Protein Bioinformatics. An Algorithmic Approach to Sequence and Structure Analysis

Description: Genomics and bioinformatics play an increasingly important and transformative role in medicine, society and agriculture. The mapping of the human genome has revealed 35,000 or so genes which might code for more than one protein, resulting in 100,000 proteins for the humans alone. Since proteins are attractive targets for developing drugs, efforts are now underway to map sequences and assign functions to many novel proteins. This book takes the novel approach to cover both the sequence and structure analysis of proteins in one volume and from an algorithmic perspective.

Key features of the book include:
- Provides a comprehensive introduction to the analysis of protein sequence and structure analysis.
- Takes an algorithmic approach, relying on computational methods rather than theoretical.
- Provides an integrated presentation of theory, examples, exercises and applications.
- Includes coverage of both protein structure, and sequence, analysis.
- Accessible enough for biologists, yet rigorous enough for computer scientists and mathematicians.
- Supported by a Web site featuring exercises, solutions, images, and computer programs.

Visit this website for exercises with solutions, computer programs, errata and additional material:
company website

Contents:

Preface.

Acknowledgements.

Part I: SEQUENCE ANALYSIS.

1. Pairwise Global Alignment of Sequences.

1.1 Alignment and Evolution.

1.2 What is an Alignment?

1.3 A Scoring Scheme for the Model.

1.4 Finding Highest-Scoring Alignments with Dynamic Programming.

1.4.1 Determine Hi,j.

1.4.2 Use of matrices.

1.4.3 Finding the alignments that give the highest score.

1.4.4 Gaps.

1.5 Scoring Matrices.

1.6 Scoring Gaps: Gap Penalties.

1.7 Dynamic Programming for General Gap Penalty.

1.8 Dynamic Programming for Affine Gap Penalty.

1.9 Alignment Score and Sequence Distance.
1.10 Exercises.

1.11 Bibliographic notes.

2 Pairwise Local Alignment and Database Search.

2.1 The Basic Operation: Comparing Two Sequences.

2.2 Dot Matrices.

2.2.1 Filtering.

2.2.2 Repeating segments.

2.3 Dynamic Programming.

2.3.1 Initialization.

2.3.2 Finding the best local alignments.

2.3.3 Algorithms.

2.3.4 Scoring matrices and gap penalties.

2.4 Database Search: BLAST.

2.4.1 The procedure.

2.4.2 Preprocess the query: make the word list.

2.4.3 Scanning the database sequences.

2.4.4 Extending to HSP.

2.4.5 Introducing gaps.

2.4.6 Algorithm.

2.5 Exercises.

2.6 Bibliographic notes.


3.1 Hypothesis Testing for Sequence Homology.

3.1.1 Random generation of sequences.

3.1.2 Use of Z values for estimating the statistical significance.

3.2 Statistical Distributions.

3.2.1 Poisson probability distribution.

3.2.2 Extreme value distributions.

3.3 Theoretical Analysis of Statistical Significance.

3.3.1 The P value has an extreme value distribution.

3.3.2 Theoretical analysis for database search.

3.4 Probability Distributions for Gapped Alignments.
3.5 Assessing and Comparing Programs for Database Search.

3.5.1 Sensitivity and specificity.

3.5.2 Discrimination power.

3.5.3 Using more sequences as queries.

3.6 Exercises.

3.7 Bibliographic notes.


4.1 Dynamic Programming.

4.1.1 SP score of multiple alignments.

4.1.2 A pruning algorithm for the DP solution.

4.2 Multiple Alignments and Phylogenetic Trees.

4.3 Phylogeny.

4.3.1 The number of different tree topologies.

4.3.2 Molecular clock theory.

4.3.3 Additive and ultrametric trees.

4.3.4 Different approaches for reconstructing phylogenetic trees.

4.3.5 Distance-based construction.

4.3.6 Rooting of trees.

4.3.7 Statistical test: bootstrapping.

4.4 Progressive Alignment.

4.4.1 Aligning two subset alignments.

4.4.2 Clustering.

4.4.3 Sequence weights.

4.4.4 CLUSTAL.

4.5 Other Approaches.

