

## Inhaltsverzeichnis

**Vorwort** *V*

**Teil I Grundlagen – Biologie und Datenbanken** *1*

<b>1</b>	<b>Biologische Grundlagen</b>	<i>5</i>
1.1	DNA	<i>6</i>
1.2	Genetischer Code und Genomkomposition	<i>8</i>
1.3	Transkription	<i>12</i>
1.4	RNA	<i>13</i>
1.5	Proteine	<i>14</i>
1.6	Peptidbindung	<i>16</i>
1.7	Konformation von Aminosäureseitenketten	<i>16</i>
1.8	Ramachandran-Plot	<i>17</i>
1.9	Hierarchische Beschreibung von Proteinstrukturen	<i>18</i>
1.10	Sekundärstrukturelemente	<i>19</i>
1.11	$\alpha$ -Helix	<i>20</i>
1.12	$\beta$ -Faltblätter	<i>20</i>
1.13	Supersekundärstrukturelemente	<i>21</i>
1.14	Proteindomänen	<i>22</i>
1.15	Proteinfamilien	<i>23</i>
1.16	Enzyme	<i>26</i>
1.17	Proteinkomplexe	<i>27</i>
1.18	Evolutionäre Prozesse	<i>28</i>
1.19	Fachbegriffe	<i>30</i>
	Literatur	<i>33</i>
<b>2</b>	<b>Sequenzen und ihre Funktion</b>	<i>37</i>
2.1	Definitionen und Operatoren	<i>38</i>
2.2	DNA-Sequenzen	<i>39</i>
2.3	Proteinsequenzen	<i>39</i>
2.4	Vergleich der Sequenzkomposition	<i>41</i>
2.5	Ontologien	<i>45</i>

x | *Inhaltsverzeichnis*

2.6	Analyse der Anreicherung von GO-Termen	48
2.7	Semantische Ähnlichkeit von GO-Termen	48
2.7.1	Bewertung mit informationstheoretischen Ansätzen	48
2.7.2	Vergleich mit einer graphentheoretischen Methode	50
	Literatur	54
<b>3</b>	<b>Datenbanken</b>	<b>57</b>
3.1	Nukleotidsequenzdatenbanken	58
3.2	RNA-Sequenz-Datenbanken	59
3.3	Proteinsequenzdatenbanken	60
3.4	3-D-Struktur-Datenbanken	60
3.5	SMART: Analyse der Domänenarchitektur	62
3.6	STRING: Proteine und ihre Interaktionen	62
3.7	SCOP: Strukturelle Klassifikation von Proteinen	63
3.8	Pfam: Kompilation von Proteinfamilien	66
3.9	COG und eggNOG: Gruppen orthologer Gene	68
3.10	KEGG: Gene, Genome und Krankheiten	68
3.11	NCBI-Datenbanken: Literatur und biologisches Wissen	69
3.12	Weitere Datenbanken	70
	Literatur	74
	<b>Teil II Lernen, Optimieren und Entscheiden</b>	<b>77</b>
<b>4</b>	<b>Grundbegriffe der Stochastik</b>	<b>81</b>
4.1	Grundbegriffe der beschreibenden Statistik	82
4.2	Zufallsvariable, Wahrscheinlichkeitsmaß	84
4.3	Urnenexperimente und diskrete Verteilungen	86
4.4	Die kolmogoroffschen Axiome	89
4.5	Bedingte Wahrscheinlichkeit, Unabhängigkeit, Satz von Bayes	89
4.6	Markov-Ketten	91
4.7	Erwartungswert, Varianz	91
4.8	Wichtige Wahrscheinlichkeitsverteilungen	92
4.8.1	Diskrete Verteilungen	92
4.8.2	Totalstetige Verteilungen	93
4.9	Schätzer	96
4.10	Grundlagen statistischer Tests	98
4.11	Eine optimale Entscheidungstheorie: die Neyman-Pearson-Methode	100
	Literatur	101
<b>5</b>	<b>Bayessche Entscheidungstheorie und Klassifikatoren</b>	<b>103</b>
5.1	Bayessche Entscheidungstheorie	103
5.1.1	Ein Beispiel: Klassifikation der Proteinoberfläche	104
5.1.2	Übergang zu bedingten Wahrscheinlichkeiten	105

