
Contents

List of Tables	xi
List of Figures	xiii
About the Author	xvii
1 Introduction to Glycobiology	1
1.1 Roles of carbohydrates	1
1.2 Glycan structures	2
1.3 Glycan classes	6
1.4 Glycan biosynthesis	13
1.4.1 N-linked glycans	13
1.4.2 O-linked glycans	16
1.4.3 Glycosaminoglycans (GAGs)	16
1.4.4 Glycosphingolipids (GSLs)	17
1.4.5 GPI anchors	19
1.4.6 LPS	19
1.5 Glycan motifs	20
1.6 Potential for drug discovery	22
2 Background	25
2.1 Glycan nomenclature	25
2.1.1 InChI TM	25
2.1.2 (Extended) IUPAC format	27
2.1.3 CarbBank format	30
2.1.4 KCF format	31
2.1.5 LINUCS format	32
2.1.6 BCSDb format	34
2.1.7 Linear Code [®]	37
2.1.8 GlycoCT format	40
2.1.9 XML representations	46
2.2 Lectin-glycan interactions	48
2.2.1 Families and types of lectins	50
2.2.2 Carbohydrate-binding mechanism of lectins	57
2.3 Carbohydrate-carbohydrate interactions	58

3	Databases	61
3.1	Glycan structure databases	61
3.1.1	KEGG GLYCAN	62
3.1.2	GLYCOSCIENCES.de	68
3.1.3	CFG	74
3.1.4	BCSDB	82
3.1.5	GLYCO3D	85
3.1.6	MonoSaccharideDB	86
3.1.7	GlycomeDB	89
3.2	Glyco-gene databases	90
3.2.1	KEGG BRITE	91
3.2.2	CFG	91
3.2.3	GGDB	94
3.2.4	CAZy	94
3.3	Lipid databases	96
3.3.1	SphingoMAP [®]	96
3.3.2	LipidBank	97
3.3.3	LMSD	98
3.4	Lectin databases	101
3.4.1	Lectines	101
3.4.2	Animal Lectin DB	101
3.5	Others	101
3.5.1	GlycoEpitopeDB	101
3.5.2	ECODAB	102
3.5.3	SugarBindDB	106
4	Glycome Informatics	107
4.1	Terminology and notations	107
4.2	Algorithmic techniques	108
4.2.1	Tree structure alignment	108
4.2.2	Linkage analysis using score matrices	110
4.2.3	Glycan variation map	112
4.3	Bioinformatic methods	114
4.3.1	Glycan structure prediction from glycogene microarrays	114
4.3.2	Glyco-gene sequence and structure analysis	116
4.3.3	Glyco-related pathway analysis	119
4.3.4	Mass spectral data annotation	124
4.4	Data mining techniques	130
4.4.1	Kernel methods	131
4.4.2	Frequent subtree mining	138
4.4.3	Probabilistic models	142
4.5	Glycomics tools	173
4.5.1	Visualization tools	173
4.5.2	Pathway analysis tools	177
4.5.3	PDB data analysis	178

4.5.4	3D analysis tools	179
4.5.5	Molecular dynamics	182
4.5.6	Spectroscopic tools	186
4.5.7	NMR tools	189
5	Potential Research Projects	193
5.1	Sequence and structural analyses	193
5.1.1	Glycan score matrix	194
5.1.2	Visualization	194
5.2	Databases and techniques to integrate heterogeneous data sets	195
5.3	Automated characterization of glycans from MS data	196
5.4	Prediction of glycans from data other than MS	196
5.5	Biomarker prediction	197
5.6	Systems analyses	197
5.7	Drug discovery	198
A	Sequence Analysis Methods	199
A.1	Pairwise sequence alignment (dynamic programming)	199
A.1.1	Dynamic programming	199
A.1.2	Sequence alignment	202
A.2	BLOSUM (BLOcks Substitution Matrix)	205
B	Machine Learning Methods	207
B.1	Kernel methods and SVMs	207
B.2	Hidden Markov models	211
B.2.1	The three problems of interest for HMMs	213
B.2.2	Expectation-Maximization (EM) algorithm	215
B.2.3	Hidden tree Markov models	216
B.2.4	Profile Hidden Markov models (profile HMMs)	218
C	Glycomics Technologies	221
C.1	Mass spectrometry (MS)	221
C.1.1	MALDI-MS	222
C.1.2	FT-ICR	223
C.1.3	LC-MS (HPLC)	224
C.1.4	Tandem MS	224
C.2	Nuclear magnetic resonance (NMR)	225
	References	227
	Index	241