Index

A
ab initio methods, protein folding, 261, 288-289
chaperonins, 294-295
energy landscape, 289-290
membrane protein folding, 295-296
reducing problem’s complexity, 291-294
ADME (absorption-distribution-metabolism-excretion) predictions, 298
ADs (activation domains), 368
transcriptional, 369
aerospace industry, in silico (experimentation by computer simulation), 2
Affymetrix, Inc., transcriptome mapping, 188
alpha (α) helix structures, 269-272
American College of Radiology, DICOM standard, 315
amino acids
alpha (α) helix structures, 269-272
beta (β) sheet structures, 269, 272-273
disulfide (-S-S-) bridges, 274
grouped by chemical properties, 265-266
overfitting, 154
peptide bonds, 266, 268
pseudocounts, 154-156
Ramachandran plot, 268-269
structure, 264-265
AND operator, sequence-analysis problems, 163-170
Anfinsen, Christian, protein folding view, 260-261
ANSI (American National Standards Institute), SQL standard, 34
APIs (application programming interfaces), vendor-specific databases, 52
apoptosis protein (XIAP), 254-255
application management tools, 350
Artemis program, 138
Atlas of Protein Sequence and Structure, 90-91
autocatalytic RNA molecules, 308
automotive industry, in silico (experimentation by computer simulation), 2

B
Basic Local Alignment and Search Tool (BLAST), 31-35, 78
beta (β) sheet structures, 269, 272-273
BIND (Biomolecular Interaction Network Database), 371
bioinformatic problems. See floating point bioinformatic problems; integer bioinformatic problems
Bioinformatic Sequence Markup Language (BSML) interchange standard, 74-79
bioinformatics
algorithms, 81
clustered computing environments, 4
computational biology, 3
database infrastructure. See database infrastructure for bioinformatics
EBI (European Bioinformatics Institute), 28-29
NCBI (National Center for Biotechnology Information), 28
new infrastructures, 16-18
parallel computing, 348, 350-354
PDB (Protein Data Bank), 29-30
public infrastructure, history, 6-7
INDEX

