

Table of Contents

Foreword VII

Preface XII

Acknowledgments XV

Introduction XVI

1 Sequence, Function, and Structure Relationships 1

- 1.1 Introduction 1
- 1.2 Protein Structure 4
- 1.3 The Properties of Amino Acids 12
- 1.4 Experimental Determination of Protein Structures 14
- 1.5 The PDB Protein Structure Data Archive 20
- 1.6 Classification of Protein Structures 22
- 1.7 The Protein-folding Problem 24
- 1.8 Inference of Function from Structure 27
- 1.9 The Evolution of Protein Function 29
- 1.10 The Evolution of Protein Structure 34
- 1.11 Relationship Between Evolution of Sequence and Evolution of Structure 37

2 Reliability of Methods for Prediction of Protein Structure 41

- 2.1 Introduction 41
- 2.2 Prediction of Secondary Structure 43
- 2.3 Prediction of Tertiary Structure 46
- 2.4 Benchmarking a Prediction Method 50
- 2.5 Blind Automatic Assessments 51
- 2.6 The CASP Experiments 51

3	Ab-initio Methods for Prediction of Protein Structures	55
3.1	The Energy of a Protein Configuration	55
3.2	Interactions and Energies	55
3.3	Covalent Interactions	56
3.4	Electrostatic Interactions	58
3.5	Potential-energy Functions	62
3.6	Statistical-mechanics Potentials	62
3.7	Energy Minimization	65
3.8	Molecular Dynamics	66
3.9	Other Search Methods: Monte Carlo and Genetic Algorithms	67
3.10	Effectiveness of Ab-initio Methods for Folding a Protein	70
4	Evolutionary-based Methods for Predicting Protein Structure: Comparative Modeling	73
4.1	Introduction	73
4.2	Theoretical Basis of Comparative Modeling	75
4.3	Detection of Evolutionary Relationships from Sequences	77
4.4	The Needleman and Wunsch Algorithm	79
4.5	Substitution Matrices	81
4.6	Template(s) Identification Part I	84
4.7	The Problem of Domains	90
4.8	Alignment	91
4.9	Template(s) Identification Part II	96
4.10	Building the Main Chain of the Core	97
4.11	Building Structurally Divergent Regions	98
4.12	A Special Case: Immunoglobulins	102
4.13	Side-chains	106
4.14	Model Optimization	107
4.15	Other Approaches	108
4.16	Effectiveness of Comparative Modeling Methods	109
5	Sequence-Structure Fitness Identification: Fold-recognition Methods	117
5.1	The Theoretical Basis of Fold-recognition	117
5.2	Profile-based Methods for Fold-recognition	119
5.3	Threading Methods	121
5.4	Profile-Profile Methods	124
5.5	Construction and Optimization of the Model	124
6	Methods Used to Predict New Folds: Fragment-based Methods	127
6.1	Introduction	127
6.2	Fragment-based Methods	128
6.3	Splitting the Sequence into Fragments and Selecting Fragments from the Database	130
6.4	Generation of Structures	135

7	Low-dimensionality Prediction: Secondary Structure and Contact Prediction	137
7.1	Introduction	137
7.2	A Short History of Secondary structure Prediction Methods	140
7.3	Automatic learning Methods	142
7.3.1	Artificial Neural Networks	142
7.3.2	Support Vector Machines	148
7.4	Secondary structure Prediction Methods Based on Automatic Learning Techniques	150
7.5	Prediction of Long-range Contacts	153
8	Membrane Proteins	159
8.1	Introduction	159
8.2	Prediction of the Secondary Structure of Membrane Proteins	162
8.3	The Hydrophobic Moment	165
8.4	Prediction of the Topology of Membrane Proteins	166
9	Applications and Examples	169
9.1	Introduction	169
9.2	Early Attempts	169
9.3	The HIV Protease	171
9.4	Leptin and Obesity	174
9.5	The Envelope Glycoprotein of the Hepatitis C Virus	176
9.6	HCV Protease	178
9.7	Cyclic Nucleotide Gated Channels	181
9.8	The Effectiveness of Models of Proteins in Drug Discovery	183
9.9	The Effectiveness of Models of Proteins in X-ray Structure Solution	186
	Conclusions	188
	Glossary	190
	Index	201