## 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
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
2.2 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

**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

### Modeling Protein Folding Pathways

**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 .. 104
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 HMMTR-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
Predicting Functional Residues in DNA Glycosylases
by Analysis of Structure and Conservation

D.O. Zharkov

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

References

Subject Index
Practical Bioinformatics
Bujnicki, J.M. (Ed.)
2004, XVII, 265 p., Hardcover
ISBN: 978-3-540-20613-2