

Link to Order: http://fivephoton.com/index.php?route=product/product&path=37&product_id=55

Price: \$109.95

Website: www.fivephoton.com. Tel: 800-462-4507

PROTEIN STRUCTURE PREDICTION

Bioinformatic Approach

edited by

IGOR F. TSIGELNY

Table of Contents:

Preface xv

List of Contributors xxi

Part I. CONCEPTS OF PROTEIN STRUCTURE PREDICTION 1

A. Prediction Methods and Systems 3

1. Computational Studies of Protein Structure and Function Using Threading Program PROSPECT 5

Dong Xu and Ying Xu

1.1. Introduction 5

1.2. Method of PROSPECT 10

1.2.1. Threading Templates 11

1.2.2. Energy Function 12

1.2.3. Threading Algorithm 13

1.2.4. Confidence Assessment of Threading Results 15

1.3. Protocols of Using PROSPECT 17

1.3.1. Pre-Processing before Running PROSPECT 18

1.3.2. Running PROSPECT 20

1.3.3. Human Evaluation 21

1.3.4. Manual Refinement 25

1.3.5. Structure-Based Functional Inference 26

1.4. Performance of PROSPECT 29

1.4.1. Testing of PROSPECT Using Known Structures in PDB 29

1.4.2. Blind Test in CASP 30

1.5. Application of PROSPECT in Protein Studies 34

1.5.1. Human Vitronectin 34

1.5.2. Human DNA-Activated Protein Kinase 35

1.5.3. Yeast PTR3 Protein 36

1.6. Summary 37

2. Bayesian Approach to Protein Fold Recognition: Building Protein Structural Models from Bits and Pieces 43

Jadwiga Bienkowska, Hongxian He, Robert G. Rogers Jr., and Lihua Yu

2.1. Introduction 45

2.2. Fundamentals of DSMs and HMMs 49

2.2.1. Representation of Protein Structure by a DSM 50

2.2.2. Mathematical Representation of a DSM 51

2.2.3. Measures of Compatibility of a Protein Sequence with a DSM 52

2.3. Automated Generation of Protein Structural Templates 53

2.3.1. Criteria for Selecting Structural Information 54

2.3.2. Candidate Structural Quantities 55

2.3.3. Classification of Structural States	57
2.4. Automated Design of a Structural DSM from a Structural Template	60
2	
2.4.1. Design Principles	60
2.4.2. Secondary Structure Submodels	62
2.4.3. Construction of DSM from the Structural Template	65
2.4.4. Using Structural Alignments and Multiple Structural Templates in Building DSM	65
2.5. Automatic Pattern Embedding in a DSM	66
2.5.1. Automated Pattern Generation and Selection	67
2.5.2. Look-Ahead	69
2.6. A Bayesian Approach to Fold Recognition	70
2.6.1. The Filtering Algorithm	70
2.6.2. Prior Model Probabilities	72
2.7. Results	75
2.7.1. Comparing the Bayesian Approach and Total Alignment Probability with Other Methods	75
2.7.2. Results of Automatic Pattern Embedding	77
2.7.3. Comparison of Different Assignments of Prior Probabilities	79
2.8. Strategies for Defeating the Combinatorial Explosion	80
3. Three-Dimensional Structure Prediction Using Simplified Structure Models and Bayesian Block Fragments	85
<i>Jun Zhu and Roland Lüthy</i>	
3.1. Introduction	87
3.2. Methods	89
3.2.1. Simplified Backbone Angle Representation of 3D Structures	89
3.2.2. Block Selection	90
3.2.3. Energy Functions	94
3.2.4. Energy Minimization	101
3.2.5. Using Information from Bayesian Blocks	102
3.2.6. Enforcing Secondary Structures	103
3.3. Examples	103
4. Protein Structure Prediction Using Hidden Markov Model Structural Libraries	109
<i>Igor Tsigelny, Yuriy Sharikov, and Lynn F. Ten Eyck</i>	
4.1. Introduction	111
4.2. Structural Hidden Markov Model Libraries	112
4.3. Decision Tree	114
4.3.1. Search for the Best HMM	114
4.3.2. Searching within the Structural Alignment	117
4.4. Program Testing	120
4.5. Prediction of Unsolved Structures	121
5. The Role of Sequence Information in Protein Structure Prediction	125
<i>Damien Devos, Florencio Pazos, Osvaldo Olmea, David de Juan, Osvaldo Graña, Jose M. Fernández, and Alfonso Valencia</i>	

