

# Indice

## Prefazione, IX

## 1 Introduzione

- 1.1** Introduzione, 1
- 1.2** Il flusso d'informazione genetica, 2
  - 1.2.1 Gli acidi nucleici: DNA e RNA, 2
  - 1.2.2 Geni e mutazioni, 3
  - 1.2.3 Proteine, 5
  - 1.2.4 La determinazione della struttura tridimensionale delle proteine, 6
- 1.3** Modelli evolutivi, 10
- 1.4** Confronto tra sequenze, 11  
*Riferimenti per saperne di più, 12*

## 2 Introduzione alle basi di dati

- 2.1** Introduzione alle basi di dati, 15
- 2.2** Acquisizione dei dati, 16
  - 2.2.1 Dati estratti dalla letteratura scientifica, 16
  - 2.2.2 Dati ottenuti attraverso sottomissione diretta, 17
  - 2.2.3 Sottomissione di grandi quantità di dati, 17
  - 2.2.4 Riferimenti incrociati, 17
- 2.3** Conservazione dei dati, 18
  - 2.3.1 XML, 18
- 2.4** Distribuzione dei dati, 18
  - 2.4.1 Internet e il World Wide Web, 20
  - 2.4.2 Interrogazione in rete delle basi di dati, 20
  - 2.4.3 SQL e linguaggi di interrogazione delle basi di dati, 23
  - 2.4.4 Operatori booleani, 23
- 2.5** Le basi di dati biologici, 24
  - 2.5.1 Tipologie delle basi di dati biologici, 24
  - 2.5.2 EMBL, 25
  - 2.5.3 GenBank, 28
  - 2.5.4 UniProt: PIR, Swiss-Prot e TrEMBL, 30
  - 2.5.5 Protein Data Bank, 31
  - 2.5.6 MSDSD, 33
  - 2.5.7 SCOP e CATH, 33
  - 2.5.8 PFAM, 34
  - 2.5.9 PROSITE, 36
  - 2.5.10 NCBI Genomes, UCSC, Ensembl, 37
  - 2.5.11 Recupero dell'informazione da banche dati biologiche: Entrez e SRS, 37
  - 2.5.12 Altre basi di dati di interesse biomedico, 40  
*Riferimenti per saperne di più, 41*

## 3 Confronto di sequenze

- 3.1** Introduzione, 43
- 3.2** Il metodo della matrice a punti (*dot plot*), 44
- 3.3** Gli algoritmi dinamici, 48
  - 3.3.1 Allineamenti globali e locali, 55
  - 3.3.2 Algoritmi dinamici per l'allineamento locale, 55
  - 3.3.3 Funzione di penalizzazione delle indel, 56
  - 3.3.4 Misura della significatività statistica di un allineamento, 57
- 3.4** La bioinformatica nella rete, 59  
*Riferimenti per saperne di più, 60*

## 4 Matrici di punteggio

- 4.1** Introduzione, 61
- 4.2** Le matrici PAM, 61
- 4.3** Le matrici BLOSUM, 65
- 4.4** Altre matrici di sostituzione, 66
- 4.5** Qual è la matrice migliore?  
*Riferimenti per saperne di più, 68*

## 5 Ricerche in banche dati con singola sequenza

- 5.1** Introduzione, 69
- 5.2** FASTA, 70
- 5.3** BLAST, 72
- 5.4** Ricerche con l'algoritmo di Waterman, 74
- 5.5** Misura della significatività statistica, 74
- 5.6** Regioni a bassa complessità, 75
- 5.7** Valutazione dell'efficienza di un algoritmo di ricerca, 75
- 5.8** La bioinformatica nella rete, 76  
*Riferimenti per saperne di più, 79*

## 6 Allineamento multiplo di sequenze

- 6.1** Introduzione, 81
- 6.2** Clustal W, 82
- 6.3** Metodi più recenti per l'allineamento multiplo di sequenze, 85
  - 6.3.1 T-Coffee, 86
  - 6.3.2 Altri metodi, 87
- 6.4** La bioinformatica nella rete, 87  
*Riferimenti per saperne di più, 89*

