

Contents

Computational Methods for Protein Structure Prediction and Fold Recognition	1
I. CYMERMAN, M. FEDER, M. PAWŁOWSKI, M.A. KUROWSKI, J.M. BUJNICKI	
1 Primary Structure Analysis	1
1.1 Database Searches	1
1.2 Protein Domain Identification	3
1.3 Prediction of Disordered Regions	5
2 Secondary Structure Prediction	5
2.1 Helices and Strands and Otherwise	5
2.2 Transmembrane Helices	8
3 Protein Fold Recognition	8
4 Predicting All-in-One-Go	12
5 Pitfalls of Fold Recognition	14
References	16
‘Meta’ Approaches to Protein Structure Prediction	23
J.M. BUJNICKI, D. FISCHER	
1 Introduction	23
2 The Utility of Servers as Standard Tools for Protein Structure Prediction	24
2.1 Consensus ‘Meta-Predictors’: Is the Whole Greater Than the Sum of the Parts?	25
2.2 Automated Meta-Predictors	26
2.3 Hybrid Methods: Going Beyond the “Simple Selection” of Models	29
3 Future Prospects	31
References	32

From Molecular Modeling to Drug Design	35
M. COHEN-GONSAUD, V. CATHERINOT, G. LABESSE, D. DOUGUET	
1 Introduction	35
1.1 General Context	35
1.2 Comparative Modeling	36
1.3 Drug Design and Screening	37
2 Comparative Modeling	38
2.1 Sequence Gathering and Alignment	38
2.1.1 Sequence Database Searches	38
2.1.2 Multiple Sequence Alignments	39
2.2 Structural Alignments	39
2.2.1 Fold Recognition	40
2.2.2 Structural Alignment Refinement	40
2.2.3 Active Site Recognition	41
2.2.4 A Biological Application	42
2.3 Complete Model Achievement	43
2.3.1 Global Structure Modeling	44
2.3.2 Optimization of Side-Chain Conformation	44
2.3.3 Insertions/Deletions Building	46
2.3.4 Modeling Protein Quaternary Structures	47
2.3.5 Energy Minimization and Molecular Dynamics	48
2.4 Model Validation	49
2.4.1 Theoretical Model Validation	49
2.4.2 Ligand-Based Model Selection	50
2.4.3 Experimental Evaluation of Models	50
2.5 Current Limitations	51
3 Model-Based Drug Design	52
3.1 Comparative Drug Design	53
3.2 Docking Methodologies	55
3.2.1 Knowledge-Based Potentials	55
3.2.2 Regression-Based (or Empirical) Methods	56
3.2.3 Physics-Based Methods	56
3.2.4 Flexible Models	57
3.2.5 Fragment-Based Drug Design	58
3.3 Virtual Screening Using Models	58
3.3.1 Docking Onto Medium Resolution Models	58
3.3.2 Docking Onto High-Resolution Models	59
3.4 Pharmacogenomic Applications	60
3.4.1 A Challenging Application: the GPCRs	60
3.4.2 Family-Wide Docking	60
3.4.3 Side Effect Predictions	61
3.4.4 Drug Metabolization Predictions	61
4 Conclusions	62
References	63

**Structure Determination of Macromolecular Complexes
by Experiment and Computation 73**
F. ALBER, N. ESWAR, A. SALI

1 Introduction 73

2 Hybrid Approaches to Determination
of Assembly Structures 77

2.1 Modeling the Low-Resolution Structures of Assemblies . . . 78

2.1.1 Representation of Molecular Assemblies 80

2.1.2 Scoring Function Consisting of Individual Spatial Restraints 80

2.1.3 Optimization of the Scoring Function 81

2.1.4 Analysis of the Models 81

3 Comparative Modeling for Structure Determination
of Macromolecular Complexes 82

3.1 Automated Comparative Protein Structure Modeling 82

3.2 Accuracy of Comparative Models 84

3.3 Prediction of Model Accuracy 86

3.4 Docking of Comparative Models
into Low-Resolution Cryo-EM Maps 86

3.5 Example 1: A Partial Molecular Model
of the 80S Ribosome from *Saccharomyces cerevisiae* 88

3.6 Example 2: A Molecular Model of the *E. coli* 70S Ribosome . 90

4 Conclusions 91

References 92

Modeling Protein Folding Pathways 97
C. BYSTROFF, Y. SHAO

1 Introduction: Darwin Versus Boltzmann 95

1.1 Protein Folding Pathway History 98

2 Knowledge-Based Models for Folding Pathways 99

2.1 I-sites: A Library of Folding Initiation Site Motifs 99

2.2 HMMSTR: A Hidden Markov Model
for Grammatical Structure 100

3 ROSETTA: Folding Simulations Using a Fragment Library . 101

3.1 Results of Fully Automated I-SITES/ROSETTA
Simulations 102

3.1.1 Summary 102

3.1.2 Topologically Correct Large Fragment
Predictions Are Found 103

3.1.3 Good Local Structure Correlates Weakly
with Good Tertiary Structure 004

3.1.4	Average Contact Order Is Too Low	105
3.1.5	How Could Automated ROSETTA Be Improved?	105
4	HMMSTR-CM: Folding Pathways Using Contact Maps	106
4.1	A Knowledge-Based Potential for Motif–Motif Interactions	106
4.2	Fold Recognition Using Contact Potential Maps	108
4.3	Consensus and Composite Contact Map Predictions	111
4.4	Ab Initio Rule-Based Pathway Predictions	111
4.5	Selected Results of HMMSTR-CM Blind Structure Predictions	112
4.5.1	A Prediction Using Templates and a Pathway	113
4.5.2	A Prediction Using Several Templates	113
4.5.3	Correct Prediction Using Only the Folding Pathway	114
4.5.4	False Prediction Using the Folding Pathway. What Went Wrong?	116
4.6	Future Directions for HMMSTR-CM	117
5	Conclusions	118
	References	118