sequence analysis. See sequence analysis technologies aiding migration from physical to in silico biology, 79
technologies critical to emergence, 24-27
workflow computing, 355-359
workflow-based clustered environment, 349-350, 354
BIOML (Biopolymer Markup Language), 78
Biomolecular Interaction Network Database (BIND), 371
BLAST (Basic Local Alignment and Search Tool), 31-35, 78
blastall program, 356-358
BLOSUM (BLOcks SUbstitution Matrix), 91
Boeing 777 aircraft, in silico design, 3
Brenner, Sydney, 107
cellular biology, 124
cen, 124
centromere, heterochromatin, 124
CGAP (Cancer Genome Anatomy Project), 28
chaperonins, 294-295
chromosomal DNA, 122-124
euchromatin morphologies, 124-125
heterochromatin inheritance, 127, 130-132
heterochromatin morphologies, 124-125
nucleosomes, 126-129
polymerase, 127, 129
telomerase functions, 124-126
chromosomes, probes in chromosomes 21 and 22, 188-189
Class, Architecture, Topology, and Homology database. See CATH database
clinical trials, criteria for success, 341
cluster analysis, 357-358
clustered computing environments, 4-5
code, genotype versus phenotype, 1
coding regions of human genome, 8
computational biology, 3
Computational Biology Department (University of Southern California), PROCRUSTES algorithm, 81
computer grids, 16
application layer, 328-330
collective layer protocols, 327-328
connectivity layer, 326
design characteristics, 325-326
fabric layer, 326
implementation techniques, 330-332
medical informatics, 323-325
resource layer, 327
computer-aided three-dimensional interactive application (CATIA), 3
Crick, Francis, discovery of the double-helix structure of DNA, 107
Danish Centre for Human Genome Research database, 371
data grids, 16-17
data management tools, 350
data type definitions (DTDs), 75
data warehousing, 18, 54
versus federated databases, 52-53
data-management framework, 308
advancement phases, 313-314
clinical medicine trends, 312-313
electronic medical records, 314-323
high-throughput observation, 309-311
hypothesis-driven scientific research, 309-311
technologies critical to emergence of bioinformatics, 25
database infrastructure for bioinformatics, 30-31
architecture, 61-64
BSML interchange standard, 74-79
data warehousing, 53-54
federated databases, 52-53
GenBank. See GenBank sequence database
indexed systems, 71
link-based systems, 68, 70-71
nucleotide sequence databases, 31-34
ontologies, 64-69
protein sequence databases, 32
queries, 56-62
RefSeq, 36
schema and query mismatches, 52, 55-56
vendor-specific databases, 52
XML databases, 71-74
Database of Interacting Proteins (DIP), 371
database-driven approaches, protein folding, 261-262
Dayhoff, Margaret, 90-91
protein structure data bases, 276
DBDs (DNA-binding domains), DBD-bait or -prey fusions, 368-369
DDBJ (DNA Database of Japan), 28
de novo approaches, drug discovery, 262
Department of Energy, PDB (Protein Data Bank), 29
Department of Health and Human Services, SNOMED, 315
Dicer enzyme, 120-123
DICOM (Digital Imaging and Communications in Medicine) standard, 315
dihedral-angle calculations, 283-284
dinucleotides, 109
DION, 63
DIP (Database of Interacting Proteins), 371
Dirichlet densities and mixtures, 156
DISCO, 63
DiscoveryLink (IBM) server, 61-64
disease, molecular, 2
disease prediction
clinical trial success, 341
polygenic diseases versus monogenic diseases, 339
presymptomatic testing, 333-338
susceptibility gene searches, 339-341
diseases
polygenic, 11-15
presymptomatic testing, 18-20
disulfide (-S-S-) bridges, 274
DNA
chromosomal DNA. See chromosomal DNA discovery and development, 107-108
double-helix structure, 107
genetic code. See genetic code
DNA Database of Japan (DDBJ), 28
DNA hybridization arrays, 194. See also microarrays
DNA-binding domains (DBDs), DBD-bait or -prey fusions, 368-369
document type definitions (DTDs), 73-74
dot-product angle calculations, 283-284
double-barreled shotgun sequencing, 143
Drosophila map, 370
Drug Information System, National Cancer Institute (DIS NCI), libraries of two- and three-dimensional structures, 298
drug-discovery process, 262
DTDs (data type definitions), 75
DTDs (document type definitions), 73-74
E
E. coli, Shine/Dalgarno sequence, 162
EBI (European Bioinformatics Institute), 7, 28-29
electronic medical records, 314-315
grid computing. See grid computing limitations, 316-323
standards, 315-317
EMBL (European Molecular Biology Laboratory), 28, 138
Emory University Medical School, 189
Entrez query language, NCI, 34-36
enzymes, Dicer, 120-123
epigenetic effects
gene-expression patterns, 120
inheritance, 127, 130-132
ePOCHs, 157
ERF (eukaryotic release factor), 242
ESTs (expressed sequence tags), 194-195
euchromatin morphology, 124-125
eukaryotic genetics, focus shift to prokaryotic genetics, 26-27
eukaryotic translation, 246-247
apoptosis process, 254-255
basics, 239-242
human genes containing secondary open reading frames gene, 247
human thrombopoietin (TPO) gene, 251-252
leaky scanning, 247, 249
reinitiation, 250-251
translation-related diseases, 251, 253
European Bioinformatics Institute (EBI), 7, 28-29
European Molecular Biology Laboratory (EMBL), 28
expressed sequence tags (ESTs), 194-195
expression profiling, 189-192
F
FASTA program, 138
federated systems, 18
versus data warehousing, 52-53
fibrous proteins, 275
floating point bioinformatics problems, 4-5, 351
fold classifications. proteins, 275
Franklyn, Rosalyn, 107
FSHIFT search algorithms, 36
INDEX

