

Inhalt

1	Einleitung	11
I	Einführung und grundlegende Algorithmen	15
2	Grundlagen der Molekularbiologie	16
2.1	Proteine	16
2.2	Nucleinsäuren	18
2.3	Erbinformation und die Proteinbiosynthese	21
2.4	Labortechniken	23
2.4.1	Grundbegriffe und Basismethoden	24
2.4.2	Vervielfältigung von DNA	24
2.4.3	Gelelektrophorese	26
2.4.4	DNA-Chips	27
2.5	Literaturhinweise	29
3	Grundlagen: Strings, Graphen und Algorithmen	30
3.1	Strings	30
3.2	Graphen	32
3.3	Algorithmen und Komplexität	35
3.4	Literaturhinweise	42
4	String-Algorithmen	43
4.1	Das String-Matching-Problem	43
4.2	String-Matching-Automaten	45
4.3	Der Boyer-Moore-Algorithmus	50
4.4	Suffix-Bäume	56
4.5	Weitere Anwendungen von Suffix-Bäumen	64
4.5.1	Verallgemeinerte Suffix-Bäume und das Teilstring-Problem	64
4.5.2	Längste gemeinsame Teilstrings	67

8		Inhalt
4.5.3	Effiziente Bestimmung von Overlaps	70
4.5.4	Wiederholungen in Strings	73
4.6	Zusammenfassung	75
4.7	Literaturhinweise	76
5	Alignment-Verfahren	78
5.1	Alignment von zwei Strings	79
5.1.1	Grundlegende Definitionen	79
5.1.2	Globales Alignment	81
5.1.3	Lokales und semiglobales Alignment	87
5.1.4	Verallgemeinerte Bewertungsfunktionen	91
5.2	Heuristiken zur Datenbanksuche	95
5.2.1	Das FASTA-Verfahren	95
5.2.2	Das BLAST-Verfahren	97
5.3	Multiple Alignments	98
5.3.1	Definition und Bewertung von multiplen Alignments	98
5.3.2	Exakte Bestimmung multipler Alignments	102
5.3.3	Zusammenfügen paarweiser Alignments	106
5.4	Zusammenfassung	110
5.5	Literaturhinweise	111
II	Sequenzierung von DNA	113
6	Problemstellung, Einleitung und Übersicht	114
7	Physikalische Kartierung	117
7.1	Restriktionsstellen-Kartierung	117
7.1.1	Das Double-Digest-Verfahren	118
7.1.2	Das Partial-Digest-Verfahren	125
7.1.3	Methoden zur Restriktionsstellen-Kartierung im Vergleich	135
7.2	Kartierung durch Hybridisierung	136
7.2.1	Kartierung mit eindeutigen Probes	139
7.2.2	Kartierung mit eindeutigen Probes und Fehlern	153
7.2.3	Kartierung mit nicht-eindeutigen Probes	158
7.3	Zusammenfassung	160
7.4	Literaturhinweise	161
8	Bestimmung der Basensequenz	163
8.1	Shotgun-Sequenzierung	163
8.1.1	Fehlerquellen und Probleme beim Fragment-Assembly	166

Inhalt	9
8.1.2	Das Shortest-Common-Superstring-Problem 168
8.1.3	Weitere Modelle für das Fragment-Assembly-Problem 188
8.2	Sequenzierung durch Hybridisierung 192
8.3	Zusammenfassung 198
8.4	Literaturhinweise 199
III	Weitere molekularbiologische Problemstellungen 201
9	Bestimmung von Signalen in DNA-Sequenzen 202
9.1	Gleiche und ähnliche Teilstrings 202
9.2	Tandem-Repeats 206
9.3	Häufige und seltene Teilstrings 212
9.4	Hidden-Markov-Modelle 216
9.5	Zusammenfassung 223
9.6	Literaturhinweise 224
10	Vergleich von Genomen 226
10.1	Modellierung 226
10.2	Sortieren ungerichteter Permutationen 229
10.3	Sortieren gerichteter Permutationen 236
10.4	Bestimmung des syntenischen Abstands 237
10.5	Zusammenfassung 244
10.6	Literaturhinweise 244
11	Phylogenetische Bäume 246
11.1	Ultrametrische Distanzen 247
11.2	Additive Bäume 254
11.3	Zweiwertige Merkmale 257
11.4	Das Parsimony-Prinzip und die Quartett-Methode 265
11.5	Zusammenfassung 274
11.6	Literaturhinweise 275
12	Molekülstrukturen 277
12.1	Vorhersagen der Sekundärstruktur von RNA 278
12.1.1	Minimierung der freien Energie 280
12.1.2	Stochastische kontextfreie Grammatiken 287

12.2	Strukturbasierter Vergleich von Biomolekülen	295
12.3	Strukturvorhersagen für Proteine	308
12.3.1	Das Gitter-Modell	311
12.3.2	Protein-Threading	320
12.4	Zusammenfassung	325
12.4.1	Vorhersagen der RNA-Sekundärstruktur	325
12.4.2	Strukturbasierter Vergleich von Molekülen	327
12.4.3	Strukturvorhersage für Proteine	327
12.5	Literaturhinweise	328
12.5.1	Vorhersagen der RNA-Sekundärstruktur	328
12.5.2	Strukturbasierter Vergleich von Molekülen	329
12.5.3	Strukturvorhersage für Proteine	330
	Literaturverzeichnis	331
	Sachverzeichnis	340