4.6 Exercises.

4.7 Bibliographic notes.

5. Scoring Matrices.

5.1 Scoring Matrices Based on Physio-Chemical Properties.

5.2 PAM Scoring Matrices.

5.2.1 The evolutionary model.
5.2.2 Calculate substitution matrix.
5.2.3 Matrices for general evolutionary time.
5.2.4 Measuring sequence similarity by use of Mt.
5.2.5 Odds matrices.
5.2.6 Scoring matrices (log-odds matrices).
5.2.7 Estimating the evolutionary distance.
5.3 BLOSUM Scoring Matrices.
5.3.1 Log-odds matrix.
5.3.2 Developing scoring matrices for different evolutionary distances.
5.4 Comparing BLOSUM and PAM Matrices.
5.5 Optimal Scoring Matrices.
5.5.1 Analysis for one sequence.
5.6 Exercises.
5.7 Bibliographic notes.
6. Profiles.
6.1 Constructing a Profile.
6.1.1 Notation.
6.1.2 Removing rows and columns.
6.1.3 Position weights.
6.1.4 Sequence weights.
6.1.5 Treating gaps.
6.2 Searching Databases with Profiles.
6.3 Iterated BLAST: PSI-BLAST.
6.3.1 Making the multiple alignment.
6.3.2 Constructing the profile.
6.4 HMM Profile.
6.4.1 Definitions for an HMM.
6.4.2 Constructing a profile HMM for a protein family.
6.4.3 Comparing a sequence with an HMM.
6.4.4 Protein family databases.
6.5 Exercises.
6.6 Bibliographic notes.
7. Sequence Patterns.
7.1 The PROSITE Language.
7.2 Exact/Approximate Matching.
7.3 Defining Pattern Classes by Imposing Constraints.
7.4 Pattern Scoring: Information Theory.
7.4.1 Information theory.
7.4.2 Scoring patterns.
7.5 Generalization and Specialization.
7.6 Pattern Discovery: Introduction.
7.7 Comparison-Based Methods.
7.7.1 Pivot-based methods.
7.7.2 Tree progressive methods.
7.8 Pattern-Driven Methods: Pratt.
7.8.1 The main procedure.
7.8.2 Preprocessing.
7.8.3 The pattern space.
7.8.4 Searching.
7.8.5 Ambiguous components.
7.8.6 Specialization.
7.8.7 Pattern scoring.
7.9 Exercises.
7.10 Bibliographic notes.

Part II: STRUCTURE ANALYSIS
8. Structures and Structure Descriptions.
8.1 Units of Structure Descriptions.
8.2 Coordinates.
8.3 Distance Matrices.
8.4 Torsion Angles.
8.5 Coarse Level Description.
8.5.1 Line segments (sticks).
8.5.2 Ellipsoid.
10.2.1 Measuring the similarity of distance (sub)matrices.
10.3 Exercises.
10.4 Bibliographic notes.

11. Clustering: Combining Local Similarities.
11.1 Compatibility and Consistency.
11.2 Searching for Seed Matches.
11.3 Consistency.
11.3.1 Test for consistency.
11.3.2 Overlapping clusters.
11.4 Clustering Algorithms.
11.4.1 Linear clustering.
11.4.2 Hierarchical clustering.
11.5 Clustering by Use of Transformations.
11.5.1 Comparing transformations.
11.5.2 Calculating the new transformation.
11.5.3 Algorithm.
11.6 Clustering by Use of Relations.
11.6.1 How many relations to compare?
11.6.2 Geometric relation.
11.6.3 Distance relation.
11.6.4 Use of graph theory.
11.7 Refinement.
11.8 Exercises.
11.9 Bibliographic notes.

12. Significance and Assessment of Structure Comparisons.
12.1 Constructing Random Structure Models.
12.1.1 Use of distance geometry.
12.2 Use of Structure Databases.
12.2.1 Constructing nonredundant subsets.
12.2.2 Demarcation line for similarity.
12.3 Reversing the Protein Chain.
12.4 Randomized Alignment Models.
12.5 Assessing Comparison and Scoring Methods.
12.6 Is RMSD Suitable for Scoring?
12.7 Scoring and Biological Significance.
12.8 Exercises.
12.9 Bibliographic notes.

13. Multiple Structure Comparison.
13.1 Multiple Superposition.
13.2 Progressive Structure Alignment.
13.2.1 Scoring.
13.2.2 Construction of consensus.
13.3 Finding a Common Core from a Multiple Alignment.
13.4 Discovering Common Cores.
13.4.1 Finding the multiple seed matches.
13.4.2 Pairwise clustering.
13.4.3 Determining common cores.
13.4.4 Scoring clusters.
13.5 Local Structure Patterns.
13.5.1 Local packing patterns.
13.5.2 Discovering packing patterns.
13.5.3 The approach.
13.5.4 Scoring the packing motifs.
13.6 Exercises.
13.7 Bibliographic notes.

