

Contents

List of contributors	xvii		
Preface	xxi		
1. Introduction to basics of bioinformatics	1	2. Biological databases and their application	17
<i>Rajesh Kumar Pathak, Dev Bukhsh Singh and Rahul Singh</i>		<i>Parva Kumar Sharma and Inderjit Singh Yadav</i>	
1.1 Introduction	1	2.1 Introduction	17
1.1.1 Concept behind bioinformatics, in silico biology, and computational biology	3	2.1.1 Relational databases	17
1.1.2 Scientific discipline and support systems for bioinformatics	5	2.1.2 Object-oriented databases	17
1.1.3 Needs of bioinformatics	5	2.2 Sequence databases	19
1.2 Historical background of bioinformatics	5	2.2.1 GenBank	19
1.3 Aim of bioinformatics	7	2.2.2 European Nucleotide Archive	19
1.4 The recent development in the field of bioinformatics	7	2.2.3 DNA Database of Japan	20
1.5 Challenges in bioinformatics	8	2.2.4 GenPept	20
1.6 Application of bioinformatics	8	2.2.5 Protein information resources	20
1.6.1 Sequence analysis	8	2.3 Composite database	21
1.6.2 Phylogenetic analysis	8	2.3.1 Universal Protein Resource (UniProt)	21
1.6.3 Prediction of protein secondary structure	9	2.3.2 Nonredundant database	22
1.6.4 Protein 3D structure prediction	9	2.4 Secondary database	22
1.6.5 Evaluation and validation of predicted protein model	10	2.4.1 PROSITE	22
1.6.6 Discovery and designing of small molecules leading to drugs/ agrochemical development	10	2.4.2 PRINTS	22
1.6.7 Next-generation sequencing data analysis	10	2.4.3 Pfam	23
1.6.8 SNP and SSR identification	10	2.4.4 InterPro	23
1.6.9 Pharmacogenomics	10	2.5 Structural databases	23
1.6.10 Metabolomics and metabolic flux analysis	11	2.5.1 Research collaboratory for structural bioinformatics protein data bank	23
1.6.11 Systems biology and omics data integration	11	2.5.2 SCOP	24
1.7 Future perspective	11	2.5.3 CATH/Gene3D	24
1.8 Conclusion	11	2.6 Specialized database	24
Conflict of interest	12	2.6.1 Clustering databases	24
References	12	2.6.2 Bibliographic databases	26
		2.6.3 Expression databases	26
		2.6.4 Taxonomy databases	26
		2.6.5 Interaction databases	26
		2.6.6 Pathway databases	27
		2.6.7 Enzyme databases	27
		2.6.8 microRNA database	27
		2.6.9 Small molecule database	27
		2.6.10 Vaccine design database	28
		2.7 Database searching and annotation	28
		2.7.1 Entrez	28
		2.7.2 The sequence retrieval system	28
		2.7.3 Annotation	28
		2.8 Conclusions	29
		Conflict of interest	29
		References	29

3. Biological sequence analysis	33	4.3 Data preprocessing	52
<i>Samvedna Shukla, Bhawana Mishra, Himanshu Avashthi and Muktesh Chandra</i>		4.3.1 Quality control of raw sequencing data	52
3.1 Introduction	33	4.3.2 Trimming and filtering of raw reads	52
3.2 Sequence alignments: determining similarity and deducing homology	34	4.4 Genome assembly approaches: types of assembly	52
3.2.1 Why construct sequence alignment?	34	4.4.1 De novo assembly approach	52
3.2.2 Similarity of sequences	34	4.4.2 Reference-based assembly approach	54
3.2.3 Homology of sequences	35	4.4.3 Hybrid assembly approach	54
3.2.4 Global sequence alignment	35	4.4.4 Meta-assembly approach	54
3.2.5 Local sequence alignment	35	4.5 Tools and software for genome assembly	56
3.2.6 Working of alignment algorithm	36	4.5.1 Genome finishing/polishing	56
3.3 Scoring matrices: construction and proper selection	36	4.5.2 Assembly quality assessment and validation	58
3.3.1 Scoring matrices	36	4.6 Pitfall in genome assemblies	58
3.3.2 PAM matrices	37	4.6.1 DNA quality	58
3.3.3 BLOSUM matrices	37	4.6.2 Library preparation	59
3.3.4 PAM versus BLOSUM	37	4.6.3 Data quality	59
3.4 Basic Local Alignment Search Tool	38	4.6.4 Repetitive DNA	59