Structural Bioinformatics and NMR Structure Determination 123

J.P. LINGE, M. NILGES

1	Introduction: NMR and Structural Bioinformatics	123
2	Algorithms for NMR Structure Calculation	124
2.1	Distance Geometry and Data Consistency	124
2.2	Nonlinear Optimization	125
2.3	Sampling Conformational Space	126
2.4	Modelling Structures with Limited Data Sets	126
3	Internal Dynamics and NMR Structure Determination	127
3.1	Calculating NMR Parameters from Molecular Dynamics Simulations	127
3.2	Inferring Dynamics from NMR Data	127
4	Structure Validation	128
5	Structural Genomics by NMR	129
5.1	Automated Assignment and Data Analysis	129
5.2	Collaborative Computing Project for NMR (CCPN)	130
5.3	SPINS	132
6	Databanks and Databases	132
6.1	BioMagResBank and PDB/RCSB	133
7	Conclusions	133
	References	134

Bioinformatics-Guided Identification and Experimental		
Characterization of Novel RNA Methyltransferases		139
J.M. BUJNICKI, L. DROOGMANS, H. GROSJEAN, S.K. PURUSHOTHAMAN, B. LAPEYRE		
1	Introduction	139
1.1	Diversity of Methylated Nucleosides in RNA	139
1.2	RNA Methyltransferases	141
1.3	Structural Biology of RNA MTases and Their Relatives . . .	142
2	Traditional and Novel Approaches to Identification of New RNA-Modification Enzymes	145
3	Bioinformatics: Terminology, Methodology, and Applications to RNA MTases	146
3.1	The Top-Down Approach	149
3.1.1	Top-Down Search for Novel RNA:m ⁵ C MTases in Yeast . . .	151
3.1.2	Top-Down Search for Bacterial and Archaeal m ¹ A MTases .	152
3.1.3	Top-Down Search for Novel Yeast 2'-O-MTases	153
3.2	The Bottom-Up Approach	155
3.2.1	Bottom-Up Search for New Yeast RNA MTases	157
4	Conclusions	160
	References	162

Finding Missing tRNA Modification Genes:		
A Comparative Genomics Goldmine		169
V. DE CRÉCY-LAGARD		

1	Missing tRNA Modification Genes	169
1.1	tRNA Modifications	169
1.2	Compilation of the Missing tRNA Modification Genes . . .	170
2	Comparative Genomics: an Emerging Tool to Identify Missing Genes	173
3	Finding Genes for Simple tRNA Modifications	175
3.1	Paralog- and Ortholog-Based Identifications	175
3.2	Comparative Genomics-Based Identifications	176
4	Finding Complex Modification Pathway Genes	178
4.1	Finding Missing Steps in Known Pathways	178
4.2	Finding Uncharacterized Pathway Genes	179
4.2.1	Identification of the preQ Biosynthesis Pathway Genes . . .	179
4.2.2	Hunting for the Wyeosine Biosynthesis Genes	182
5	Conclusions	183
	References	184

Evolution and Function of Processosome, the Complex That Assembles Ribosomes in Eukaryotes: Clues from Comparative Sequence Analysis 191

A. MUSHEGIAN

1	Introduction	191
2	Sequence Analysis of the Processosome Components	192
2.1	Intrinsic Features	193
2.2	Evolutionarily Conserved Sequence Domains	195
2.2.1	Kre33p, or Possibly AtAc: Protein with Multiple Predicted Activities	204
2.2.2	Imp4/Ssf1/Rpf1/Brx1/Peter Pan Family of Proteins	209
2.2.4	Diverse RNA-Binding Domains and Limited Repertoire of Globular Protein Interaction Modules	211
3	Phyletic Patterns	212
4	Concluding Remarks	216
	References	217

Bioinformatics-Guided Experimental Characterization of Mismatch-Repair Enzymes and Their Relatives 221

P. FRIEDHOFF

1	Introduction	221
1.1	Sau3AI and Related Restriction Endonucleases	222
1.2	DNA Mismatch Repair	223
1.3	Nicking Endonuclease MutH	224
2	Sau3AI – Similar Folds for N- and C-Terminal Domains	225
2.1	Fold Recognition for the C-terminal of Sau3AI	225
2.2	Biochemical and Biophysical Analysis – Evidence for a Pseudotetramer That Induces DNA Looping	227
3	Identification of the Methylation Sensor of MutH	232
3.1	Evolutionary Trace Analysis	233
3.2	Superposition of MutH with REases in Complexes with DNA	235
3.3	Mutational Analysis of MutH	236
4	Conclusions	238
	References	239

**Predicting Functional Residues in DNA Glycosylases
by Analysis of Structure and Conservation 243**
D.O. ZHARKOV

1	Introduction	243
2	Generating Predictions: Sequence Selection and Analysis . .	244
3	Testing the Predictions: Mutational Analysis of Residues Defining Substrate Specificity in Formamidopyrimidine-DNA Glycosylase	251
4	Refining the Predictions: Analysis of Substrate Specificity in the Endonuclease III Family	254
	References	259

Subject Index 263