

Arthur M. Lesk

BIOINFORMATIK

Eine Einführung

Aus dem Englischen übersetzt von Sebastian Vogel

Mit einem Vorwort zur deutschen Ausgabe von Hans Werner Mewes,
Institut für Bioinformatik, GSF, Neuherberg

In nature's infinite book of secrecy
A little I can read.

Antonius und Kleopatra, William Shakespeare

Inhalt

1 Einführung

Ein Szenario 3

Leben in Raum und Zeit 4

Dogmen: zentrale und periphere 5

Beobachtbare Daten und ihre Archivierung 8

Datenbankpflege, Annotation und Qualitätskontrolle 10

Das World Wide Web 12

Der URL-Dschungel 13

Elektronisches Publizieren 13

Computer und Informatik 14

Programmierung 15

Biologische Klassifikation und Nomenklatur 19

Nutzung von Sequenzen zur Klärung stammesgeschichtlicher

Verwandtschaftsverhältnisse 22

SINES, LINES und die Klärung stammesgeschichtlicher Verwandtschaftsbeziehungen 29

Die Suche nach ähnlichen Sequenzen in Datenbanken: PSI-BLAST 31

Proteinstruktur: eine Einführung 39

Der hierarchische Aufbau der Proteinstruktur 41

Klassifikation von Proteinstrukturen 43

Vorhersage und künstliche Konstruktion von Proteinstrukturen 49

Kritische Beurteilung von Strukturvoraussagen (CASP) 50

Proteindesign 51

Medizinische Anwendungen 51

Die Zukunft 54

Empfohlene Literatur 55

Übungsaufgaben, Anwendungsaufgaben und Web-Aufgaben 56

2 Organisation und Evolution von Genomen

Genomik und Proteomik 66

Gene 66

Proteine 69

Proteome 70

Belascht: die Übertragung der genetischen Information 72

Kartierung zwischen den Karten 77

Hoch auflösende Karten 77

Die Suche nach Genen in Genomen 80

Prokaryotengenome 81

Das Genom des Bakteriums *Escherichia coli* 82

Das Genom des Archaeons *Methanococcus jannaschii* 85

Das Genom eines der einfachsten Lebewesen: <i>Mycoplasma genitalium</i>	86
Eukaryotengenome	86
Das Genom der Bäckerhefe <i>Saccharomyces cerevisiae</i>	91
Das Genom von <i>Caenorhabditis elegans</i>	92
Das Genom von <i>Drosophila melanogaster</i>	94
Das Genom von <i>Arabidopsis thaliana</i>	95
Das Genom des Menschen (<i>Homo sapiens</i>)	96
Proteincodierende Gene	97
Wiederholungssequenzen	99
RNA	99
Einzelnucleotid-Polymorphismen (SNPs)	100
Genetische Vielfalt in der Anthropologie	102
Genetische Vielfalt und Personenidentifizierung	103
Genetische Analyse der Domestizierung von Rindern	104
Evolution von Genomen	104
Gene bitte weitergeben: horizontale Genübertragung	108
Vergleichende Genomanalyse bei Eukaryoten	110
Empfohlene Literatur	112
Übungsaufgaben, Anwendungsaufgaben und Web-Aufgaben	112
3 Archive und die Abfrage von Informationen	
Einleitung	118
Datenbankindexierung und Festlegung von Suchbegriffen	118
Nachgefragt	119
Analyse der gewonnenen Daten	120
Die Archive	120
Datenbanken mit Nucleinsäuresequenzen	121
Genomdatenbanken	123
Proteinsequenzdatenbanken	123
Strukturdatenbanken	126
Spezialisierte oder „Boutique“-Datenbanken	134
Datenbanken für Expression und Proteomik	135
Datenbanken mit Stoffwechselwegen	137
Bibliografische Datenbanken	138
Überblick über molekularbiologische Datenbanken und Server	138
Einstiegsseiten für Archive (Gateways)	139
Zugang zu molekularbiologischen Datenbanken	140
ENTREZ	140
Das Sequence Retrieval System (SRS)	149
Die Protein Information Resource (PIR)	151
ExPASy: das Expert Protein Analysis System	154
Ensembl	155
Und wie geht es weiter?	157
Empfohlene Literatur	158
Übungsaufgaben, Anwendungsaufgaben und Web-Aufgaben	158

4 Alignments und phylogenetische Stammbäume

Sequenz-Alignment: eine Einführung 162

Der Dotplot 163

Dotplots und Sequenz-Alignments 169

Maße für Sequenzähnlichkeit 173

Scoring-Schemata 174

Berechnung des Alignment für zwei Sequenzen 178

Abwandlungen und Verallgemeinerungen 178

Näherungsverfahren zum schnellen Durchsuchen von Datenbanken 179

Der Algorithmus des *dynamic programming* zum optimalen Alignment von Sequenzpaaren 179

Die Bedeutung von Alignments 185

Alignment mehrerer Sequenzen (multiples Sequenz-Alignment) 188

Erkenntnisse über die Struktur, abgeleitet aus dem multiplen Alignment mehrerer Sequenzen 189

Anwendung des multiplen Sequenz-Alignment bei der Suche in Datenbanken 191

Profile 191

PSI-BLAST 193

Hidden-Markov-Modelle (HMMs) 195

Phylogenie 198

Phylogenetische Stammbäume 202

Methoden der Clusterbildung 204

Kladistische Verfahren 206

Das Problem der unterschiedlichen Evolutionsraten 207

Fragen der Berechnung 208

Empfohlene Literatur 210

Übungsaufgaben, Anwendungsaufgaben und Web-Aufgaben 210

5 Proteinstruktur und Medikamentenentwicklung

Einleitung 220

Stabilität und Faltung von Proteinen 222

Das Sasisekharan-Ramakrishnan-Ramachandran-Diagramm beschreibt die erlaubten Konformationen der Hauptkette 222

Die Seitenketten 224

Stabilität und Denaturierung von Proteinen 225

Proteinfaltung 227

Anwendungen der Hydrophobizität 229

Superposition von Strukturen und Struktur-Alignment 233

DALI (*Distance-matrix ALIgnment*) 235

Evolution von Proteinstrukturen 236

Klassifikation von Proteinstrukturen 239

SCOP 239

Vorhersage und Modellierung von Proteinstrukturen 240

Critical Assessment of Structure Prediction (CASP) 243

Vorhersage von Sekundärstrukturen 244

Homologiemodellierung 249

Erkennung von Faltungen 251

Erkennung von Faltungsmustern im Rahmen von CASP2000 254

Berechnung der Konformationsenergie und molekulare Dynamik 255

ROSETTA 258

LINUS 258

Zuordnung von Proteinstrukturen zu Genomen 261

Voraussage von Proteinfunktionen 263

Auseinanderentwicklung von Funktionen: orthologe und paraloge Proteine 264

Entdeckung und Entwicklung von Medikamenten 266

Die Leitstruktur 269

Computergestütztes Wirkstoffdesign 271

Empfohlene Literatur 274

Übungsaufgaben, Anwendungsaufgaben und Web-Aufgaben 275

Zum Schluss 281

Index 282

Farbtafeln