3.4.1 Seeding step	39	4.7 A mathematical calculation for depth/ coverage	60
3.4.2 Ungapped extension step	39	4.8 Genome annotation	60
3.4.3 Gapped extension step	39	4.8.1 Structural annotation	61
3.4.4 Types of BLAST	39	4.8.2 Functional annotation	62
3.4.5 BLAT versus BLAST	39	4.9 Application and future prospects of genome assembly	62
3.5 Multiple sequence alignment (MSA)	40	4.10 Conclusion	63
3.5.1 Need for MSA	40	Conflict of interest	63
3.5.2 Multiple sequence alignment algorithm	41	References	63
3.5.3 ClustalW	41		
3.6 Phylogenetic analysis	42	5. Computational molecular phylogeny: concepts and applications	67
3.6.1 Types of trees	43	<i>Krishna Kumar Ojha, Swapnil Mishra and Vijay Kumar Singh</i>	
3.6.2 Algorithms for phylogenetic analysis	43	5.1 Introduction	67
3.6.3 Terminology of phylogenetic tree	43	5.2 Convergent and divergent evolution	67
3.7 Application of sequence alignment	44	5.3 Concept of cladistics and systematics	68
3.8 Conclusion	44	5.4 Phylogenetic trees' terminology	69
Conflict of interest	45	5.4.1 Phylogenetic tree	69
References	45	5.4.2 Taxon	69
		5.4.3 Node	69
4. Genome assembly and annotation	49	5.4.4 Leaf	69
<i>Pallavi Mishra, Ranjeet Maurya, Himanshu Avashthi, Shikha Mittal, Muktesh Chandra and Pramod Wasudeo Ramteke</i>		5.4.5 Internal node	70
4.1 Introduction	49	5.4.6 Edge	70
4.1.1 How do you reassemble a genome after sequencing?	50	5.4.7 Topology	70
4.1.2 Assembler technology: historical landscape	50	5.4.8 Root	70
4.2 Genome assembly algorithms	51	5.4.9 Rooted tree	70
4.2.1 The overlap-layout-consensus/string graph assemblers	51	5.4.10 Unrooted tree	70
4.2.2 De Bruijn graph assemblers	51	5.4.11 Binary tree	70
		5.4.12 Clade	71
		5.4.13 Monophyletic taxon	71

5.4.14 Paraphyletic taxon	71	6.7 Applications of RNA-sequencing	100
5.4.15 Polyphyletic taxon	71	6.8 Advantages of transcriptome sequencing over microarray technology	100
5.4.16 Split	72	6.9 Limitations and future perspective of RNA-sequencing	100
5.4.17 Subtree	72	6.10 Conclusion	100
5.4.18 Edge length	72	Conflict of interest	101
5.4.19 Ultrametric tree	72	References	101
5.4.20 Cladogram	73		
5.4.21 Phylogram	73		
5.4.22 Dendrogram	73		
5.5 Evolutionary inference of phylogenetic trees	73	7. RNA-seq for revealing the function of the transcriptome	105
5.5.1 Importance of shared derived characters in phylogeny	74	<i>Kavita Goswami and Neeti Sanan-Mishra</i>	
5.6 Tree construction methods	74	7.1 Introduction	105
5.6.1 UPGMA	74	7.2 Next-generation sequencing platforms and their technologies	105
5.6.2 Neighbor-joining algorithm	76	7.2.1 Illumina	106
5.6.3 Maximum parsimony	80	7.2.2 Roche 454	106
5.6.4 Maximum likelihood phylogeny	82	7.2.3 ION torrent	107
5.6.5 Bayesian phylogeny	83	7.2.4 SoLid ABI	107
5.7 Estimating reliability of phylogenetic tree	84	7.2.5 Illumina Tru-seq SLR technology	108
5.7.1 Bootstrapping	85	7.2.6 Pacific Biosciences (PacBio) SMRT sequencing	108
5.8 Phylogenetic tools	85	7.2.7 Oxford Nanopore	108
5.9 Application of molecular phylogeny	86	7.3 Analyzing the RNA-seq data	109
5.9.1 Classification	86	7.3.1 Quality and depth of raw sequencing data	109
5.9.2 Identifying the origin of pathogens	87	7.3.2 Adapter removal	110
5.9.3 Species conservation	87	7.3.3 Level of alignment	110
5.10 Conclusion	87	7.3.4 Redundancy rate of reads	111
Conflict of interest	87	7.4 RNA-seq applications	111
References	87	7.4.1 Transcriptome assembly	111
		7.4.2 Identification of novel protein-coding genes	112
		7.4.3 Identification of other classes of RNAs	113
		7.4.4 Profiling expression patterns	115
		7.4.5 Degradome sequencing	119
		7.4.6 Variants detection and allele-specific expression	120