5.1.3	Erweitern auf $m$ Eigenschaften	107
5.2	Marginalisieren	109
5.3	Boosting	110
5.4	ROC-Kurven	112
5.4.1	Bewerten von Fehlklassifikationen	112
5.4.2	Aufnehmen einer ROC-Kurve	112
5.5	Testmethoden für kleine Trainingsmengen	115
	Literatur	117
<b>6</b>	<b>Klassische Cluster- und Klassifikationsverfahren</b>	<b>119</b>
6.1	Metriken und Clusteranalyse	120
6.2	Das mittlere Fehlerquadrat als Gütemaß	120
6.3	Ein einfaches iteratives Clusterverfahren	121
6.4	$k$ -Means-Clusterverfahren	123
6.5	Hierarchische Clusterverfahren	126
6.6	Affinity propagation	127
6.7	Bewertung der Clusterverfahren	129
6.8	Überlappende Cluster	130
6.9	Nächster-Nachbar-Klassifikation	130
6.10	$k$ -nächste-Nachbarn-Klassifikation	132
	Literatur	133
<b>7</b>	<b>Neuronale Netze</b>	<b>135</b>
7.1	Architektur von neuronalen Netzen	136
7.2	Das Perzeptron	136
7.3	Modellieren boolescher Funktionen	138
7.4	Lösbarkeit von Klassifikationsaufgaben	139
7.5	Universelle Approximation	141
7.6	Lernen in neuronalen Netzen	143
7.7	Der Backpropagation-Algorithmus	144
7.8	Codieren der Eingabe	147
7.9	Selbstorganisierende Karten	148
7.10	Tiefe Architekturen	150
7.11	Ein einfaches Neuron, die <i>rectified linear unit</i>	151
7.12	Das Neocognitron als alternatives Modellierparadigma	152
7.13	Faltung mithilfe von CNNs	153
7.14	Längerfristiges Speichern von Eingabedaten	157
7.15	Attention-basierte Netze	161
	Literatur	168
<b>8</b>	<b>Genetische Algorithmen</b>	<b>171</b>
8.1	Objekte und Funktionen	173
8.2	Ablauf des Verfahrens	174
8.3	Codieren der Problemstellung	176
8.4	Der Begriff des Schemas	176

**XII** | *Inhaltsverzeichnis*

- 8.5 Dynamik der Anzahl von Schemata 177
- 8.6 Limitationen genetischer Algorithmen 179
- 8.7 Genetisches Programmieren 180
- Literatur 183

**Teil III Algorithmen und Modelle der Bioinformatik 185**

- 9 Paarweiser Sequenzvergleich 189**
  - 9.1 Dotplots 191
    - 9.1.1 Definition 191
    - 9.1.2 Beispiel 192
    - 9.1.3 Implementierung 193
    - 9.1.4 Abschätzen der Laufzeit 194
    - 9.1.5 Anwendungen 195
    - 9.1.6 Einschränkungen und Ausblick 196
  - 9.2 Entwickeln eines optimalen Alignment-Verfahrens 198
    - 9.2.1 Paarweise und multiple Sequenzalignments 200
    - 9.2.2 Dynamisches Programmieren 200
    - 9.2.3 Distanzen und Metriken 202
    - 9.2.4 Die Minkowski-Metrik 203
    - 9.2.5 Die Hamming-Distanz 203
  - 9.3 Levenshtein-Distanz 204
    - 9.3.1 Berechnungsverfahren 206
    - 9.3.2 Ableiten des Alignments 210
  - 9.4 Bestimmen der Ähnlichkeit von Sequenzen 210
    - 9.4.1 Globales Alignment 210
    - 9.4.2 Lokales Sequenzalignment 211
  - 9.5 Optimales Bewerten von Lücken 212
    - 9.5.1 Eigenschaften affiner Kostenfunktionen 213
    - 9.5.2 Integration in Algorithmen 213
  - 9.6 Einordnung der Algorithmen 215
  - Literatur 216
- 10 Sequenzmotive 219**
  - 10.1 Signaturen 220
  - 10.2 Die PROSITE-Datenbank 221
  - 10.3 Die BLOCKS-Datenbank 221
  - 10.4 Sequenzprofile 222
  - 10.5 Scores für Promotorsequenzen 224
  - 10.6 Möglichkeiten und Grenzen profilbasierter Klassifikation 224
  - 10.7 Sequenzlogos 225
  - 10.8 Konsensussequenzen 226
  - 10.9 Sequenzen niedriger Komplexität 227
  - 10.10 Der SEG-Algorithmus 228
  - Literatur 231