G
GenBank sequence database, 7, 28, 34-36, 138
components, 38
accession number, 42-45
Base count field, 50
Definition field, 41
Direct Submission, 46
Features field, 46-50
Locus field, 40-41
Origin field, 51-52
divisions, 37
sequence size limits, 37
gene chips. See DNA hybridization arrays
Gene Map of the Human Genome, 28
Gene Recognition and Assembly Internet Link (GRAIL), 140
gene-expression process, 8
molecular medicine, 12
patterns, epigenetic effects, 120
GeneMark program, 140
Generalized Markup Language (GML), 73
genetic code, 108-109
features, 109-111
solution with synthetic polynucleotides, 110
standard genetic code, 111-113
translation systems, 117
genotype, 1
GenScan (MIT), 138-140
GGF (Global Grid Forum), 330
Gilbert, Walter, nucleotide sequencing, 26
Glimmer and GlimmerM, 140-141
global alignment tools, 93-100
versus local alignment, 86-88
Global Grid Forum (GGF), 330
globular proteins, 275-276
Globus Toolkit, 330
glycine, analysis of phi/psi angles and steric hindrance, 267
GML (Generalized Markup Language), 73
Goldfarb, Charles, SGML origins, 72
GRAIL (Gene Recognition and Assembly Internet Link), 140
GRATH (graphical representation of CATH structures) program, 281-282, 285
grid computing
application layer, 328-330
collective layer protocols, 327-328
connectivity layer, 326
design characteristics, 325-326
drug discovery, 331-332
fabric layer, 326
implementation techniques, 330
medical informatics, 323-325
resource layer, 327
GRID database, 371
H
hairpin loops, 272
haplotypes, 170-172
HapMap XML, 78
health, molecular, 2
Health information Portability and Accountability Act (HIPAA), 321
Health Level-7 (HL-7) standard, 315
Henikoff, Steve, BLOSUM, 91
Hermes, 63
heterochromatin inheritance, 127, 130-132
heterochromatin morphology, 124-125
hidden Markov models. See HMMs
hierarchical clustering, 200-202
advantages, 202-204
disadvantages, 204-207
gene expression-level information, 207-208
gene-clustering analysis of ten sequences, 204-206
gene-clustering experiments, 204-205
two-dimensional dendrogram, 203
high-throughput genome sequencing, 141-145
high-throughput observation, 309-311
HIPAA (Health information Portability and Accountability Act), 321
hmmbuild program, 357-358
hmmcalibrate program, 357-358
HMMER, 151
HMMs (hidden Markov models)
definition, 150
eexample, 152-154
HMMER, 151
overfitting, 154
profile HMM, 151
pseudocounts of amino acids, 154-156
SAM (Sequence Alignment and Modeling Software System), 151
sequence alignment, 150
hmmssearch program, 357-358
HTML (Hypertext Markup Language), 71-74
human genome, 4
chromosomal DNA. See chromosomal DNA coding regions, 8
computational techniques for identification, 132-136
codon distribution, 138
content-based approach, 136-138
database comparison approach, 137, 140
feature analysis approach, 137-138
identifying repeating sequence patterns, 140-141
high-throughput genome sequencing, 141-145
layers of complexity, 7-10
regulatory RNAmolecules, 119-123
structure, 118
human thrombopoietin (TPO) gene, 251-252
Hypertext Markup Language (HTML), 71-74
hypothesis-driven scientific research, 309-311
reasons for complexity, 310-311
IBM DiscoveryLink server, 61-64
IBM, expression profiling, 189
imaging data analysis, 16
IMMs (interpolated Markov models), 140-141
in silico biology, technologies aiding migration to
physical biology, 79
in silico modeling
in industry, 1-2
protein-small molecule interactions, 262
in vitro (test tube experimentation), 1
in vivo (observations of real life), 1
InCellico, Inc., CELL platform, 67-68
Incyte Genomics (Proteome BioKnowledge Library) database, 371
information technology. See IT
information-based medicine, 3, 308
advancement phases, 313-314
clinical medicine trends, 312-313
electronic medical records. See electronic medical records
high-throughput observation, 309-311
hypothesis-driven scientific research, 309-311
new infrastructures, 16-18
proactive predictability, 18-20
Institute for Systems Biology, expression profiling, 191
integer bioinformatics problems, 4-5, 351
International Nucleotide Sequence Databases Collaboration, 28
interpolated Markov models (IMMs), 140-141
IT (information technology)
mRNA expression profiling, 13
technologies critical to emergence of bioinformatics, 25-26
J-K
JDBC (Java Database Connect), 61
k-means clustering, 207-209
Kendrew, John, myoglobin structure, 263
Kent, James, WABA (wobble-aware block alignment algorithm), 83
Khorana, Har Gobind, synthetic polynucleotides in solution to genetic code, 109
leaky scanning, 242, 247-249
Leder, Philip, GUU codes for valine and UUG codes for leucine, 108
linear sequence-analysis problems, XOR operator solutions, 164-165
linkage disequilibrium, 170-171
living, definition, 105
Load Sharing Facility (LSF), Platform, 349
local alignment tools, 88-92
versus global alignment, 86-88
loop structures, 272
LSF (Load Sharing Facility), Platform, 349
Lynx Therapeutics, transcriptional profiling, 220
MAGE XML (Microarray / Gene Expression Markup Language), 78
MAML (Microarray Markup Language), 78
Markov models, programs, 357
Massively Parallel Signature Sequencing (MPSS), 220, 223, 226-227, 230
Megaclone-MPSS strategy, 230-234
MAXSEGS alignment program, 36
medical informatics, grid computing, 323-325
application layer, 328-330
collective layer protocols, 327-328
connectivity layer, 326
design characteristics, 325-326
drug discovery, 331-332
fabric layer, 326
implementation techniques, 330
resource layer, 327
medical records. See electronic medical records
medicine history, 14
new infrastructures, 16-18
one-protein-one-disease model, 11
Megaclone, 221-225
Megaclone-MPSS strategy, 230-234
membrane protein folding, 295-296
membrane proteins, 275
messenger molecules, 107-108
metabolic pathways, delineation, 361-367
methylation process, 130-132
INDEX