		7.4.7 Expression quantitative trait loci	120
		7.5 Databases and software for small RNA analysis	121
		7.5.1 miRBase	121
		7.5.2 PMRD	121
		7.5.3 Armour	121
		7.5.4 UEA sRNA workbench	121
		7.5.5 sRNAtoolbox	122
		7.5.6 sRNAbench	122
		7.5.7 miRNAconsTarget	122
		7.5.8 Massively parallel signature sequencing database	122
		7.5.9 CLC Genomics Workbench	122
6. Applications and challenges of microarray and RNA-sequencing	91		
<i>Ankita Negi, Abhimati Shukla, Akanksha Jaiswar, Jatin Shrinet and Rahul Singh Jasrotia</i>			
6.1 Introduction	91		
6.2 Evolution of microarray	92		
6.2.1 Automated arrays and cDNA cloning to microarray technology	92		
6.2.2 Principle of microarray	92		
6.2.3 List of microarray tools and their utility	94		
6.3 DNA sequencing	94		
6.3.1 First generation	94		
6.3.2 Second generation	95		
6.3.3 Third generation	96		
6.4 RNA-sequencing	96		
6.4.1 Library preparation and sequencing	96		
6.4.2 Pipeline and usage of RNA-sequencing	97		
6.5 Biological databases for data submission	99		
6.6 Applications of microarray	99		

7.6 Conclusion	122	9.3.5 DEEPred	149
Conflict of interest	123	9.3.6 SDNGO	149
References	123	9.3.7 AmiGO	149
8. Analysis of SSR and SNP markers	131	9.3.8 OBO-Edit	149
<i>Ankita Mishra, Pramod Kumar Singh, Abhishek Bhandawat, Vinay Sharma, Vikas Sharma, Pradeep Singh, Joy Roy and Himanshu Sharma</i>		9.3.9 Gene Ontology Functional Enrichment Annotation Tool	149
8.1 Introduction	131	9.4 GO enrichment analysis	149
8.2 Analysis of SSR markers	132	9.4.1 DAVID (Database for Annotation, Visualization, and Integrated Discovery)	150
8.2.1 Benefits and limitations of microsatellite markers	132	9.4.2 PANTHER (Protein ANalysis THrough Evolutionary Relationships)	150
8.2.2 SSR mining and primer designing	133	9.4.3 g:Profiler	150
8.2.3 Classification and genomic localization	133	9.4.4 clusterProfiler	151
8.2.4 Functional annotations of SSR containing sequences	133	9.4.5 Enrichr	151
8.2.5 SSR amplification and evaluation of polymorphic potential	133	9.4.6 ToppGene	151
8.2.6 Cross-transferability of SSR markers	134	9.4.7 QuickGo	152
8.2.7 Future perspective	135	9.4.8 REVIGO (Reduce & Visualize Gene Ontology)	152
8.3 Analysis of SNP markers	135	9.4.9 WEGO (Web Gene Ontology)	152
8.3.1 Approaches for sequence data generation	135	9.4.10 ShinyGO	152
8.3.2 Sample preparation	135	9.4.11 ViSEAGO	152
8.3.3 Sequencing of complex genomes	135	9.4.12 PoGo (Prediction of Gene Ontology)	152
8.3.4 Assessment of sequence quality	136	9.4.13 GOrilla (Gene Ontology enRichment anaLysis and visuaLizAtion tool)	152
8.3.5 Reads assembly and mapping to reference genome	136	9.4.14 EasyGO	152
8.3.6 Postprocessing of mapped reads	138	9.4.15 GOEAST (Gene Ontology Enrichment Analysis Software Toolkit)	153
8.3.7 Variant calling and filtration	138	9.4.16 GOAT (Gene Ontology Annotation Tool)	153
8.3.8 Functional annotation of variant	138	9.4.17 GOLEM (Gene Ontology Local Exploration Map)	153
8.4 Conclusion	140	9.4.18 GOsTo (Gene Ontology semantic similarity Tool)	153
Acknowledgments	140	9.4.19 NaviGO	153
Conflict of interest	140	9.5 Applications	153
References	140	9.6 Future prospects	154
Further reading	144	9.7 Conclusion	154
9. Gene Ontology: application and importance in functional annotation of the genomic data	145	Acknowledgment	154
<i>Reshu Saxena, Ritika Bishnoi and Deepak Singla</i>		Conflict of interest	155
9.1 Background	145	Author contributions	155
9.2 Gene Ontology–based classification	146	References	155
9.2.1 Cellular component	147	10. Metagenomics: the boon for microbial world knowledge and current challenges	159