5.1. Introduction	127
5.1.1. Information Contained in Multiple Sequence Alignments in Protein Families	127
	3
5.2. Automated Generation of Protein Structural Templates	128
5.3. Distribution of Informative Positions in Protein Structures	130
5.4. Informative Positions in Protein Structure Models	132
5.5. A Threading Server That Filters Models with Multiple Sequence Alignments Information	128
5.6. A First Field Evaluation of the Server, the CAFASP Results	130
5.7. A CAFASP Example of the Use of Sequence Information	132
5.8. Training Neural Networks for the Discrimination of Wrong Threading Models Using Sequence	130
5.9. Conclusions	132
6. Protein Fold Recognition and Comparative Modeling Using HOMSTRAD, JOY, and FUGUE	143
<i>Ricardo Núñez Miguel, Jiye Shi, and Kenji Mizuguchi</i>	
6.1. Introduction	145
6.2. Overview	149
6.3. Identification of Homologues	150
6.4. Generating Sequence-Structure Alignment	152
6.5. Example	153
6.5.1. Searching for Homologues	153
6.5.2. Alignment	157
6.5.3. Modeling	161
6.5.4. Heteroatoms	161
6.5.5. Refinements	162
6.5.6. Model Validation	162
6.5.7. Model	163
6.6. Conclusion	165
7. Fully Automated Protein Tertiary Structure Prediction Using Fourier Transform Spectral Methods	171
<i>Carlos Adriel Del Carpio Muñoz and Atsushi Yoshimori</i>	
7.1. Sequence Alignment and Protein Structure Modeling	173
7.2. Protein Function and Structure Elucidation by Spectral Analysis	176
7.3. Spectral Analysis and Folding Pattern Recognition	179
7.3.1. Spectral Representation of Protein Primary Structures	180
7.3.2. Spectral Alignment and Protein Structure Similarity	184
7.3.3. Automatic Protein Folding Pattern Recognition	186
7.4. Automatic Classification of Protein Foldings	188
7.4.1. Dominant Physicochemical Parameters	188
7.4.2. Classification of Protein Folding by Spectral Analysis	191
7.5. Protein Folding Pattern Recognition by Spectral Analysis	195
8. From the Building Blocks Folding Model to Protein Structure Prediction	201
<i>Nurit Haspel, Chung-Jung Tsai, Haim Wolfson, and Ruth Nussinov</i>	
8.1. Introduction	203