Microarray / Gene Expression Markup Language (MAGE XML), 78
Microarray Markup Language (MAML), 78
microarrays, 196, 199-200. See also DNA hybridization arrays
pharmaceutical experiment steps, 195-197
SAGE (Serial Analysis of Gene Expression), 218-219
scatterplot, 201
MIPS (Munich Information Center for Protein Sequences) database, 371
miRNAs (microRNAs), 119-122
MM (mismatch) probes, 198
MMDB (Molecular Modeling Database), 28
molecular biology
central dogma, 107-108
integer bioinformatic problems, 4
public infrastructure, history, 6-7
molecular docking, 262
molecular dynamics (MD) computations, 292
molecular genetics, one-gene-one-protein model, 11
molecular medicine, 2
gene-expression process, 12
Molecular Modeling Database (MMDB), 28
molecules, messenger molecules, 107-108
monogenic diseases, versus polygenic diseases, 339
MPSS (Massively Parallel Signature Sequencing), 220, 223, 226-227, 230
Megaclone-MPSS strategy, 230-234
mRNA expression profiling
information technology, 12
mRNA species, size classes, 15
Munich Information Center for Protein Sequences (MIPS) database, 371
myoglobin structure, 263

N
N-H bond, 268
National Biomedical Research Foundation (NBRF), PIR (Protein Information Resource), 276
National Cancer Institute. See NCI
National Center for Biotechnology Information. See NCBI
National Electrical Manufacturers Association DICOM standard, 315
National Human Genome Research Institute (NHGRI), 74
National Institute of General Medical Sciences, PDB (Protein Data Bank), 29
National Institute of Standards and Technology (NIST), HIPAA compliance, 321
National Institutes of Health (NIH), 7, 28
National Library of Medicine, PDB (Protein Data Bank), 29
National Library of Medicine (NLM), 7, 28
National Science Foundation, PDB (Protein Data Bank), 29
NBRF (National Biomedical Research Foundation, PIR (Protein Information Resource), 276
NCBI (National Center for Biotechnology Information), 7, 27-28
BLAST (Basic Local Alignment and Search Tool), 31-33
blastall program, 356
clustalw program, 357
Entrez query language, 34-36
GenBank DNA sequence database, 140
NCI (National Cancer Institute)
CGAP (Cancer Genome Anatomy Project), 28
DIS (Drug Information System), libraries of two- and three-dimensional structures, 298
transcriptome mapping, 188
Needleman-Wunsch algorithm, 94-100
Needleman-Wunsch alignments, 281
neural networks, perceptron algorithm, 156-170
NHGRI (National Human Genome Research Institute), 74
NIH (National Institutes of Health), 7, 28
Nirenberg, Marshall
GUU codes for valine and UUG codes for leucine, 108
synthetic polynucleotides in solution to genetic code, 109
NIST (National Institute of Standards and Technology), HIPAA compliance, 321
NLM (National Library of Medicine), 7 nonlinear sequence-analysis problems, XOR (exclusive OR) operator solution, 164-165
nucleosomes, 126-129
nucleotide sequence databases, record structure, 31-34
GenBank, 34-52
RefSeq, 36
NuTec Sciences, expression profiling, 189
NWGAP alignment program, 37