9.2.2 Molecular function	147	<i>J.K. Choudhari, J. Choubey, M.K. Verma, T. Chatterjee and B.P. Sahariah</i>	
9.2.3 Biological process	147	10.1 Introduction: an overview of metagenomics	159
9.3 Annotation of unknown gene/genome	147		
9.3.1 Blast2GO	148		
9.3.2 IPRscan (InterProScan)	148		
9.3.3 Genome Assembly and Annotation Package	149		
9.3.4 GOnet	149		

10.2	Resources in metagenomics	161	11.5.1	Mutation studies	185
10.3	Challenges in metagenomics	161	11.5.2	Unfolding studies	185
10.4	The workflow in metagenome analysis	162	11.5.3	Binding site prediction	185
10.5	Dataset acquire and processing	162	11.5.4	Protein docking and virtual screening	185
10.6	Quality control analysis	162	11.5.5	Understanding the dynamics of protein or protein–ligand complex	186
10.6.1	Base quality score	162	11.5.6	Structure evolution analysis	186
10.6.2	Sequence quality score	162	11.6	Conclusion	186
10.6.3	Overrepresented sequences	163		Conflict of interest	186
10.6.4	Per base N content	163		References	186
10.6.5	Duplicate sequences	163		Further reading	188
10.6.6	Read length distribution	163			
10.6.7	Per base sequence content	163			
10.6.8	Guanine–cytosine content	163			
10.6.9	Overrepresented k-mers	163			
10.6.10	Quality analysis and improving software tools	163			
10.7	Genome assembly tools in metagenomics	164	12.	Structural and functional analysis of protein	189
10.8	Binning tools in metagenomics	164		<i>Neetu Singh Yadav, Pawan Kumar and Indra Singh</i>	
10.8.1	Statistical analysis	166	12.1	Protein preliminaries	189
10.9	Data storage and sharing	166	12.2	Growth of the protein structural database	189
10.10	Metagenomics analysis: a case study	166	12.3	Structural topology and fold classification scheme	190
10.11	Material, methodology, and outcome	166	12.4	D-Structure quality assessment protocol	191
10.11.1	Metagenome dataset	166	12.5	Protein 3D structure prediction	191
10.11.2	Sequencing quality analysis	166	12.5.1	Energy functions	192
10.11.3	Hits distribution of metagenome from the database sources	168	12.6	Machine learning in PSP	197
10.11.4	Hits distribution of functional group	168	12.6.1	Feature engineering and representation	197
10.11.5	Taxonomic hits distribution	170	12.6.2	Feature selection	198
10.11.6	Rarefaction curve	170	12.6.3	ML algorithms	198
10.11.7	Alpha diversity	170	12.6.4	ML models' implementation and evaluation	199
10.12	Advantages of metagenomics study	172	12.7	Conclusion	200
10.13	Limitations and future perspective	172		Conflict of interest	201
10.14	Conclusion	172		References	201
	Conflict of interest	173			
	References	173			
11.	Protein structure prediction	177	13.	Computational approaches in drug designing	207
	<i>Shikha Agnihotry, Rajesh Kumar Pathak, Dev Bukhsh Singh, Apoorv Tiwari and Imran Hussain</i>			<i>Anshul Tiwari and Sakshi Singh</i>	
11.1	Introduction	177	13.1	Introduction	207
11.2	Protein structure prediction	177	13.2	Computer-aided drug designing	207
11.3	Method of protein structure prediction	178	13.2.1	Structure-based drug design	207
11.3.1	Homology modeling (comparative modeling)	178	13.2.2	Ligand-based drug design	209
11.3.2	Threading or fold recognition	180	13.3	Computational approaches	209
11.3.3	Ab initio approach	182	13.3.1	Molecular modeling	209
11.4	Evaluation of predicted protein structure	184	13.3.2	Binding site and cavity prediction	211
11.5	Applications of structure prediction	185	13.3.3	Computational ligand designing and searching	211

13.3.4	Pharmacophore modeling	211	15.3.2	Conformational flexibility	236
13.3.5	Molecular docking	211	15.3.3	Scaffold hopping	236
13.3.6	Molecular dynamics simulation	212	15.3.4	Fragment-based drug design	236
