

Contents

Preface	xi
1 Introduction	1
1.1 Prologue	1
1.1.1 Scope and aims	1
1.1.2 Why proteins?	2
1.1.3 Outline of the work	4
1.2 Basic principles of protein structure	5
1.2.1 The shapes and sizes of proteins	6
1.2.2 The hydrophobic core	7
1.2.3 Secondary structure	8
1.2.4 Packed layers	9
1.2.5 Barrel structures and β -helices	11
1.2.6 Protein topology	13
1.2.7 Domain structure	15
 PART 1	
GEOMETRY	17
2 Ellipsoids and embedding	19
2.1 Geometric representations of structure	19
2.1.1 Atomic coordinates	19
2.1.2 Torsion angles	22
2.1.3 Distances	26
2.1.4 Inertial axes	27
2.1.5 The shapes of proteins	32
2.2 Distance geometry	34
2.2.1 Out of hyperspace	35
2.2.2 Interpretation of the eigenvalues	40
2.2.3 Hierarchical inertial embedding	41
2.2.4 Gradual projection	45
2.2.5 Practical method specification	47

3	Sticks to strings	49
3.1	Secondary structure geometries	49
3.1.1	Secondary structure line segments	50
3.1.2	Secondary structure definition	55
3.1.3	Comparison to standard definitions	57
3.1.4	Applications and further developments	61
3.2	Simplified architectures	63
3.2.1	Stick packing	63
3.2.2	Calibrating segment packing	65
3.2.3	Layer architectures	68
3.2.4	Layer-based stick models	70
3.2.5	Polyhedra-based stick models	73
3.2.6	Packing nomenclatures	82
3.2.7	From 3D to 2D	85
3.2.8	From 2D to 1D	87
3.2.9	Uniqueness of string descriptors	87
3.2.10	Predicting helix contacts	89
4	Sheets and barrels	93
4.1	β -sheet geometry	93
4.1.1	Sheet chirality	93
4.1.2	Geometric models for a twisted sheet	94
4.1.3	The surface of a twisted sheet	99
4.1.4	Sheet bend and curl	111
4.1.5	β -sandwiches	113
4.2	β -barrels	118
4.2.1	Hyperbolic surfaces	118
4.2.2	Shear and stagger	120
4.2.3	Cylindrical β -barrels	121
4.2.4	Hyperbolic β -barrels	125
4.2.5	Optimized barrels	129
PART 2		
CLASSIFICATION		131
5	Networks and domains	133
5.1	Hydrogen bond networks	133
5.1.1	α -carbon-based β -sheet definition	134
5.1.2	β -sheet dimensions	136
5.1.3	TBC angle distribution	136
5.1.4	β -barrel identification	138
5.1.5	β -sheet classification	140
5.2	Protein domain definitions	143
5.2.1	Physical methods	143

5.2.2	An Ising model	144
5.2.3	Model evolution and domain extraction	146
5.2.4	Conforming to expectation	147
5.2.5	Setting the granularity level	149
5.2.6	Performance and examples	151
5.2.7	Simultaneous definition on multiple structures	155
5.2.8	Probabilistic definitions	156
5.2.9	Reparsing domains	157
6	Protein structure comparison	159
6.1	Overview of comparison methods	159
6.1.1	Structure representations and degrees of difficulty	160
6.1.2	The DALI method	161
6.1.3	Geometric hashing approach	162
6.1.4	Using structural superposition	164
6.1.5	The SSAP program	165
6.1.6	Iterated double dynamic programming	167
6.1.7	Secondary structure graph matching	169
6.1.8	Stick-figure comparisons	170
6.2	Assessment of significance	172
6.2.1	Score distributions from known structures	173
6.2.2	Random structural models	174
6.2.3	Randomized alignment models	175
6.2.4	Scoring and biological significance	176
6.2.5	Examples	177
7	Classification and fold spaces	180
7.1	Protein structure classification	180
7.1.1	Practical approaches to classification	180
7.1.2	Organization of the classifications	182
7.1.3	Analysis of the classifications	183
7.2	Protein fold spaces	185
7.2.1	Distance geometry projection	185
7.2.2	Simplified fold space	187
7.3	A 'periodic table' for protein structures	187
7.3.1	Classification using ideal stick forms	187
7.3.2	Structure layers become valance shells	188
7.3.3	Matching against all stick forms	191
7.3.4	Reintroducing topology	195
7.3.5	Expanding the classification tables	197
7.4	'Evolutionary' steps in fold space	198
7.4.1	Matching ideal forms	200
7.4.2	Largest common fold	200
7.4.3	Trees of structures	203
7.4.4	Links and islands in fold space	207

PART 3	
TOPOLOGY	209
8 Folds, tangles and knots	211
8.1 Topology and knots	211
8.1.1 Introduction	211
8.1.2 Chemical topology	212
8.1.3 Polymer topology	212
8.1.4 True topology of proteins	214
8.1.5 Pseudo-topology of proteins	216
8.1.6 Topology of weak links in proteins	223
8.1.7 Generalized protein knots	225
8.1.8 Knots in random chains	227
8.2 Random walks in fold space	227
8.2.1 Random walks	229
8.2.2 Secondary structure based fake proteins	230
8.2.3 Off-lattice fold combinatorics	232
8.2.4 Classifying topology in fake proteins	236
8.2.5 Local <i>versus</i> global folding	237
8.3 Protein fold complexity	238
8.3.1 Topological indices	238
8.3.2 Local and non-local packing	243
8.3.3 Smoothing folds away	248
9 Structure prediction and modelling	249
9.1 Random folds from distance geometry	249
9.1.1 Outline of the projection strategy	249
9.1.2 Model specification	250
9.1.3 Hydrogen bonds	252
9.1.4 Chirality	255
9.1.5 Scoring	255
9.1.6 Simple models	255
9.2 Modelling with distance geometry	258
9.2.1 Generic preferences	258
9.2.2 Specific interactions	259
9.2.3 Sources of real data	260
9.3 Protein tertiary structure prediction	261
9.3.1 <i>Ab initio</i> prediction	262
9.3.2 Empirical methods	263
9.3.3 Model evaluation	265
9.3.4 Genetic algorithm approach	267

PART 4	
SYMMETRY	269
10 Structural symmetry	271
10.1 Symmetry	271
10.1.1 Symmetry from domain duplication	272
10.1.2 Symmetries from secondary structure	277
10.2 A Fourier analysis of symmetry	280
10.2.1 Internal protein structure comparison	281
10.2.2 Smoothing the score matrix	282
10.2.3 Fourier transform	283
10.2.4 Visualizing repeats	285
10.2.5 Analysis of total power	286
10.2.6 Removing expected symmetries	288
10.2.7 Assessment of the Fourier approach	289
10.2.8 Origin of structural symmetry	290
11 Evolution and origins	293
11.1 Evolution of structure and function	293
11.1.1 Gene duplication and fusion	294
11.1.2 Introns and exons	295
11.1.3 Models of structure evolution	297
11.1.4 Evolution of function	300
11.1.5 Selection on random folds	301
11.2 The origins of proteins	303
11.2.1 The emergence of proteins in an RNA world	303
11.2.2 Functions for protoproteins	304
11.3 The secret of life	306
References	308
Contents	325