Oak Ridge National Laboratory, GRAIL (Gene Recognition and Assembly Internet Link), 140
Ochoa, Severo, synthetic polynucleotides in solution to genetic code, 109
ODBC (Open Database Connect), 61
OGSA (Open Grid Services Architecture), 330

384
oligonucleotide arrays, 198
omega (Ω) loops, 272
OMIM (Online Mendelian Inheritance in Man) database, 28
one-gene-one-protein model of molecular genetics, 11
one-protein-one-disease model of medicine, 11
ontologies, 18, 64-69
Open Database Connect (ODBC), 61
Open Grid Services Architecture (OGSA), 330
organ systems modeling, 372-373
overfitting, 154
oxidative phosphorylation, 364-365

P
PACS (Picture Archiving and Communications Systems), 315
PAM matrices, 90-92
parallel applications, 351-353
parallel computing, 350, 354
parallel applications, 351-353
symmetric multiprocessing systems, replacement with parallel computing, 348, 351
parallel computing environments. See clustered computing environments
pattern dictionaries, 178
pattern-discovery problems
clustered computing environments, 5
weighting functions, 162
PBS (Portable Batch System), 349
PCR (polymerase chain reaction) primer sequence, 215-216
PDB (Protein Data Bank), 7, 29-30, 277
Pegasus, 63
peptide bonds, 266, 268
perceptron algorithm, 156-162
misalignments, 161
XOR (exclusive OR) operator solution, 163-170
perfect-match (PM) probes, 198
personalized medicine, 10-11
Perutz, Max
hemoglobin structure, 276
myoglobin structure, 263
petroleum industry, in silico (experimentation by computer simulation), 2
phenotypes, 1
phi (ϕ) angle, Ca-N bond, 266-268
physical biology, technologies aiding migration to in silico biology, 79
Picture Archiving and Communications Systems (PACS), 315
PIR (Protein Information Resource), 7, 276
Pittsburgh Supercomputing Center, 36
PM (perfect-match) probes, 198
polygenic diseases, 11-15
versus monogenic diseases, 339
polymerase chain reaction (PCR) primer sequence, 215-216
polymerase DNA sequences, 127, 129
Portable Batch System (PBS), 349
presymptomatic testing, 18-20, 333-334
for metabolic diseases, 334-335
for viral and bacterial infection, 334
implementation strategies, 335-338
PRF (Protein Research Foundation), 7
primary structure, proteins, 263-269
prions, 308
proactive predictability, 18
PROCRUSTES algorithm, 81-82, 140
prokaryotic genetics, focus shift to eukaryotic genetics, 26-27
proline, 264
analysis of phi/psi angles and steric hindrance, 267
Protein Data Bank (PDB), 7, 29-30, 277
protein folding, 5, 275
ab initio methods. See ab initio methods
database-driven approaches, 261-262
Protein Information Resource (PIR), 7, 276
Protein Research Foundation (PRF), 7
protein sequence databases, record structure, 32
protein structure
basics, 263
lead predictions, 296-299
primary structure, 263-269
quaternary structure, 263, 276
secondary structure, 263, 269-273
tertiary structure, 263, 273-275
protein structure databases, 276-277
CASP (Critical Assessment of Protein Structure Prediction), 289
CATH, 277-285
SCOP, 285-288
protein structure prediction, 262
protein-interaction maps, 368
databases, 371
Drosophila map, 370
organ systems modeling, 372-373
time slices, 370
protein-translation process
characteristics, 117
eukaryotic mechanism, 239-242
eukaryotic translation. See eukaryotic translation fidelity, 244-245
protein-protein and protein-nucleic acid interactions, 245
ribosomes. See ribosomes
sequence-specific enzymatic reaction, 245
tRNA-amino acid control, 246
Proteome BioKnowledge Library (Incyte Genomics) database, 371
proteomics, 367
pseudocounts of amino acids, 154-156
public infrastructure, bioinformatics history, 6-7
Q-R
QM (quantum mechanical model), 291-292
quaternary structure, proteins, 263, 276
R groups, 264
Ramachandran plot, phi/psi angles, 267-269
random coils, 272
RefSeq database, 36
regulatory RNA molecules, 119-123
reinitiation, 244, 250-251
relational databases, technologies critical to emergence of bioinformatics, 25
resource management tools, 349-350
response profiles for circulating proteins, 337
retroviruses, 108
reverse transcriptases, 108
reverse turns, 272
ribosomes
definition, 239
eukaryotic translation. See eukaryotic translation key functions, 238
protein synthesis, 243
protein-translation fidelity, 244-246
rRNA complexities, 242-244
RISC (RNA-induced silencing complex), 119-121
RNA (mRNA) messenger molecules, 108
RNA interference (RNAi), 119
RNA molecules, 119-123
RNA-induced silencing complex (RISC), 119-121
RNAi (RNA interference), 119
rRNA ribosomes, 242-244
S
SAGE (Serial Analysis of Gene Expression), 214-216
generating tags, 214-216
in presence of active form of p53
in Rat Fibroblast, 221
oncogene in human epithelial cancers cells, 217, 222
microarray analyses, 218-219
splice-variant problem, 232
tags displaying redundancy, 222
SAM (Sequence Alignment and Modeling Software System), 151
Sanger Center, Artemis program, 138
Sanger, Frederick
nucleotide sequencing, 26
protein structure, 264
SCOP (Structural Classification of Proteins) database, 285-287
domains, 287-288
family level, 286
fold level, 286
superfamily level, 286
secondary structure, proteins, 263, 269-270
alpha (α) helix structures, 270-272
beta (β) sheet structures, 272-273
seed pairs, 356
SELDI (surface-enhanced laser desorption ionization) spectroscopy, 337
self-organizing maps (SOM) analysis, 209-212
eight-node analysis, 211
sequence alignment
HMMs (hidden Markov models). See HMMs
perceptron algorithm, 156-170
SNPs (single nucleotide polymorphisms), 170-178
Sequence Alignment and Modeling Software System (SAM), 151
sequence analysis, 81
global alignment tools, 93-100
global alignment versus local alignment, 86-88
local alignment tools, 88-92
sequence aberrations, 85-86
two-dimensional dot-matrix comparison, 82-84, 87
sequence classification problems, 356-358
sequence databases
architecture, 61-64
BSML interchange standard, 74-79
CGAP (Cancer Genome Anatomy Project), 28
data warehousing, 53-54
DDBJ (DNA Database of Japan), 28
EBI (European Bioinformatics Institute), 28-29
EMBL (European Molecular Biology Laboratory), 28
federated databases, 52-53
GenBank. See GenBank sequence database
Gene Map of the Human Genome, 28
indexed systems, 71
International Nucleotide Sequence Databases Collaboration, 28
link-based systems, 68-71
INDEX

