
Contents

Outline	1
----------------------	---

Part I Mutations and Alignments

1 Introduction	5
1.1 Mutation and Alignment	5
1.1.1 Classification of Biological Sequences	5
1.1.2 Definition of Mutations and Alignments	6
1.1.3 Progress on Alignment Algorithms and Problems to Be Solved	8
1.1.4 Mathematical Problems Driven by Alignment and Structural Analysis	12
1.2 Basic Concepts in Alignment and Mathematical Models	13
1.2.1 Mutation and Alignment	13
1.3 Dynamic Programming-Based Algorithms for Pairwise Alignment	17
1.3.1 Introduction to Dynamic Programming-Based Algorithms	17
1.3.2 The Needleman–Wunsch Algorithm, the Global Alignment Algorithm	18
1.3.3 The Smith–Waterman Algorithm	21
1.4 Other Notations	24
1.4.1 Correlation Functions of Local Sequences	24
1.4.2 Pairwise Alignment Matrices Among Multiple Sequences	25
1.5 Remarks	26
1.6 Exercises, Analyses, and Computation	27

2 Stochastic Models of Mutations and Structural Analysis . . . 29

2.1 Stochastic Sequences and Independent Sequence Pairs 29

2.1.1 Definitions and Notations of Stochastic Sequences 29

2.1.2 Independently and Identically Distributed Sequences 31

2.1.3 Independent Stochastic Sequence Pairs 33

2.1.4 Local Penalty Function and Limit Properties
of 2-Dimensional Stochastic Sequences 36

2.2 Stochastic Models of Flow Raised by Sequence Mutations 37

2.2.1 Bernoulli Processes 37

2.2.2 Poisson Flow 40

2.2.3 Mutated Flows Resulting from the Four Mutation
Types 43

2.3 Stochastic Models of Type-I Mutated Sequences 45

2.3.1 Description of Type-I Mutation 45

2.3.2 Properties of Type-I Mutations 47

2.4 Type-II Mutated Sequences 50

2.4.1 Description of Type-II Mutated Sequences 51

2.4.2 Stochastic Models of Type-II Mutated Sequences 51

2.4.3 Error Analysis of Type-II Mutated Sequences 54

2.4.4 The Mixed Stochastic Models Caused by Type-I
and Type-II Mutations 57

2.5 Mutated Sequences Resulting from Type-III and Type-IV
Mutations 58

2.5.1 Stochastic Models of Type-III and Type-IV Mutated
Sequences 58

2.5.2 Estimation of the Errors Caused by Type-III
and Type-IV Mutations 59

2.5.3 Stochastic Models of Mixed Mutations 61

2.6 Exercises 64

3 Modulus Structure Theory 67

3.1 Modulus Structure of Expanded and Compressed Sequences . . 67

3.1.1 The Modulus Structures of Expanded Sequences
and Compressed Sequences 67

3.1.2 The Order Relation and the Binary Operators
on the Set of Expanded Modes 71

3.1.3 Operators Induced by Modes 73

3.2 Modulus Structure of Sequence Alignment 76

3.2.1 Modulus Structures Resulting from Multiple Alignment 76

3.2.2 Structure Analysis of Pairwise Alignment 78

3.2.3 Properties of Pairwise Alignment 81

3.2.4 The Order Relation and the Operator Induced
by Modulus Structure 84

3.3	Analysis of Modulus Structures Resulting from Sequence Mutations	85
3.3.1	Mixed Sequence Mutations	85
3.3.2	Structure Analysis of Purely Shifting Mutations	87
3.3.3	Structural Representation of Mixed Mutation	93
3.4	The Binary Operations of Sequence Mutations	93
3.4.1	The Order Relationship Among the Modes of Shifting Mutations	93
3.4.2	Operators Induced by Modes of Shifting Mutations	96
3.5	Error Analysis for Pairwise Alignment	100
3.5.1	Uniform Alignment of Mutation Sequences	100
3.5.2	Optimal Alignment and Uniform Alignment	102
3.5.3	Error Analysis of Uniform Alignment	104
3.5.4	Local Modification of Sequence Alignment	106
3.6	Exercises	107
4	Super Pairwise Alignment	109
4.1	Principle of Statistical Decision-Based Algorithms for Pairwise Sequences	109
4.1.1	Uniform Alignment and Parameter Estimation for Pairwise Sequences	109
4.1.2	The Locally Uniform Alignment Resulting from Local Mutation	111
4.1.3	The Estimations of Mutation Position and Length	113
4.2	Operation Steps of the SPA and Its Improvement	115
4.2.1	Operation Steps of the SPA	115
4.2.2	Some Unsolved Problems and Discussions of SPA	118
4.2.3	Local Modifications for Sequence Alignment	121
4.3	Index Analysis of SPA	122
4.3.1	The Statistical Features of Estimations	122
4.3.2	Improvement of the Algorithm to Estimate \hat{s}^*	128
4.3.3	The Computational Complexity of the SPA	131
4.3.4	Estimation for the Error of Uniform Alignment Induced by a Hybrid Stochastic Mutation Sequence	132
4.4	Applications of Sequence Alignment and Examples	135
4.4.1	Several Applications of Sequence Alignment	135
4.4.2	Examples of Pairwise Alignment	137
4.5	Exercises	146
5	Multiple Sequence Alignment	149
5.1	Pairwise Alignment Among Multiple Sequences	149
5.1.1	Using Pairwise Alignment to Process Multiple Sequences	149
5.1.2	Topological Space Induced by Pairwise Alignment of Multiple Sequences	150

