Index

3D-Jury 147

ABIpro-h 146
ambiguous interaction restraints (AIRs) 217, 225
angle bending 119
artificial neural networks (ANNs) 44, 51, 204

Bayesian statistics 122–4

BLAST
de novo protein modeling 129, 135
enzymes 192
fold recognition 91–6
protein gene function 64–5, 69, 70, 72–8
protein sequence analysis 14–16, 17–18, 22, 29–30
protein structure prediction 44
protein–ligand interactions 164
BLOSUM matrix 13–14
Boltzmann-inversion formula 122–4
bond stretching 119
Born-Oppenheimer approximation 118–19

CAPRI see Critical Assessment of Predicted Interactions

CASP7 see Critical Assessment of Structure Prediction
Catalytic Site Atlas (CSA) 192, 197, 200
CD see circular dichroism
Circle-QA 146–7, 155
circular dichroism (CD) 47
CLANS searches 18
CLUSS 19
cluster analysis see graph clustering; sequence clustering
coiled coil structures 49–50
comparative genomics 74–5
comparative modeling 88–110
accuracy and errors 106–10
de novo protein modeling 90, 95, 104
evolutionary principles 88–9
fold recognition 89–99
fragment assembly 95
loop modeling 90, 103–5
meta-predictors 94–5, 97
model building 90, 100–2
model quality assessment 90, 98–9, 106
model refinement 90, 102–3
side-chain modeling 90, 104, 106–7
stepwise procedure 89–90
target-template alignment 88–90, 99–100
template identification 90, 97–9
comparative performance 252–3
COMPASS 15–16
computational docking 264
conformational sampling 124–7
consensus analysis 94, 147, 148–9, 154
continuous-space reduced models 127–30
controlled vocabularies 66–8
convergent evolution 190, 191–2
Critical Assessment of Predicted Interactions (CAPRI) 211–12, 216–21, 226
Critical Assessment of Structure Prediction (CASP7) 143–4, 149–54, 272–3
cross docking 213
cryo-electron microscopy 271
CSA see Catalytic Site Atlas
cut-based methods 237, 238–42
database searches 3–4, 7–17, 104
de novo protein modeling 90, 95, 104, 117–41
angle bending 119
bond stretching 119
conformational sampling 124–7
continuous-space reduced models 127–30
dihedral angles 120–1
electrostatic interactions 119
empirical force fields 118–21
experimental and predicted restraints 133–6
force fields models 118–24
Index

de novo protein modeling (Continued)
  high resolution lattice models 130–1
  in silico 117
  knowledge-based force fields 118, 122–4, 130
  molecular docking 123
  molecular dynamics 122–3, 128
  molecular mechanics 121–2
  multiscale modeling 131–3
  non-bonded interactions 119
  potential energy surfaces 121–2
  prediction algorithms 124–6
  protein folding processes 122–3
  protein gene function 261–3, 264
  quantum mechanics 118, 123
  reduced models 123, 124–30
  scoring functions 118–24
  sparse experimental data 133–4


dihedral angles 120–1
  displacements of local structures 109
  distance-based hierarchical clustering 237, 248–9
  distant homologues 190, 192–3
  disulfide bonds 51, 53, 129
  divisive hierarchical network clustering 246–7

DNA-binding domains 4–5, 111

see also protein–DNA interactions

Docking methods
  computational 264
  data use 216–18
  explicit flexibility 214, 224
  flexibility 213–16, 224
  implicit flexibility 213–14
  interpretation of results 225–6
  macromolecular complexes 211–30
  multiple docking 214–15
  performance 221–2
  practicalities 222–5
  protein–DNA interactions 214, 215–18, 222
  protein–ligand interactions 164, 222–4
  protein–protein interactions 178–80
  state-of-the-art 218–21

Docking problem 178–9

Domains 4–6
  databases 7–10
  DNA-binding 4–5, 111
  prediction 26–9
  sequence clustering 18

DSSP see Dictionary of Secondary Structure in Proteins

EC see Enzyme Commission

Electrostatic interactions 119

Empirical force fields 118–21

Ensemble docking 213–14

Enzyme Commission (EC) classification
  enzymes 187–8, 191
  protein gene function 66, 67, 260–1, 263, 273–4

Enzymes
  convergent evolution 190, 191–2
  distant homologues 190, 192–3
  domains 5
  identifying homologues 188–90, 205–6
  meta-servers 204–6
  overall protein structure 188–90
  protein gene function 66, 187–8
  meta-servers 204–6
  not inferred from homologues 202–4
  strategy 205–6
  template based prediction 191
  protein structure prediction 187–209
  structure-sequence motifs 200–2
  surface patch methods 202
  template based prediction 190–200

ERRAT 145

Evoluation frameworks 250–2

Evolutionary template identification 98

Explicit flexibility 214, 224

FASTA
  fold recognition 93
  protein gene function 64–5, 77–8
  protein sequence analysis 14–15
  flexibility 167, 213–16, 224
  flow-based methods 237, 242
  FlowerPower 19
  fold recognition (FR) 89–99
  force fields models 118–24
  Fourier transform infrared (FTIR) 47
  FR see fold recognition
  fragment assembly 95
  FTIR see Fourier transform infrared
  functional catalogue (FunCat) 66, 67–8
  functional genomics 231
  function analysis see protein gene function