- 11 Scoring-Schemata 233**
  - 11.1 Theorie von Scoring-Matrizen 234
  - 11.2 Algorithmenbedingte Anforderungen 237
  - 11.3 Identitätsmatrizen 237
  - 11.4 PAM-Einheit 238
  - 11.5 PAM-Matrizen 238
  - 11.6 Ein moderner PAM-Ersatz: die JTT-Matrix 240
  - 11.7 BLOSUM-Matrizen 241
  - 11.8 Matrixentropie 243
  - 11.9 Scoring-Schemata und Anwendungen 244
  - 11.10 Flexible Erweiterung: Scoring-Funktionen 245
    - Literatur 247
  
- 12 FASTA und die BLAST-Suite 249**
  - 12.1 FASTA 250
    - 12.1.1 Programmablauf 250
    - 12.1.2 Statistische Bewertung der Treffer 252
  - 12.2 BLAST 255
    - 12.2.1 Konzepte und Umsetzung 256
    - 12.2.2 Statistik von Alignments 259
    - 12.2.3 Ausgabe der Treffer 264
  - 12.3 Vergleich der Empfindlichkeit von FASTA und BLAST 264
  - 12.4 Ansätze zur Performanzsteigerung 265
  - 12.5 Profilbasierter Sequenzvergleich 266
  - 12.6 PSI-BLAST 267
  - 12.7 Sensitivität verschiedener Sequenzvergleichsmethoden 269
  - 12.8 Vergleich von Profilen und Konsensussequenzen 270
  - 12.9 DELTA-BLAST 271
  - 12.10 Alternative Ansätze 275
    - Literatur 276
  
- 13 Multiple Sequenzalignments und Anwendungen 279**
  - 13.1 Berechnen von Scores für multiple Sequenzalignments 281
  - 13.2 Iteratives Berechnen eines Alignments 282
  - 13.3 ClustalW: Ein klassischer Algorithmus 283
    - 13.3.1 Grundlegende Konzepte 283
    - 13.3.2 Algorithmus 283
    - 13.3.3 Ein Beispiel: MSA für Trypsininhibitoren 284
  - 13.4 T-Coffee 286
  - 13.5 M-Coffee und 3D-Coffee 289
  - 13.6 Alternative Ansätze 291
  - 13.7 Alignieren großer Datensätze mit Clustal Omega 292
  - 13.8 Alignieren großer Proteinsequenzdatensätze mit DECIPHER 293
  - 13.9 Charakterisierung von Residuen mithilfe von Alignments 296
    - 13.9.1 Entwickeln der Scoring-Funktion 297

**XIV** | *Inhaltsverzeichnis*

- 13.9.2 FRpred: Vorhersage funktionell wichtiger Residuen 297
- 13.9.3 SDPpred: Vergleich homologer Proteine mit unterschiedlicher Spezifität 298
- 13.10 Alignment von DNA- und RNA-Sequenzen 300
  - Literatur 301
  
- 14 Grundlagen phylogenetischer Analysen 303**
  - 14.1 Einteilung phylogenetischer Ansätze 307
  - 14.2 Distanzbasierte Verfahren 307
    - 14.2.1 Ultrametrische Matrizen 308
    - 14.2.2 Additive Matrizen 309
  - 14.3 Linkage-Algorithmen 311
  - 14.4 Der Neighbour-Joining-Algorithmus 313
  - 14.5 Parsimony-Methoden 314
  - 14.6 Maximum-Likelihood-Ansätze 317
    - 14.6.1 Übergangswahrscheinlichkeiten für DNA-Sequenzen 318
    - 14.6.2 Empirische Modelle der Proteinevolution 319
    - 14.6.3 Berechnen der Likelihood eines Baumes 321
    - 14.6.4 Quartett-Puzzle: Heuristik zum Finden einer Topologie 323
  - 14.7 Grundannahmen phylogenetischer Algorithmen 325
  - 14.8 Statistische Bewertung phylogenetischer Bäume 326
    - 14.8.1 Validierung durch Outgroups 327
    - 14.8.2 Bootstrap-Verfahren und A-posteriori-Wahrscheinlichkeiten 327
    - 14.8.3 Alternativen und Ergebnisse 329
    - Literatur 332
  