MMDB (Molecular Modeling Database), 28
NCBI (National Center for Biotechnology Information), 28
OMIM (Online Mendelian Inheritance in Man) database, 28
ontologies, 64-69
PDB (Protein Data Bank), 29
queries, 56-62
RefSeq, 36
schema and query mismatches, 52, 55-56
structure, 31-34
Taxonomy Browser, 28
UniGene (Unique Human Gene Sequence Collection), 28
vendor-specific databases, 52
XML databases, 71-74
Sequence Retrieval system (SRS), 71
sequence-analysis problems
SNPs (single nucleotide polymorphisms), 173-178
genetic code extensions, 173
haplotype identification, 170-172
XOR (exclusive or) operator solution, 163-170
sequence-encoding process, 157
sequence-homology problems, clustered computing environments, 5
sequential structure alignment program (SSAP), 278-285
Serial Analysis of Gene Expression. See SAGE
SGML (Standard Generalized Markup Language), 72-73
Shine-Dalgarno sequence, 162, 244
shotgun sequencing, 141-145
single nucleotide polymorphisms. See SNPs
siRNAs (small interfering RNAs), 119-122
small temporal RNAs (stRNAs), 119, 122-123
SMP (symmetric multiprocessor) systems, parallel computing, 348, 351
SNNs (supervised neural networks) systems, 211
eight-node analysis, 211
splice-variant problem, 232-234
technologies, 192-193
transcriptional profiling applications, 193-194
ESTs (expressed sequence tags), 194-195
hierarchical clustering. See hierarchical clustering
k-means clustering, 207-209
Megaclone, 221-223
Megaclone-MPSS strategy, 230-232
microarrays, 195-201
MPSS (Massively Parallel Signature Sequencing), 220, 223, 226-227, 230
oligonucleotide arrays, 198
SAGE (Serial Analysis of Gene Expression). See SAGE
SNNs (supervised neural networks), 212-213
SOM (self-organizing maps) analysis, 209-212
SQL (Structured Query Language) databases, 34
SRS (Sequence Retrieval system), 71
SSAP (sequential structure alignment program), 278-281, 285
Standard Generalized Markup Language (SGML), 72-73
standard genetic code, 111-113
STRING database, 371
stRNAs (small temporal RNAs), 119-123
Structural Classification of Proteins database. See SCOP database
Structured Query Language (SQL) databases, 34
Sun Grid Engine, 349
supervised neural networks (SNNs), 212-213
surface-enhanced laser desorption ionization (SELDI) spectroscopy, 337
susceptibility gene searches, 339-341
Swiss-PROT protein information database, 7, 28-29
tertiary structure, proteins, 263, 273-275
thrombocythaemia, 251
TPO (thrombopoietin) gene, 251-252
transcriptional profiling applications, 193-194
ESTs (expressed sequence tags), 194-195
hierarchical clustering. See hierarchical clustering
k-means clustering, 207-209
Megaclone, 221-223
Megaclone-MPSS strategy, 230-232
microarrays, 195-201
MPSS (Massively Parallel Signature Sequencing), 220, 223, 226-227, 230
oligonucleotide arrays, 198
SAGE (Serial Analysis of Gene Expression). See SAGE
SNNs (supervised neural networks), 212-213
SOM (self-organizing maps) analysis, 209-212
eight-node analysis, 211
splice-variant problem, 232-234
technologies, 192-193
transcriptome basics, 186-187
expression profiling, 189, 191-192
mapping, 188-189
INDEX