13.3.7	Quantitative structure–activity relationship	213	15.4 Pharmacophore mapping		236
13.3.8	Lead optimization	214	15.4.1	Diverse conformation generation	236
13.4 Limitations		214	15.4.2	Generation of 3D pharmacophore	237
13.5 Recent trends in drug designing		214	15.4.3	Validation of the pharmacophore model	237
13.6 Conclusion		215	15.5 Pharmacophore classifications		238
Conflict of interest		215	15.5.1	Ligand-based pharmacophore modeling	238
References		215	15.5.2	Structure-based pharmacophore modeling	239
14. Structure-based drug designing		219	15.6 Application of pharmacophore in virtual screening and de novo design		239
<i>Shubham Pant, Shivani Verma, Rajesh Kumar Pathak and Dev Bukhsh Singh</i>			15.6.1	Virtual screening	239
14.1	Introduction	219	15.6.2	De novo pathway	240
14.2	Background of structure-based drug design	220	15.7 Advancement in exploring 3D pharmacophore principles over the above limitations		240
14.3	Process of SBDD	221	15.8 Quantitative structure–activity relationship		240
14.3.1	Target identification	221	15.8.1	Designation of QSAR	241
14.3.2	Target structure determination/prediction	221	15.8.2	Backbone of chemical similarity	241
14.3.3	Cavity/binding site prediction	223	15.8.3	QSAR data	242
14.3.4	Ligand structure preparation/retrieval	223	15.8.4	Classification of 3D QSAR approaches	243
14.3.5	Molecular docking and virtual screening	224	15.8.5	Molecular interactions and energies	243
14.3.6	ADMET analysis	226	15.8.6	QSAR modeling	245
14.3.7	Molecular dynamics simulation	226	15.8.7	Concept of applicability domain and QSAR approaches	246
14.3.8	Binding-free-energy calculation-MM-PBSA	227	15.9 Development of new QSAR: HQSAR		248
14.4 Recent development in SBDD		227	15.10 Application of QSAR/SAR		248
14.5 Challenges and limitations		228	15.10.1	Synthetic organic chemistry and QSAR	248
14.6 Future prospective		228	15.10.2	Prediction of kinetic and thermodynamic parameters	249
14.7 Conclusion		229	15.10.3	Drug development and other applications	249
Conflict of interest		229	15.11 Conclusion		250
References		229	Conflict of interest		250
15. Ligand-based drug designing		233	References		250
<i>Suchitra M. Ajjrapu, Apoorv Tiwari, Pramod Wasudeo Ramteke, Dev Bukhsh Singh and Sundip Kumar</i>			16. Discovery and optimization of lead molecules in drug designing		253
15.1	Introduction	233	<i>Shivani Verma and Rajesh Kumar Pathak</i>		
15.2	Pharmacophore	234	16.1	Introduction	253
15.2.1	Build pharmacophore hypothesis	234	16.2	Principles of CADD	254
15.2.2	Alignment of molecules	235			
15.2.3	Similarity search methods	235			
15.2.4	2D finger prints	235			
15.3 3D fingerprints		235			
15.3.1	3D similarities depend on the alignment	236			

16.2.1 Case study: inhibition of lipoxygenases	255	17.5.2 Prediction of pharmacokinetic properties	279
16.3 Discovery of the lead molecule	256	17.5.3 Target identification and de novo ligand design using pharmacophore approaches	279
16.4 Types of lead molecules	256	17.5.4 Protein functionality studies	279
16.4.1 Natural lead compounds	256	17.5.5 3D pharmacophore modeling using a web platform	279
16.4.2 Synthetic lead molecules	257	17.5.6 Pharmacophore methods in light of molecular dynamics simulations	280
16.4.3 Semisynthetic drugs	259	17.6 Applications of pharmacophore modeling	281
16.5 Lead optimization and strategies	259	17.6.1 Generation of e-pharmacophore for virtual screening of drug molecules	281
16.5.1 By using organic synthetic chemistry	259	17.6.2 ADME-Tox analysis	282
16.5.2 Structure simplification	261	17.6.3 Generation of a multitarget ligand	282
16.5.3 Structure modification	261	17.6.4 Modulation of the immune system	282
