
Developing Bioinformatics Computer Skills

Cynthia Gibas and Per Jambeck

O'REILLY®

Beijing · Cambridge · Farnham · Köln · Paris · Sebastopol · Taipei · Tokyo

Table of Contents

<i>Preface</i>	<i>ix</i>
<i>1. Introduction</i>	<i>1</i>
<i>1. Biology in the Computer Age</i>	<i>3</i>
How Is Computing Changing Biology?	<i>4</i>
Isn't Bioinformatics Just About Building Databases?	<i>8</i>
What Does Informatics Mean to Biologists?	<i>12</i>
What Challenges Does Biology Offer Computer Scientists?	<i>13</i>
What Skills Should a Bioinformatician Have?	<i>13</i>
Why Should Biologists Use Computers?	<i>14</i>
How Can I Configure a PC to Do Bioinformatics Research?	<i>16</i>
What Information and Software Are Available?	<i>18</i>
Can I Learn a Programming Language Without Classes?	<i>18</i>
How Can I Use Web Information?	<i>19</i>
How Do I Understand Sequence Alignment Data?	<i>20</i>
How Do I Write a Program to Align Two Biological Sequences?	<i>20</i>
How Do I Predict Protein Structure from Sequence?	<i>21</i>
What Questions Can Bioinformatics Answer?	<i>21</i>
<i>2. Computational Approaches to Biological Questions</i>	<i>22</i>
Molecular Biology's Central Dogma	<i>22</i>
What Biologists Model	<i>27</i>
Why Biologists Model	<i>31</i>

Computational Methods Covered in This Book	32
A Computational Biology Experiment	38
II. <i>The Bioinformatics Workstation</i>	45
3. <i>Setting Up Your Workstation</i>	47
Working on a Unix system	47
Setting Up a Linux Workstation	50
How to Get Software Working	56
What Software Is Needed?	62
4. <i>Files and Directories in Unix</i>	64
Filesystem Basics	64
Commands for Working with Directories and Files	71
Working in a Multiuser Environment	79
5. <i>Working on a Unix System</i>	87
The Unix Shell	87
Issuing Commands on a Unix System	89
Viewing and Editing Files	94
Transformations and Filters	101
File Statistics and Comparisons	108
The Language of Regular Expressions	110
Unix Shell Scripts	113
Communicating with Other Computers	114
Playing Nicely with Others in a Shared Environment	119
III. <i>Tools for Bioinformatics</i>	131
6. <i>Biological Research on the Web</i>	133
Using Search Engines	134
Finding Scientific Articles	136
The Public Biological Databases	140
Searching Biological Databases	147
Depositing Data into the Public Databases	155
Finding Software	155
Judging the Quality of Information	156

7. Sequence Analysis, Pairwise Alignment, and Database Searching	159
Chemical Composition of Biomolecules	160
Composition of DNA and RNA	161
Watson and Crick Solve the Structure of DNA	161
Development of DNA Sequencing Methods	164
Genefinders and Feature Detection in DNA	169
DNA Translation	171
Pairwise Sequence Comparison	172
Sequence Queries Against Biological Databases	182
Multifunctional Tools for Sequence Analysis	188
8. Multiple Sequence Alignments, Trees, and Profiles	191
The Morphological to the Molecular	191
Multiple Sequence Alignment	193
Phylogenetic Analysis	199
Profiles and Motifs	205
9. Visualizing Protein Structures and Computing Structural Properties	215
A Word About Protein Structure Data	216
The Chemistry of Proteins	217
Web-Based Protein Structure Tools	229
Structure Visualization	231
Structure Classification	241
Structural Alignment	246
Structure Analysis	250
Solvent Accessibility and Interactions	254
Computing Physicochemical Properties	258
Structure Optimization	260
Protein Resource Databases	263
Putting It All Together	265
10. Predicting Protein Structure and Function from Sequence	268
Determining the Structures of Proteins	269
Predicting the Structures of Proteins	273
From 3D to 1D	275
Feature Detection in Protein Sequences	276
Secondary Structure Prediction	277
Predicting 3D Structure	283

Putting It All Together: A Protein Modeling Project	287
Summary	293
11. Tools for Genomics and Proteomics	294
From Sequencing Genes to Sequencing Genomes	296
Sequence Assembly	301
Accessing Genome Information on the Web	303
Annotating and Analyzing Whole Genome Sequences	307
Functional Genomics: New Data Analysis Challenges	310
Proteomics	317
Biochemical Pathway Databases	321
Modeling Kinetics and Physiology	325
Summary	327
IV. Databases and Visualization	329
12. Automating Data Analysis with Perl	331
Why Perl?	331
Perl Basics	332
Pattern Matching and Regular Expressions	339
Parsing BLAST Output Using Perl	340
Applying Perl to Bioinformatics	345
13. Building Biological Databases	350
Types of Databases	351
Database Software	359
Introduction to SQL	361
Installing the MySQL DBMS	366
Database Design	371
Developing Web-Based Software That Interacts with Databases	375
14. Visualization and Data Mining	383
Preparing Your Data	384
Viewing Graphics	385
Sequence Data Visualization	386
Networks and Pathway Visualization	388
Working with Numerical Data	390
Visualization: Summary	396
Data Mining and Biological Information	397
Bibliography	403
Index	409