- 15 Markov-Ketten und Hidden-Markov-Modelle 335**
  - 15.1 Ein epigenetisches Signal: CpG-Inseln 335
  - 15.2 Finite Markov-Ketten 336
  - 15.3 Kombination zweier Ketten zu einem Klassifikator 337
  - 15.4 Genvorhersage mithilfe inhomogener Ketten 340
  - 15.5 Hidden-Markov-Modelle 343
  - 15.6 Der Viterbi-Pfad 346
  - 15.7 Ein HMM zur Erkennung von CpG-Inseln 348
  - 15.8 Der Vorwärts- und der Rückwärtsalgorithmus 349
  - 15.9 Schätzen von Parametern 351
  - 15.10 Der Baum-Welch-Algorithmus 352
  - 15.11 Entwurf von HMMs 354
  - 15.12 Verwendung und Grenzen von HMMs 356
  - 15.13 Wichtige Eigenschaften von Markov-Ketten 357
  - 15.14 Markov-Ketten-Monte-Carlo-Verfahren 359
    - 15.14.1 Monte-Carlo-Integration 359
    - 15.14.2 Metropolis-Hastings-Algorithmus 360
    - 15.14.3 Simulated annealing 361
    - 15.14.4 Gibbs-Sampler 362

- 15.15 Weitere Anwendungen von Markov-Ketten 362
  - Literatur 366
  
- 16 Profil-HMMs 369**
  - 16.1 HMM-Struktur zur Beschreibung von Proteinfamilien 370
  - 16.2 Suche nach homologen Sequenzen 373
  - 16.3 Modellbau für Profil-HMMs 376
  - 16.4 Approximieren von Wahrscheinlichkeitsdichten 380
  - 16.5 HHsearch: Vergleich zweier Profil-HMMs 386
    - 16.5.1 Grundlagen des Alignments von zwei Hidden-Markov-Ketten 387
    - 16.5.2 Paarweises Alignment von HMMs 390
    - 16.5.3 Performanz von HHsearch 391
    - 16.5.4 Strukturvorhersage mit HHsearch 393
    - Literatur 395
  
- 17 Support-Vektor-Maschinen 397**
  - 17.1 Beschreibung des Klassifikationsproblems 398
  - 17.2 Lineare Klassifikatoren 399
  - 17.3 Klassifizieren mit großer Margin 403
  - 17.4 Kernel-Funktionen und Merkmalsräume 405
  - 17.5 Implizite Abbildung in den Merkmalsraum 407
  - 17.6 Eigenschaften von Kernel-Funktionen 408
  - 17.7 Häufig verwendete Kernel-Funktionen 409
  - 17.8 Aus Merkmalen abgeleitete Kernel-Funktionen 410
  - 17.9 Support-Vektor-Maschinen in der Anwendung 416
  - 17.10 Multiklassen-SVM 419
  - 17.11 Theoretischer Hintergrund 420
    - Literatur 424
  
- 18 Vorhersage der Sekundärstruktur 427**
  - 18.1 Vorhersage der Proteinsekundärstruktur 427
    - 18.1.1 Ein früher Ansatz: Chou-Fasman-Verfahren 428
    - 18.1.2 PHD: profilbasierte Vorhersage 429
  - 18.2 Vorhersage der RNA-Sekundärstruktur 436
    - 18.2.1 RNA-Sequenzen und -Strukturen 438
    - 18.2.2 Freie Energie und Strukturen 439
    - 18.2.3 Sekundärstrukturvorhersage durch Energieminimierung 440
    - 18.2.4 Strukturen mit Schleifen 442
    - 18.2.5 MEA-Verfahren zur Vorhersage von Strukturen mit Pseudoknoten 444
    - 18.2.6 Strukturvorhersage mithilfe von multiplen Sequenzalignments 447
    - Literatur 449
  
- 19 Vergleich von Protein-3-D-Strukturen 451**
  - 19.1 Grundlagen des Strukturvergleichs 453
  - 19.2 Simulated annealing 455
  - 19.3 DALI: fragmentbasierte Superposition 458

**XVI** | Inhaltsverzeichnis

- 19.3.1 Scores für Substrukturen 459
- 19.3.2 Alignieren von Substrukturen 459
- 19.4 Fr-TM-align: Alignieren von Fragmenten 461
- 19.5 SPalignNS: optimales Kombinieren von Residuenpaaren 462
- 19.6 FAST: Vergleich der lokalen Geometrie 463
- 19.7 DeepAlign: Verwenden eines Strukturalphabets 466
- 19.8 Multiple Superpositionen 471
  - Literatur 474
  