## 7 Alberi filogenetici

- 7.1 Introduzione, 91
  - 7.2 Il metodo *neighbor joining*, 92
  - 7.3 Significatività statistica della topologia di un albero: analisi di *bootstrap*, 94
  - 7.4 Aggiungere la radice all'albero, 95
  - 7.5 La bioinformatica nella rete, 96
- Riferimenti per saperne di più*, 98

## 8 Ricerche in banche dati con allineamenti multipli

- 8.1 Introduzione, 99
  - 8.2 Calcolo dei profili, 99
    - 8.2.1 Il profilo del lavoro, 103
  - 8.3 PSI-BLAST, 104
  - 8.4 Modelli di Markov nascosti, 105
    - 8.4.1 Introduzione, 105
    - 8.4.2 Profili di allineamenti multipli basati su HMM, 108
    - 8.4.3 Uso dei profili di HMM, 110
  - 8.5 Visualizzazione di PSSM e di profili di HMM, 112
  - 8.6 Confronto sequenza-PSSM o sequenza-HMM, 114
  - 8.7 Confronto profilo-profilo, 115
  - 8.8 La bioinformatica nella rete, 116
- Riferimenti per saperne di più*, 117

## 9 Prevedere le proprietà strutturali delle proteine a partire dalle sequenze

- 9.1 Introduzione, 119
  - 9.2 Le reti neurali, 119
    - 9.2.1 Struttura delle reti neurali, 120
    - 9.2.2 Architettura delle reti neurali, 121
    - 9.2.3 Apprendimento da parte delle reti neurali, 121
  - 9.3 Applicazioni biologiche delle reti neurali, 123
    - 9.3.1 La struttura secondaria delle proteine, 123
    - 9.3.2 Assegnazione delle strutture secondarie in base alle coordinate atomiche, 123
    - 9.3.3 Previsione della struttura secondaria delle proteine, 124
    - 9.3.4 Previsione dell'accessibilità al solvente, 130
    - 9.3.5 Previsione delle eliche transmembrana di proteine, 130
    - 9.3.6 Previsione dei siti di interazione tra proteine, 131
    - 9.3.7 Previsione dei siti di scissione proteolitica dei peptidi segnale, 133
  - 9.4 La bioinformatica nella rete, 134
- Per saperne di più*, 135

## 10 Confronto tra strutture di proteine

- 10.1 Introduzione, 137
- 10.2 Le strutture tridimensionali delle proteine, 137
- 10.3 Sovrapposizione strutturale delle proteine, 138
- 10.4 Confronto strutturale attraverso proprietà geometriche locali, 141
  - 10.4.1 Confronto di strutture con algoritmi dinamici (SSAP), 142
  - 10.4.2 Confronto di strutture attraverso le mappe di contatto (DALI), 144

- 10.4.3 Confronto di strutture attraverso elementi di struttura secondaria, 146
- 10.4.4 Confronto e allineamento per estensione combinatoria (CE), 147

- 10.5 Confronti strutturali multipli, 148
  - 10.6 Ricerca di motivi di catene laterali, 149
  - 10.7 La bioinformatica nella rete, 152
- Riferimenti per saperne di più*, 152

## 11 Dalla sequenza alla struttura e alla funzione di una proteina

- 11.1 Introduzione, 153
  - 11.2 Metodi di riconoscimento di ripiegamento (*fold recognition*), 153
    - 11.2.1 Profili-3D, 155
    - 11.2.2 3D-PSSM, 159
    - 11.2.3 Potenziali di coppia e campi di forze medi, 160
    - 11.2.4 GenTHREADER, 161
    - 11.2.5 PROSPECT, 162
    - 11.2.6 Combinazione di più metodi: i *meta-server* e i *consensus-server*, 163
  - 11.3 Modellizzazione per omologia, 165
    - 11.3.1 Ricerca degli stampi strutturali, 166
    - 11.3.2 Selezione degli stampi strutturali, 166
    - 11.3.3 Allineamento tra la sequenza bersaglio e lo stampo o gli stampi strutturali, 166
    - 11.3.4 Costruzione del modello, 167
    - 11.3.5 Valutazione del modello, 172
    - 11.3.6 Accuratezza dei modelli, 178
  - 11.4 Previsione della funzione di una proteina, 178
    - 11.4.1 Analisi della topografia superficiale della proteina, 179
    - 11.4.2 Ricerca di motivi strutturali, 180
    - 11.4.3 Uso di analisi filogenetiche, 180
    - 11.4.4 Previsione di siti di interazione proteina-proteina, 181
    - 11.4.5 Combinazione di più metodi, 182
  - 11.5 La bioinformatica nella rete, 182
- Riferimenti per saperne di più*, 183