8.2. Protein Folding: A Process of Intra-Molecular Building Block Recognition **205**

8.3. Experimental and Theoretical Support for the Building Block Concept **206**

8.4. The Building Block Cutting Algorithm **209**

8.5. The Scoring Function **210**

4

8.6. The Cutting Procedure **211**

8.7. Critical Building Blocks **213**

8.8. From the Building Block Folding Model to Structure Prediction: The Scheme **214**

8.9. Conclusions **220**

9. Protein Threading Statistics: An Attempt to Assess the Significance of a Fold Assignment to a Sequence 227

Antoine Marin, Joël Pothier, Karel Zimmermann, and Jean-François Gibrat

9.1. Introduction **229**

9.2. Method **232**

9.2.1. Library of “Cores” 232

9.2.2. Development of a Score Function 233

9.2.3. Combinatorial Optimization Algorithm 239

9.2.4. Empirical Distribution of Scores 241

9.2.5. Development of a Benchmark Database 244

9.3. Results **247**

9.4. Discussion **254**

9.4.1. Use of Filters 254

9.4.2. Difficulty of the Benchmark 255

9.4.3. Statistical Criterion 256

9.4.4. Present Limits of the Method 258

9.5. Conclusion **259**

10. Protein Structure Prediction by Threading: Force Field Philosophy, Approaches to Alignment 263

Thomas Huber and Andrew E. Torda

10.1. Introduction **265**

10.3.1. Common Methodology 267

10.2. Force Field Based Scoring **269**

10.3. Parameterizing Force Fields **271**

10.3.1. Physically-Based Potential Energies 271

10.3.2. Potentials of Mean Force 272

10.3.3. Optimized Force Fields 273

10.4. Alignment Philosophy **278**

10.4.1. Common Alignment and Score Methods 278

10.4.2. Sausage Alignments 279

10.5. Beyond Pairwise Terms **280**

10.6. Template Libraries **285**

10.7. Further Outlook and Speculation **289**

11. Predicting Protein Structure Using SAM, UCSC’s Hidden Markov Model Tools 297

Kevin Karplus

- 11.1. A Naive View of Protein Structure Prediction **299**
- 11.2. Fold Recognition **301**
- 11.3. Hidden Markov Models **302**
 - 11.3.1. Multitrack Hidden Markov Models 305
 - 11.3.2. Statistical Significance for Hidden Markov Models 307
- 5
- 11.4. Using SAM-T2K for Superfamily Modeling **308**
- 11.5. Improved Verification of Homology **312**
- 11.6. Family-Level Multiple Alignments **314**
- 11.7. Modeling Non-Contiguous Domains **315**
- 11.8. Building an HMM from a Structural Alignment **316**
- 11.9. Improving Existing Multiple Alignments **319**
- 11.10. Creating a Multiple Alignment from Unaligned Sequences **319**
- 11.11. Conclusions **320**
- 12. Local Genome Organization, Gene Expression, and Structural Genomics: Evolution at Work 325**
Wayne Volkmuth and Nickolai Alexandrov
 - 12.1. Introduction **327**
 - 12.2. Methods **329**
 - 12.2.1. Genomes 329
 - 12.2.2. Microarray Expression Data 329
 - 12.2.3. Fold Assignment 331
 - 12.2.4. Non-Redundant Set of Proteins 333
 - 12.2.5. Fold Enrichment Along the Genome 333
 - 12.2.6. Fold Enrichment for Genes with Similar Patterns of Expression 333
 - 12.3. Results **333**
 - 12.3.1. Fold Enrichment Along the Genome 333
 - 12.3.2. Fold Enrichment for Genes with Similar Patterns of Expression 333
 - 12.3. Summary and Conclusions **334**
- 13. Protein Structure Prediction on the Basis of Combinatorial Peptide Library Screening 341**
Igor Tsigelny, Yuriy Sharikov, Vladimir Kotlovyi, Michael Kelner, and Lynn F. Ten Eyck
 - 13.1. Concept of the Comprehensive System **343**
 - 13.2. HMM-ELONGATOR **345**
 - 13.2.1. Problem Description 345
 - 13.2.2. Elongation Strategies 346
- B. Consensus Structure Prediction 353**
- 14. A User's Guide to Fold Recognition 355**
Naomi Siew and Daniel Fischer
 - 14.1. Introduction **357**
 - 14.2. Examples of Using Fold Recognition for Biological Research **358**
 - 14.2.1. Plant Resistance Gene Products 359
 - 14.2.2. Acetohydroxyacid Synthase 360
 - 14.2.3. Endothelial Cell Protein C/Activated Protein C Receptor 361
 - 14.3. How to Fold Recognize? **363**
 - 14.3.1. Searching for Homologues of Known Structure 364

14.3.2. Running Your Favorite Fold Recognition Method 365

14.3.3. Running Other Methods 368

14.3.4. Why Run More Than One Method? 369

14.3.5. 3D-Shotgun Meta-Predictor 370

14.4. Summary **370**

15. Structure Prediction Meta Server 377

6

Leszek Rychlewski

15.1. Introduction **379**

15.2. The Meta Server **381**

15.2.1. User Input and Job Status Display 382

15.2.2. Job Deposition and Administration 382

15.2.3. Request Submission Queuing 384

15.2.4. Blast-Filter 385

15.2.5. Local and Remote Prediction Services 385

15.2.6. Raw Output Converters 387

15.2.7. Visualization and Linking 389

15.2.8. Interfaces 389

15.3. Discussion **390**

Part II. METHODS OF STRUCTURE AND SEQUENCE ALIGNMENT 395

16. Improved Fold Recognition by Using the PCONS Consensus Approach 397

Huisheng Fang, Björn Wallin, Jesper Lundström, Christer von Wöwern, and Arne Elofsson