- 20 Vorhersage der Protein-3-D-Struktur, Proteindesign und Moleküldynamik 477**
  - 20.1 Threading-Verfahren 482
  - 20.2 *3D-1D-Profile*: profilbasiertes Threading 484
    - 20.2.1 Bestimmen der lokalen Umgebung 484
    - 20.2.2 Erzeugen eines 3-D-1-D-Profiles 486
  - 20.3 Wissensbasierte Kraftfelder 489
    - 20.3.1 Theoretische Grundlagen 490
    - 20.3.2 Ableiten der Potenziale 493
  - 20.4 Rotamerbibliotheken 494
  - 20.5 MODELLER 499
  - 20.6 Bewerten der Modellqualität 504
  - 20.7 Alternative Modellieransätze 504
  - 20.8 ROSETTA/ROBETTA 505
    - 20.8.1 *De-novo*-Strukturvorhersage mit ROSETTA 506
    - 20.8.2 Verfeinerung der Fragmentinsertion 508
    - 20.8.3 Modellieren strukturell variabler Regionen 508
    - 20.8.4 Proteindesign mithilfe von ROSETTA 510
  - 20.9 Moleküldynamiksimulationen 517
    - 20.9.1 Physikalische Grundlagen von MD-Simulationen 518
    - 20.9.2 Berechnungsverfahren 519
    - 20.9.3 Berechnen der Interaktionen mithilfe von Kraftfeldern 521
    - 20.9.4 Spezielle Hardware beschleunigt die Simulationen 522
    - Literatur 523
  
- 21 Analyse integraler Membranproteine 527**
  - 21.1 Architektur integraler Membranproteine 528
  - 21.2 Spezifische Probleme beim Sequenzvergleich 530
  - 21.3 Vorhersage der Topologie von  $\alpha$ -helikalen IMPs 530
    - 21.3.1 HMMTOP 531
    - 21.3.2 MEMSAT-SVM 532
    - 21.3.3 Ein Metaansatz: TOPCONS2 534
  - 21.4 Vorhersage der Struktur von  $\beta$ -Fässern 535
    - 21.4.1 TMBpro 535
    - 21.4.2 PRED-TMBB2 537
    - 21.4.3 BOCTOPUS2 539



- 21.4.4 Alternative Ansätze und Homologiemodellierung 541
  - Literatur 541
  
- 22 Entschlüsselung von Genomen 545**
  - 22.1 Shotgun-Sequenzierung 549
  - 22.2 Erwartete Anzahl von Contigs beim Shotgun-Ansatz 550
  - 22.3 Basecalling und Sequenzqualität 551
  - 22.4 Der klassische Assemblieransatz 553
    - 22.4.1 Phase eins: Bestimmen überlappender Präfix-Suffix-Regionen 554
    - 22.4.2 Phase zwei: Erzeugen von Contigs 556
    - 22.4.3 Phase drei: Generieren der Konsensussequenz 556
  - 22.5 Assemblieren kurzer Fragmente 558
  - 22.6 Assemblieren langer und fehlerbehafteter Reads 561
  - 22.7 Annotation kompletter Genome 565
  - 22.8 Metagenomik 570
    - 22.8.1 Spezielle Anforderungen an die Bioinformatik 571
    - 22.8.2 Minimalanforderungen für die Metagenomannotation 573
    - Literatur 574
  
- 23 Auswertung von Transkriptomdaten 579**
  - 23.1 DNA-Chip-Technologie 579
    - 23.1.1 Datenbanken für Transkriptomdaten 581
    - 23.1.2 Grenzen der Technologie 582
  - 23.2 Analyse von DNA-Chip-Signalen 583
    - 23.2.1 Quantifizierung von Expressionswerten 583
    - 23.2.2 Normalisieren und Datenreduktion 584
  - 23.3 Identifizieren differenziell exprimierter Gene 586
  - 23.4 RNA-Sequenzierung 587
  - 23.5 Analyse der RNA-Sequenzen 588
  - 23.6 Einzelzell-RNA-Sequenzierung 591
  - 23.7 Metriken zum Vergleich von Expressionsdaten 591
  - 23.8 Analyse kompletter Expressionsdatensätze 593
    - 23.8.1 Anwenden von Clusterverfahren 593
    - 23.8.2 Validierung und Alternativen 593
  - 23.9 Hauptkomponentenanalyse 594
  - 23.10 Biclusterverfahren 597
    - 23.10.1 ISA: ein performantes Biclusterverfahren 597
    - 23.10.2 Der Signaturalgorithmus 598
    - 23.10.3 Iterative Optimierung 601
    - 23.10.4 QUBIC2: Ein graphenbasiertes Biclusterverfahren 602
  - 23.11 Grenzen und Alternativen bei der Expressionsanalyse 604
  - 23.12 Genexpressions-Profiling 605
    - 23.13 Visualisieren mithilfe von Wärmekarten 606
      - 23.13.1 Der klassische Ansatz 607
  - 23.14 Datenaufbereitung für systembiologische Fragestellungen 607