translation process
- characteristics, 117
- eukaryotic mechanism, 239-242
- eukaryotic translation. See eukaryotic translation
- fidelity, 244-245
- protein-protein and protein-nucleic acid interactions, 245
- ribosomes. See ribosomes
- sequence-specific enzymatic reaction, 245
- tRNA-amino acid control, 246
- Transparent Access to Multiple Bioinformatics Information Sources (TAMBIS), 66
- trinucleotides, 109
- tRNA-amino acid control, 246
- TSIMMIS, 63
- TwinScan, 138, 140
- two-dimensional dot-matrix comparison, 82-84, 87
- two-hybrid protein-interaction, 368-369

U
- UniGene (Unique Human Gene Sequence Collection), 28
- University of Manchester, TAMBIS (Transparent Access to Multiple Bioinformatics Information Sources), 66

V-W
- vendor-specific databases, 52
- VOs (virtual organizations), 325-326
- W3C (World Wide Web Consortium) standards, 71
- WABA (wobble-aware block alignment algorithm), 83
- Watson, James, discovery of DNA double-helix structure, 107
- weather forecasting, database-driven approaches, protein folding, 262
- weighting functions
  - pattern discovery, 162
  - perceptron algorithm, 156-170
- whole genome shotgun sequencing, 4
- Wilkins, Maurice, 107
- Winship Cancer Center, expression profiling, 189
- wobble-aware block alignment algorithm (WABA), 83
- workflow computing, 355-359
- workflow-based clustered environments, 349-350, 354
  - application management tools, 349-350
  - data management tools, 349-350
  - resource management tools, 349-350
- World Wide Web Consortium (W3C) standards, 71

X-Z
- XIAP (apoptosis protein), 254
- XML (eXtensible Markup Language) databases, 32, 71-74
- XOR (exclusive or) operator
  - sequence-analysis problems, 163-170
  - truth table, 169
- Zahler, Alan, WABA (wobble-aware block alignment algorithm), 83