16.5.4 Functional group interconversion	261	17.6.5 Pharmacophore-guided drug target identification	282
16.5.5 Bonding strength and selectivity	262	17.7 Future perspectives of pharmacophore models	283
16.5.6 Using thermodynamic, pharmacodynamics, and pharmacokinetic parameters	262	17.7.1 Fragment-based drug design	283
16.6 Computational lead optimization	263	17.7.2 Protein–protein interaction inhibition	283
16.7 Advantages of computational lead designing	263	17.7.3 A potential role in protein design	283
16.8 Future perspectives	263	17.8 Conclusion	284
16.9 Conclusion	264	Conflict of interest	284
Conflict of interest	265	References	284
References	265	18. Molecular docking and molecular dynamics simulation	291
17. Pharmacophore modeling and its applications	269	<i>Sakshi Singh, Qanita Bani Baker and Dev Bukhsh Singh</i>	
<i>Rashmi Tyagi, Amisha Singh, Kamal Kumar Chaudhary and Manoj Kumar Yadav</i>		18.1 Introduction	291
17.1 Introduction	269	18.2 Molecular docking	292
17.2 Basics of pharmacophore modeling	271	18.2.1 Algorithms	292
17.2.1 Division of initial data into diverse datasets	271	18.2.2 Scoring functions	294
17.2.2 Conformational analysis with three-dimensional structures	272	18.2.3 Knowledge-based SFs	294
17.3 Different methods of pharmacophore generation	273	18.3 Docking methodologies	295
17.3.1 Ligand-based pharmacophore modeling	273	18.3.1 Flexible docking	295
17.3.2 Structure-based pharmacophore modeling	275	18.3.2 Semiflexible docking	295
17.4 Validation of pharmacophore models	276	18.3.3 Virtual screening of high-throughput docking	295
17.4.1 Receiver operating characteristic curve	276	18.3.4 Fragment docking	296
17.4.2 Structure-based pharmacophore modeling approach for the design of azaindole derivatives as DprE1 inhibitors for tuberculosis: a case study	277	18.3.5 Machine learning in docking	296
17.5 Recent trends in pharmacophore generation	278	18.3.6 Docking tools and their features	297
17.5.1 Machine-learning models incorporated with pharmacophore descriptors	278	18.4 Molecular dynamics simulation	297
		18.4.1 Postdocking refinement	298
		18.4.2 Binding-free energy calculations: MM-GBSA/MM-PBSA	299

18.5 Challenges in molecular docking and MD simulation techniques	300	20.3.2 Prediction of linear B-cell epitopes	321
18.6 Conclusion	301	20.4 <i>In silico</i> method for T-cell epitope prediction	322
Conflict of interest	301	20.4.1 MHC class I binder prediction	323
References	301	20.4.2 MHC class II binder prediction	324
		20.4.3 MHC gene diversity and its importance in T-cell epitope prediction	325
19. Pharmacokinetics and pharmacodynamics analysis of drug candidates	305	20.5 Adjuvant and linker selection	325
<i>Satendra Singh, Dev Bukhsh Singh, Budhayash Gautam, Anamika Singh and Namrata Yadav</i>		20.6 Building 3D model and validation of fusion vaccine construct	326
19.1 Introduction	305	20.7 Miscellaneous properties	326
19.2 Postgenomic era and drug discovery	306	20.7.1 Peptide toxicity	326
19.3 Pharmacokinetics	307	20.7.2 Half-life or stability	327
19.3.1 Drug absorption	308	20.7.3 Delivery methodology	327
19.3.2 Drug distribution (binding/localization/storage)	308	20.8 Role of next-generation sequencing technology in vaccine design	328
19.3.3 Drug metabolism	309	20.9 Computer-aided vaccine development example	328
19.4 Pharmacodynamics	311	20.10 Conclusion	329
19.4.1 Drug toxicity	311	Acknowledgments	329
19.5 Computational approaches for ADMET prediction	311	Conflict of interest	330
19.6 Translational bioinformatics	313	References	330
19.7 Drug repurposing	313		