- 5.2 Optimization Criteria of MA 156
 - 5.2.1 The Definition of MA 156
 - 5.2.2 Uniform Alignment Criteria and SP-Optimization
Criteria for Multiple Sequences..... 156
 - 5.2.3 Discussion of the Optimization Criterion of MA 160
 - 5.2.4 Optimization Problem Based on Shannon Entropy 164
 - 5.2.5 The Similarity Rate and the Rate of Virtual Symbols .. 170
- 5.3 Super Multiple Alignment 172
 - 5.3.1 The Situation for MA 172
 - 5.3.2 Algorithm of SMA 174
 - 5.3.3 Comparison Among Several Algorithms 179
- 5.4 Exercises, Analyses, and Computation 180

- 6 Network Structures of Multiple Sequences Induced
by Mutation 183**
 - 6.1 General Method of Constructing the Phylogenetic Tree 183
 - 6.1.1 Summary..... 183
 - 6.1.2 Distance-Based Methods 184
 - 6.1.3 Feature-Based (Maximum Parsimony) Methods 188
 - 6.1.4 Maximum-Likelihood Method and the Bayes Method .. 191
 - 6.2 Network Structure Generated by MA 197
 - 6.2.1 Graph and Tree Generated by MA 197
 - 6.2.2 Network System Generated by Mutations
of Multiple Sequences 200
 - 6.3 The Application of Mutation Network Analysis..... 206
 - 6.3.1 Selection of the Data Sample 206
 - 6.3.2 The Basic Steps to Analyze the Sequences 208
 - 6.3.3 Remarks on the Alignment and Output Analysis 210
 - 6.4 Exercises, Analyses, and Computation 216

- 7 Alignment Space 219**
 - 7.1 Construction of Alignment Space and Its Basic Theorems..... 219
 - 7.1.1 What Is Alignment Space? 219
 - 7.1.2 The Alignment Space Under General Metric 221
 - 7.2 The Analysis of Data Structures in Alignment Spaces 224
 - 7.2.1 Maximum Score Alignment and Minimum Penalty
Alignment 224
 - 7.2.2 The Structure Mode of the Envelope of Pairwise
Sequences 226
 - 7.2.3 Uniqueness of the Maximum Core and Minimum
Envelope of Pairwise Sequences 230
 - 7.2.4 The Envelope and Core of Pairwise Sequences..... 231
 - 7.2.5 The Envelope of Pairwise Sequences and Its Alignment
Sequences 233