14. Protein Structure Classification.
14.1 Protein Domains.
14.2 An Ising Model for Domain Identification.
14.3 Domain Classes.
14.3.1 Mainly-? domains.
14.3.2 Mainly-? domains.
14.3.3 ?-? domains.
14.4 Folds.
14.5 Automatic Approaches to Classification.
14.6 Databases for Structure Classification.
14.7 FSSP-Dali Domain Dictionary.
14.8 CATH.
14.8.1 Domains.
14.8.2 Class.
14.8.3 Architecture.
14.8.4 Topology (fold family).
14.8.5 Homologous superfamily.
14.8.6 Sequence families.
14.8.7 The CATH classification procedure.
14.9 Classification Based on Sticks.
14.10 Exercises.
14.11 Bibliographic notes.

Part III: SEQUENCE-STRUCTURE ANALYSIS.

15.1 Protein Secondary Structure Prediction.
15.1.1 Artificial neural networks.
15.1.2 The PHD program.
15.1.3 Accuracy in secondary structure prediction.
15.2 Threading.
15.3 Methods Based on Sequence Alignment.
15.3.1 The 3D–1D matching method.
15.3.2 The FUGUE method.
15.4 Methods Using 3D Interactions.
15.4.1 Potentials of mean force.
15.4.2 Towards modelling methods.
15.5 Alignment Methods.
15.5.1 Frozen approximation.
15.5.2 Double Dynamic Programming.
15.6 Multiple Sequence/Structure Threading.
Index.


Order by Fax - using the form below

Order by Post - print the order form below and send to

Research and Markets,
Guinness Centre,
Taylors Lane,
Dublin 8,
Ireland.
Fax Order Form
To place an order via fax simply print this form, fill in the information below and fax the completed form to 646-607-1907 (from USA) or +353-1-481-1716 (from Rest of World). If you have any questions please visit http://www.researchandmarkets.com/contact/

Order Information
Please verify that the product information is correct.

<table>
<thead>
<tr>
<th>Product Name:</th>
<th>Protein Bioinformatics. An Algorithmic Approach to Sequence and Structure Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web Address:</td>
<td><a href="http://www.researchandmarkets.com/reports/2180857/">http://www.researchandmarkets.com/reports/2180857/</a></td>
</tr>
<tr>
<td>Office Code:</td>
<td>SC93RZTX</td>
</tr>
</tbody>
</table>

Product Format
Please select the product format and quantity you require:

<table>
<thead>
<tr>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard Copy (Hard Back):</td>
</tr>
</tbody>
</table>

* Shipping/Handling is only charged once per order.

Contact Information
Please enter all the information below in BLOCK CAPITALS

<table>
<thead>
<tr>
<th>Title:</th>
<th>Mr</th>
<th>Mrs</th>
<th>Dr</th>
<th>Miss</th>
<th>Ms</th>
<th>Prof</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Name:</td>
<td></td>
<td></td>
<td></td>
<td>Last Name:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email Address: *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job Title:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organisation:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Address:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>City:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postal / Zip Code:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phone Number:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fax Number:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Please refrain from using free email accounts when ordering (e.g. Yahoo, Hotmail, AOL)
Please indicate the payment method you would like to use by selecting the appropriate box.

☐ Pay by credit card: You will receive an email with a link to a secure webpage to enter your credit card details.

☐ Pay by check: Please post the check, accompanied by this form, to:

Research and Markets,
Guinness Center,
Taylors Lane,
Dublin 8,
Ireland.

☐ Pay by wire transfer: Please transfer funds to:

Account number 833 130 83
Sort code 98-53-30
Swift code ULSBIE2D
IBAN number IE78ULSB98533083313083
Bank Address Ulster Bank,
27-35 Main Street,
Blackrock,
Co. Dublin,
Ireland.

If you have a Marketing Code please enter it below:

Marketing Code: _______________________

Please note that by ordering from Research and Markets you are agreeing to our Terms and Conditions at http://www.researchandmarkets.com/info/terms.asp

Please fax this form to:
(646) 607-1907 or (646) 964-6609 - From USA
+353-1-481-1716 or +353-1-653-1571 - From Rest of World