19.7.1 Benefits of drug repurposing	313	21. Metabolomics and flux balance analysis	337
19.7.2 Computational drug repurposing	314	<i>Priyanka Narad, G. Naresh and Abhishek Sengupta</i>	
19.8 Role of pharmacogenomics in precision medicine	314	21.1 Introduction	337
19.9 Chemical diversity of natural products: a source for computer-aided drug discovery	315	21.2 Definition of metabolomics	338
19.10 Conclusion	315	21.3 MS- and NMR-based techniques in metabolomics	338
Conflict of interest	315	21.4 Data processing in metabolomics	339
References	315	21.4.1 Nuclear magnetic resonance spectroscopy	339
Further reading	316	21.4.2 Workflow of MS-based metabolomics	339
		21.4.3 Limitations of NMR and MS methods	340
20. Computational approaches for vaccine designing	317	21.4.4 Recent advances in MS- and NMR-based metabolomics	341
<i>Animesh Awasthi, Gaurav Sharma and Piyush Agrawal</i>		21.4.5 Challenges and affecting factors	341
20.1 Introduction	317	21.5 Applications of metabolomics	341
20.2 Antigen selection and immunological databases	318	21.5.1 Microbial science	341
20.2.1 Exogenous antigens	318	21.5.2 Plant science	341
20.2.2 Endogenous antigens	319	21.5.3 Animal science	342
20.2.3 Autoantigens	319	21.5.4 Medical science	342
20.3 <i>In silico</i> method for B-cell epitope prediction	319	21.5.5 Food and herbal medicines	343
20.3.1 Prediction of conformational B-cell epitopes	320	21.6 Flux balance analysis	343
		21.7 Metabolic networks and model construction	344

21.7.1	Metabolic model: construction and refinement	345	22.4.1	Graph theory	371
21.7.2	Mass balance	348	22.4.2	Adjacency matrices	373
21.7.3	Steady state	349	22.5 Topological parameters	373	
21.7.4	Types of constraints	349	22.5.1	Node degree	374
21.7.5	Optimization	350	22.5.2	The average of shortest path length	374
21.7.6	Steps in FBA	351	22.5.3	Clustering coefficient	375
21.8 Metabolic control analysis and isotopic steady state/carbon flux analysis	352		22.5.4	Betweenness centrality	375
21.8.1	Metabolic control analysis	352	22.5.5	Statistical comparison	375
21.8.2	Carbon flux analysis	352	22.6 Biological significance of network motifs	375	
21.8.3	Different types of flux balance analysis at different conditions	353	22.7 Applications of network biology	376	
21.9 Some important tools of flux balance analysis	355		22.8 Limitations and challenges	376	
21.9.1	OptKnock	355	22.9 Conclusion	376	
21.9.2	OptGene	355	Conflict of interest	377	
21.9.3	OptStrain	356	References	377	
21.9.4	COBRA Toolbox	356			
21.9.5	MetaboAnalyst 4.0	356	23. Network biology and applications	381	
21.9.6	OptFlux	356	<i>Neeru Redhu and Zoozeal Thakur</i>		
21.9.7	OpenFlux	357	23.1 Introduction to biological networks	381	
21.9.8	CellNetAnalyzer	357	23.2 Biological networks properties	381	
21.9.9	SBRT	357	23.2.1	Path, average path length, and diameter	381
21.9.10	Escher-FBA	357	23.2.2	Degree aka connectivity	382
21.9.11	MetaFluxNet	357	23.2.3	Scale free	382
21.10 Applications, challenges, and future perspectives of FBA	358		23.2.4	Small world network	382
21.10.1	Applications	358	23.2.5	Date and party hub	382
21.10.2	Challenges and future perspectives	360	23.2.6	Network motifs	382
21.11 Case study: applications of metabolomics and flux balance analysis in industrially important microorganisms	361		23.3 Types of biological networks	382	
21.11.1	Lactococcus lactis	361	23.3.1	Ecological networks	383
21.11.2	Saccharomyces cerevisiae	361	23.3.2	Gene (genetic) regulatory network	383
21.11.3	Escherichia coli	362	23.3.3	Protein–protein interaction network	384
21.12 Conclusion	363		23.3.4	Metabolic networks	385
References	363		23.3.5	Cellular signaling network	385