7.3 The Counting Theorem of the Optimal Alignment
and Alignment Spheroid 237

7.3.1 The Counting Theorem of the Optimal Alignment 237

7.3.2 Alignment Spheroid 238

7.4 The Virtual Symbol Operation in the Alignment Space 241

7.4.1 The Definition of the Virtual Symbol Operator 241

7.4.2 The Modulus Structure of the Virtual Symbol
Operator 243

7.4.3 The Isometric Operation and Micro-Adapted
Operation of Virtual Symbols 247

7.5 Exercises, Analyses, and Computation 250

Part II Protein Configuration Analysis

**8 Background Information Concerning the Properties
of Proteins 255**

8.1 Amino Acids and Peptide Chains 255

8.1.1 Amino Acids 255

8.1.2 Basic Structure of Peptide Chains 257

8.2 Brief Introduction of Protein Configuration Analysis 259

8.2.1 Protein Structure Database 259

8.2.2 Brief Introduction to Protein Structure Analysis 260

8.3 Analysis and Exploration 263

**9 Informational and Statistical Iterative Analysis of Protein
Secondary Structure Prediction 265**

9.1 Protein Secondary Structure Prediction and Informational
and Statistical Iterative Analysis 265

9.1.1 Protein Secondary Structure Prediction 265

9.1.2 Data Source and Informational Statistical Model
of PSSP 267

9.1.3 Informational and Statistical Characteristic Analysis
on Protein Secondary Structure 269

9.2 Informational and Statistical Calculation Algorithms
for PSSP 271

9.2.1 Informational and Statistical Calculation for PSSP 271

9.2.2 Informational and Statistical Calculation Algorithm
for PSSP 273

9.2.3 Discussion of the Results 275

9.3 Exercises, Analyses, and Computation 277

10	Three-Dimensional Structure Analysis of the Protein Backbone and Side Chains	279
10.1	Space Conformation Theory of Four-Atom Points	279
10.1.1	Conformation Parameter System of Four-Atom Space Points	279
10.1.2	Phase Analysis on Four-Atom Space Points	283
10.1.3	Four-Atom Construction of Protein 3D Structure	286
10.2	Triangle Splicing Structure of Protein Backbones	288
10.2.1	Triangle Splicing Belts of Protein Backbones	288
10.2.2	Characteristic Analysis of the Protein Backbone Triangle Splicing Belts	291
10.3	Structure Analysis of Protein Side Chains	292
10.3.1	The Setting of Oxygen Atom O and Atom C_B on the Backbones	293
10.3.2	Statistical Analysis of the Structures of Tetrahedrons V_O, V_B	295
10.4	Exercises, Analyses, and Computation	297
11	Alignment of Primary and Three-Dimensional Structures of Proteins	299
11.1	Structure Analysis for Protein Sequences	299
11.1.1	Alignment of Protein Sequences	299
11.1.2	Differences and Similarities Between the Alignment of Protein Sequences and of DNA Sequences	301
11.1.3	The Penalty Functions for the Alignment of Protein Sequences	302
11.1.4	Key Points of the Alignment Algorithms of Protein Sequences	306
11.2	Alignment of Protein Three-Dimensional Structures	307
11.2.1	Basic Idea and Method of the Alignment of Three-Dimensional Structures	307
11.2.2	Example of Computation in the Discrete Case	310
11.2.3	Example of Computation in Consecutive Case	314
11.3	Exercises, Analyses, and Computation	323
12	Depth Analysis of Protein Spatial Structure	325
12.1	Depth Analysis of Amino Acids in Proteins	325
12.1.1	Introduction to the Concept of Depth	325
12.1.2	Definition and Calculation of Depth	327
12.1.3	Proof of Theorem 40	329
12.1.4	Proof of Theorem 41	334
12.2	Statistical Depth Analysis of Protein Spatial Particles	335
12.2.1	Calculation for Depth Tendency Factor of Amino Acid	335
12.2.2	Analysis of Depth Tendency Factor of Amino Acid	338

12.2.3	Prediction for Depth of Multiple Peptide Chains in Protein Spatial Structure	342
12.2.4	The Level Function in Spatial Particle System	347
12.2.5	An Example	347
12.3	Exercises	353
13	Analysis of the Morphological Features of Protein Spatial Structure	355
13.1	Introduction	355
13.1.1	Morphological Features of Protein Spatial Structure ...	355
13.1.2	Several Basic Definitions and Symbols	357
13.1.3	Preliminary Analysis of the Morphology of Spatial Particle System	360
13.1.4	Example	362
13.2	Structural Analysis of Cavities and Channels in a Particle System	366
13.2.1	Definition, Classification and Calculation of Cavity ...	366
13.2.2	Relationship Between Cavities	368
13.2.3	Example	371
13.3	Analysis of γ -Accessible Radius in Spatial Particle System ...	371
13.3.1	Structural Analysis of a Directed Polyhedron	371
13.3.2	Definition and Basic Properties of γ -Accessible Radius .	377
13.3.3	Basic Principles and Methods of γ -Accessible Radius ...	378
13.4	Recursive Algorithm of γ -Function	382
13.4.1	Calculation of the γ -Function Generated by 0-Level Convex Hull	382
13.4.2	Recursive Calculation of γ -Function	384
13.4.3	Example	387
13.5	Proof of Relative Theorems and Reasoning of Computational Formulas	387
13.5.1	Proofs of Several Theorems	387
13.5.2	Reasoning of Several Formulas	390
13.6	Exercises	394
14	Semantic Analysis for Protein Primary Structure	395
14.1	Semantic Analysis for Protein Primary Structure	395
14.1.1	The Definition of Semantic Analysis for Protein Primary Structure	395
14.1.2	Information-Based Stochastic Models in Semantic Analysis	397
14.1.3	Determination of Local Words Using Informational and Statistical Means and the Relative Entropy Density Function for the Second and Third Ranked Words	400
14.1.4	Semantic Analysis for Protein Primary Structure	402

- 14.2 Permutation and Combination Methods for Semantic Structures 412
 - 14.2.1 Notation Used in Combinatorial Graph Theory 416
 - 14.2.2 The Complexity of Databases 420
 - 14.2.3 Key Words and Core Words in a Database 421
 - 14.2.4 Applications of Combinatorial Analysis 425
- 14.3 Exercises, Analyses, and Computation 429

- Epilogue** 431

- References** 433

- Index** 441