divergence ....................................................... 236
11.1.1 Ortholog Identification ............................................... 236
11.1.2 Sequence Alignment .................................................... 237
11.1.3 Sequence Distance Calculation ........................................ 237
11.2 Gene Expression Divergence ............................................. 240
11.2.1 Database Sources ....................................................... 241
11.2.2 Probe-to-Gene Mapping ............................................... 241
11.2.3 Structure of the Data ................................................... 242
11.2.4 Transformation and Normalization ..................................... 242
11.2.5 Measuring Divergence .................................................. 243
11.2.6 Clustering and Visualization .......................................... 245
11.3 Integrated Analysis ....................................................... 246
11.3.1 Sequence vs. Expression Divergence ................................ 246
11.3.2 Neutral Changes in Gene Expression ................................. 247
11.3.3 Evolutionary Conservation of Gene Expression ................. 250
References ................................................................. 251

12 From Biophysics to Evolutionary Genetics:
Statistical Aspects of Gene Regulation
M. Lässig ................................................................. 253
12.1 Introduction .............................................................. 253
12.2 Biophysics of Transcriptional Regulation .............................. 254
12.2.1 Factor-DNA Binding Energies ...................................... 255
12.2.2 Energy Distribution in the Genome .................................. 257
12.2.3 Search Kinetics ....................................................... 258
12.2.4 Thermodynamics of Factor Binding ................................ 258
12.2.5 Sensitivity and Genomic Design of Regulation ................. 260
12.2.6 Programmability and Evolvability of Regulatory Networks ........................................ 260
12.3 Bioinformatics of Regulatory DNA .................................... 261
12.3.1 Markov Model for Background Sequence .......................... 261
12.3.2 Probabilistic Model for Functional Sites ............................ 262
12.3.3 Bayesian Model for Genomic Loci .................................. 263
12.3.4 Dynamic Programming and Sequence Analysis .................. 264
12.4 Evolution of Regulatory DNA ......................................... 266
12.4.1 Deterministic Population Dynamics and Fitness ............... 267
# Contents

14.3.4 Neutrality ................................................................. 326
14.4 Dynamics of Adaptation ................................................ 327
  14.4.1 Peak Shifts and Punctuated Evolution ......................... 328
  14.4.2 Evolutionary Trajectories for the Quasispecies Model .... 328
  14.4.3 Dynamics in Smooth Fitness Landscapes ...................... 332
14.5 Evolution in the Laboratory ........................................... 333
  14.5.1 RNA Evolution In Vitro ............................................ 333
  14.5.2 Quasispecies Formation in RNA Viruses ..................... 334
  14.5.3 Dynamics of Microbial Evolution ............................... 334
14.6 Conclusions ................................................................. 335
References ............................................................................. 336

# Genetic Variability in RNA Viruses:
Consequences in Epidemiology and in the Development
of New Strategies for the Extinction of Infectivity

15.1 Introduction .................................................................. 341
15.2 Replication of RNA Viruses
  and Generation of Genetic Variability ................................. 343
15.3 Structure of Viral Populations ........................................ 344
15.4 Viral Quasi-Species and Adaptation ................................ 345
15.5 Population Dynamics of Host–Pathogen Interactions .......... 348
15.6 The Limit of the Error Rate ............................................ 350
  15.6.1 Increases in the Error Rate of Replication
    Lethal Mutagenesis As a New Antiviral Strategy ............... 352
  15.6.2 Evolution of Viral Populations
    Through Successive Bottlenecks ...................................... 355
15.7 Conclusions ................................................................ 359
References ............................................................................. 360

Index ...................................................................................... 363
Structural Approaches to Sequence Evolution
Molecules, Networks, Populations
Bastolla, U.; Porto, M.; Roman, E.; Vendruscolo, M.
(Eds.)
2007, XIX, 367 p., Hardcover
ISBN: 978-3-540-35305-8