**XVIII** | *Inhaltsverzeichnis*

- 23.14.1 Bündelung von Datenbankinformation 608
- 23.14.2 Statistische Analyse der Termverteilung 609
- 23.14.3 Verwendbarkeit der Verfahren 610
  - Literatur 612
  
- 24 Analyse von Protein-Protein-Interaktionen 615**
  - 24.1 Biologische Bedeutung des Interaktoms 615
  - 24.2 Methoden zum Bestimmen des Interaktoms 616
  - 24.3 Vergleich von Codonhäufigkeiten 618
  - 24.4 Analyse des Genominhaltes 619
    - 24.4.1 Genfusion 619
    - 24.4.2 Phyletische Profile 620
    - 24.4.3 Analyse von Genfolgen 622
    - 24.4.4 Performanz sequenzbasierter Methoden 623
  - 24.5 Suche nach korrelierten Mutationen 624
    - 24.5.1 Erzeugen sortierter MSA-Paare 624
    - 24.5.2 Identifizieren korrelierter Mutationen 625
  - 24.6 Vergleich phylogenetischer Bäume 627
    - 24.6.1 Die *Mirror-Tree*-Methode 627
    - 24.6.2 Korrektur des Hintergrundsignals 629
    - 24.6.3 Ein alternativer Ansatz, der auf einem Nullmodell basiert 630
  - 24.7 Vorhersage des Interaktoms der Hefe 631
  - 24.8 Strukturbasierte Protein-Protein-Interaktionsvorhersagen 634
    - 24.8.1 Vorhersagen basierend auf Strukturinformation 635
    - 24.8.2 PrePPI: Integration zusätzlicher Merkmale 637
  - 24.9 Netzwerkbasierte Protein-Protein-Interaktionsvorhersagen 640
    - Literatur 642
  
- 25 Big Data und Deep Learning: neue Herausforderungen und Möglichkeiten 645**
  - 25.1 Klassifikation mit Random Forests 647
    - 25.1.1 Entscheidungsbäume 647
    - 25.1.2 Berechnen der Topologie 649
    - 25.1.3 RF-Algorithmus 652
    - 25.1.4 Theoretische Klassifikationsleistung eines RFs 653
    - 25.1.5 Problemlösungen für konkrete Anwendungen 654
    - 25.1.6 Auswahl informativer Eigenschaften 655
    - 25.1.7 Bioinformatische Anwendungen 657
  - 25.2 Sequenzbasierte Vorhersage der Protein-3-D-Struktur 658
    - 25.2.1 Experimentelle Proteinstrukturaufklärung 659
    - 25.2.2 Berechnen von Co-Variationssignalen 660
    - 25.2.3 PSICOV: Vorhersage räumlich benachbarter Residuenpaare 663
    - 25.2.4 Vorhersage der 3-D-Struktur mithilfe von Kontaktinformation 665
    - 25.2.5 Alternative Nutzung von Kopplungssignalen 665
  - 25.3 Berechnen einer Feinstruktur großer Proteinfamilien 666

25.3.1	MCL: Clustern mithilfe stochastischer Matrizen	667
25.3.2	Cytoscape: Visualisierung von Netzwerkclustern	669
25.4	Positionierung von Nukleosomen	670
25.4.1	Chromatin und Nukleosomen	671
25.4.2	<i>NucleoFinder</i> : ein statistischer Ansatz zur Vorhersage von Nukleosomenpositionen	672
25.5	Auswertung großer Datensätze mit tiefen Lernverfahren	676
25.5.1	DL-basierte Vorhersage der Proteinstruktur	677
25.5.2	AlphaFold2 und RoseTTAFold	680
25.5.3	Erkennen von Translationsinitiationsstellen	683
25.5.4	DeepCpG bestimmt den Methylierungsstatus in einzelnen Zellen	684
25.6	Analyse des menschlichen Genoms mithilfe von ENCODE-Daten	686
25.6.1	Datentypen	687
25.6.2	Genome Browser	689
	Literatur	692
<b>26</b>	<b>Zum Schluss</b>	699
26.1	Informatik in schwierigem Umfeld	699
26.2	Ungelöste Probleme und Herausforderungen	701
	Literatur	704
	<b>Stichwortverzeichnis</b>	705