16.1. Introduction **399**

16.2. Why are Manual Predictions Better? **401**

16.2.1. Biological Knowledge 401

16.2.2. Structural Analysis 401

16.2.3. Consensus Analysis 402

16.3. Consensus Predictions in CASP4 **403**

16.4. Pcons **405**

16.4.1. Collection of Publicly Available Models 406

16.4.2. Structural Comparison 406

16.4.3. Prediction of Quality of the Models 407

16.5. Performance of Pcons **408**

16.5.1. Performance in LiveBench-2 409

16.5.2. Why Does Pcons Perform Better? 411

16.6. Pcons-II **412**

16.6.1. Improvements Using More Servers 412

16.6.2. Speed-Up of Structural Comparisons 412

16.6.3. Using Better Statistics 413

16.6.4. Improvements Using Linear Regression 413

16.7. Summary **414**

17. New Insights into Protein Fold Space and Sequence-Structure Relationships 417

Ilya N. Shindyalov and Philip E. Bourne

17.1. Introduction **419**

- 17.2. Overview of CE Sequence-Structure Space **420**
- 17.3. Scop vs. CE Fold Space Comparison **421**
- 17.4. Analysis of Structure Redundancy **422**
 - 17.4.1. Size of NR Set as a Function of Criteria Used 423
 - 17.4.2. Characterization of Chains Excluded from the Set 423
 - 17.4.3. Characterization of Similarity Between Chains in the Set 424
 - 17.4.4. Complementary Sequence and Structure NR Sets 428
 - 7
 - 17.4.5. Combined NR Set 428
- 18. A Flexible Method for Structural Alignment: Applications to Structure Prediction Assessments 431**
Vladimir Kotlovyy, Igor Tsigelny, and Lynn Ten Eyck
 - 18.1. Introduction **433**
 - 18.2. Theoretical Background **435**
 - 18.3. Algorithms and Their Implementation **438**
 - 18.4. Representation of Data in XML Forms **440**
 - 18.5. Timing **442**
 - 18.6. Web-Servers **443**
 - 18.7. Illustrative Examples **447**
- 19. Comparative Analysis of Protein Structure: New Concepts and Approaches for Multiple Structure Alignment 449**
Chittibabu Guda, Eric D. Scheeff, Philip E. Bourne, and Ilya N. Shindyalov
 - 19.1. Introduction **451**
 - 19.2. Algorithm for Aligning Multiple Protein Structures Using Monte Carlo Optimization **452**
 - 19.2.1. Scoring Function **452**
 - 19.3. Approaches for Optimization of Multiple Structure Alignment **453**
 - 19.3.1. Effect of Weights Based on Number of Residues on Alignment Length and Alignment Distance 453
 - 19.4. Analysis of Specific Protein Families **455**
 - 19.4.1. Analysis of an Alignment of Protein Kinases 455
 - 19.4.2. Analysis of an Alignment of Aspartic Proteinases 458
 - 19.5. Summary **459**
- 20. Comparative Analysis of Protein Structure: Automated vs. Manual Alignment of the Protein Kinase Family 463**
Eric D. Scheeff, Philip E. Bourne, and Ilya N. Shindyalov
 - 20.1. Introduction **465**
 - 20.2. The Challenge of Automated Protein Structure Alignment **466**
 - 20.3. A Case Study: Alignment of the Eukaryotic Protein Kinases and Their Relatives **467**
 - 20.4. An Example of an Automated Alignment: The Combinatorial Extension Algorithm **468**
 - 20.5. Parameters for the Determination of an “Optimal” Structure Alignment **470**
 - 20.6. Comparison of CE Alignments with Manual Alignments **471**

20.7. Conclusion **475**
Index **479**