			23.3.6	Gene coexpression network	387
22. Topological parameters, patterns, and motifs in biological networks	367		23.4 Experimental methods in network biology	387	
<i>Arvind Kumar Yadav, Rohit Shukla and Tiratha Raj Singh</i>			23.4.1	Microarray	387
22.1 Introduction	367		23.4.2	Deep mRNA sequencing	388
22.2 Biological networks	368		23.4.3	Exome sequencing	388
22.2.1	Construction of biological networks	368	23.4.4	ChIP-seq	389
22.3 Network motifs and patterns	369		23.4.5	Genome-wide bisulfite sequencing	389
22.3.1	Motif discovery and counting	369	23.4.6	Yeast two-hybrid	390
22.4 Analysis of biological network	370		23.4.7	Mass spectrometry-based proteomics	391
			23.4.8	Flow and mass cytometry	392
			23.4.9	Live cell imaging	392
			23.5 Resources for biological network-based studies	393	

25.6.1	Network-based approaches for systems pharmacology	435	27.3.2	Estimation	458
25.7	Case study related to transcriptome data analysis	435	27.3.3	Prediction	459
25.8	Limitations in big data analysis	437	27.3.4	Association	459
25.9	Conclusion	438	27.3.5	Clustering	459
	Acknowledgment	438	27.3.6	Description and visualization	460
	Conflict of interest	438	27.3.7	Case studies using Waikato environment for knowledge analysis	461
	References	438	27.4	Feature selection technique in data mining	461
26. Machine learning in bioinformatics		443	27.4.1	Objective of feature selection	462
	<i>Indrajeet Kumar, Surya Pratap Singh and Shivam</i>		27.5	Major data mining algorithms applicable to biological data	462
26.1	Introduction	443	27.5.1	C4.5 algorithm	462
26.2	Supervised learning	443	27.5.2	K-means algorithm	462
26.2.1	Classification	444	27.5.3	Support vector machine	463
26.2.2	Supervised machine learning in bioinformatics	444	27.6	Biological data evolution and related issues	463
26.3	Unsupervised machine learning	445	27.6.1	Biological data availability	463
26.4	Problems to understand supervised learning and unsupervised learning	446	27.6.2	Biological data availability in computer-readable form	464
26.5	Regression	446	27.6.3	Biological data cleaning	464
26.5.1	Hypothesis	447	27.6.4	Biological data quality	464
26.6	Clustering	450	27.6.5	Biological data dimensionality	464
26.6.1	K-means clustering	450	27.6.6	Biological data knowledge discovery	466
26.6.2	Density-based clustering	450	27.7	Bioinformatics research areas and tools	466
26.6.3	Distribution-based clustering	451	27.7.1	Sequence searching, comparison, and evolutionary analysis	467
26.6.4	Fuzzy clustering	451	27.7.2	Annotation of gene/protein structure and function	467
26.7	Unsupervised learning in bioinformatics	452	27.7.3	Gene and protein expression analysis	468
26.8	Application of machine learning	452	27.7.4	Mutation and disease association study	468
26.8.1	Genome annotation	452	27.7.5	Protein structure prediction, docking, and protein–protein interaction analysis	468
26.8.2	Protein structure prediction	452	27.7.6	Biological systems modeling and network analysis	468
26.8.3	Research area in bioinformatics with deep learning	453	27.7.7	Expressed sequence tag analysis	469
26.9	Discussion	453	27.7.8	MicroRNA and target prediction	469
26.10	Conclusion	454	27.7.9	Medical and health data analysis and clinical decision support system	469
	Conflict of interest	454	27.8	Limitations	469
	References	454	27.9	Conclusion	470
27. Bioinformatics and biological data mining		457		Conflict of interest	470
	<i>Aditya Harbola, Deepti Negi, Mahesh Manchanda and Rajesh Kumar Kesharwani</i>			References	470
27.1	Biological data mining	457			
27.2	Data mining applications	457			
27.3	Data mining process	458			
27.3.1	Classification	458	Index		473