## 12 Meccanica e dinamica molecolare

- 12.1 Introduzione, 187
    - 12.1.1 Campi di forze, 187
    - 12.1.2 Parametrizzazione di tipi atomici: i potenziali e le cariche parziali, 190
    - 12.1.3 *Cut-off* delle energie di non legame nei campi di forze, 191
    - 12.1.4 Considerazioni generali sull'applicazione dei campi di forze alla simulazione e modellizzazione molecolare, 192
    - 12.1.5 Calcolo della superficie di energia potenziale attraverso i campi di forze, 192
  - 12.2 La minimizzazione energetica, 193
    - 12.2.1 Considerazioni generali sull'applicazione della minimizzazione energetica alla simulazione e modellizzazione molecolare, 196
  - 12.3 La dinamica molecolare, 197
    - 12.3.1 La presenza del solvente nelle simulazioni di dinamica molecolare, 200
  - 12.4 I limiti della meccanica e della dinamica molecolare, 201
  - 12.5 Meccanica e dinamica molecolare in rete, 202
- Riferimenti per saperne di più*, 203

## 13 La bioinformatica nella progettazione razionale di farmaci

- 13.1 Introduzione, 205
- 13.2 Approcci di CADD basati sulla conoscenza della struttura del ligando, 206
  - 13.2.1 Approcci di CADD basati sulla struttura 2D nota del ligando, 207
  - 13.2.2 Approcci di CADD basati sulla struttura 3D nota del ligando, 212
- 13.3 Approcci di CADD basati sulla struttura 3D nota del recettore, 216
  - 13.3.1 Trasformazioni geometriche e gradi di libertà nel *docking*, 217
  - 13.3.2 Algoritmi di ricerca sistematici e stocastici nel *docking*, 218
  - 13.3.3 Le funzioni di punteggio nel *docking*, 222
  - 13.3.4 Il recettore nel *docking*: mappe a griglia e flessibilità, 223
  - 13.3.5 Potenzialità e limiti nel *docking*, 224
  - 13.3.6 Il *docking* nello studio dell'interazione tra proteine, 225
- 13.4 CADD in rete, 226

*Riferimenti per saperne di più*, 227

### Appendice A

#### Introduzione alla programmazione con Python

- 1 Introduzione, 229
  - 1.1 Le qualità di Python, 229
  - 1.2 Installazione di Python, 230
- 2 Utilizzo di Python, 231
- 3 Tipi di dati, 231
  - 3.1 I numeri, 231

- 3.2 Le stringhe, 232
  - 3.3 Le liste, 234
  - 3.5 I *set* e i *frozenset*, 235
  - 3.6 I dizionari, 236
  - 4 Istruzioni di controllo, 237
    - 4.1 L'istruzione *if*, 237
    - 4.2 L'istruzione *while*, 238
    - 4.3 L'istruzione *for*, 238
    - 4.4 Le istruzioni *break*, *continue* e *pass*, 239
  - 5 Le funzioni, 239
    - 5.1 Gli argomenti opzionali, 240
    - 5.2 Variabili locali e globali, 240
    - 5.3 Documentare le funzioni, 241
  - 6 Le classi, 241
    - 6.1 Attributi e metodi delle classi, 242
    - 6.2 Attributi dell'oggetto e della classe, 242
    - 6.3 Attributi e metodi pubblici e privati, 243
    - 6.4 Composizione ed ereditarietà, 243
    - 6.5 Metodi speciali nelle classi, 245
  - 7 I moduli, 245
    - 7.1 Come importare i moduli, 245
    - 7.2 I pacchetti, 247
    - 7.3 NumPy, SciPy e Biopython, 248
  - 8 Input, output, eccezioni, 251
- Riferimenti per saperne di più*, 252

### Appendice B

#### Esercitazioni

Disponibile online

Indice analitico, 253