Gene function see protein gene function

Gene fusion 265

Gene neighborhood methods 236–8, 239, 265, 269

Gene Ontology (GO) 66–74

3D structures 264
  assessment of prediction quality 251, 273
  automated methods 69–70
  benchmark results 72–4
  enzymes 187–8
  function prediction competitions 76
homology-based prediction 260, 262
large-scale experimental methods/networks 269
protein function prediction algorithm 70–4
protein–protein interactions 233–4
quantifying functional similarity 68
genome-sequencing 259, 265–7
Girvan-Newman (GN) algorithm 237
global Needleman-Wunsch algorithms 13
global SPC (gSPC) 19
globular domains 4, 26–9
GN see Girvan-Newman
GO see Gene Ontology
graph clustering 237–8, 245–9, 251–2
graph cut methods 237, 238–42
gSPC see global SPC
HADDOCK model 212, 214–25
helix-turn-helix domains 4
HHsearch 15–16
HHsenser 15–16
hidden Markov models (HMMs)
de novo protein modeling 134–5
fold recognition 91, 93, 96
protein gene function 65, 260
protein sequence analysis 12, 14–16, 18–19, 22–3
protein structure prediction 44
protein–ligand interactions 164
hierarchical clustering 237–8, 245–9
high resolution lattice models 130–1
high-throughput methods
protein gene function 259, 268, 270–1
protein–ligand interactions 177–8
protein–protein interactions 232, 235
HMMs see hidden Markov models
homology-based prediction
enzymes 188–90, 192–3, 205–6
homology modeling 88
limitations and problems 259–61
protein gene function 63–5, 259–61
protein sequence prediction 259, 263, 265
see also protein sequence analysis
hybrid methods of loop modeling 104
IDPs see intrinsically disordered proteins
ILP see integer linear programming
implicit flexibility 213–14
in silico protein structure prediction 117
integer linear programming (ILP) 241
interaction networks 267–9
interaction reliability 235–6
interactomes see protein–protein interactions
intermediate sequence searches (ISS) 14, 16, 96
intrinsically disordered proteins (IDPs) 5–6, 39–40, 50–1, 55
ISS see intermediate sequence searches
k-clique percolation clusters 247
kernel-based learning methods 249–50
knowledge-based force fields 118, 122–4, 130
knowledge-based template identification 98–9
large-scale experimental methods/networks 267–9
lattice models 130–1
linear motifs (LMs) 6–7, 11–12, 29–30
local graph clustering 237, 245
local Smith-Waterman algorithms 13, 17, 64–5
long-range contacts 51–3
loop conformations 214
loop modeling 90, 103–5
macromolecular complexes
data use in docking 216–18
docking methods 211–30
explicit flexibility 214, 224
flexibility 213–16, 224
implicit flexibility 213–14
interpretation of docking results 225–6
multiple docking 214–15
performance of docking servers 221–2
practicalities of docking methods 222–5
protein–DNA interactions 214, 215–18, 222
protein–ligand interactions 222–4
state-of-the-art methods 218–21
Markov cluster (MCL) algorithm 18–19
MCL see Markov cluster
MD see molecular dynamics
MetaMQAP 147
meta-predictors 94–5, 97
meta-servers 28–9
enzymes 204–6
fold recognition 94
protein structure prediction 45–6, 55
microarrays 66, 178
misalignment errors 108–9
misfolded loops/insertions 109
MM see molecular mechanics
model quality assessment programs (MQAPs) 143–57
3D-Jury 147
ABIPro-h 146
CASP7 143–4, 149–54, 272–4
Circle-QA 146–7, 155
comparative modeling 90, 98–9, 106
comparison of methods 154–5
model quality assessment programs (MQAPs) (Continued)
consensus based 147, 148–9, 154
ERRAT 145
global quality assessment 149, 150–2
historical overview 144–6
local quality assessment 149–50, 152–4
local versus global predictors 148–9
meta methods 147
predictive 146–9
PROCHECK 144, 145–6
ProQ 146, 148, 150–5
ProsaII 145
recommended improvements 155
Verify3D 144–5
Victor/FRST 146
WHATCHECK 144, 145–6
module discovery 237–8, 244–9, 251–2
molecular docking 123
molecular dynamics (MD)
de novo protein modeling 128
docking methods 213–14, 218
flexibility 213–14
knowledge-based force fields 122–3
macromolecular complexes 213–14, 218
molecular mechanics (MM) 121–2
Monte Carlo methods 124, 127, 129
motifs 6–12
MQAPs see model quality assessment programs
MSAs see multiple sequence alignments
multiple docking 214–15
multiple sequence alignments (MSAs)
de novo protein modeling 134–5
fold recognition 91–2
protein gene function 65, 76
protein sequence analysis 17, 19–26, 30
protein structure prediction 44, 53–4
multiscale modeling 131–3
nearest-neighbor searches (NNS) 44
see also gene neighborhood methods
Needleman-Wunsch algorithms 13
network alignments 238, 249
comparisons 267–9
dynamics 178
visualization software 233
network-based hierarchical clustering 237, 245–8
neural networks 44, 51, 204
new domain prediction 26–8
Newton–Raphson method 122
NNS see nearest-neighbor searches
non-bonded interactions 119
nuclear pore complexes 271–2
nucleic acid binding sites 111
nucleic acid polymerases 5
open reference frames (ORFs) 266
orthology 63–4, 260
pairwise alignments 10–17
pairwise potentials 130
PAM matrix 13
Parallel Tempering (PT) 127
paralogy 2–3, 64, 260
Pcons 147, 148–9, 154
PDB see Protein Databank
PDFs see probability density functions
PES see potential energy surfaces
PFP see protein function prediction
phylogenetic analysis 17, 23–6, 231, 265–7
phylogenomics 266
physical interaction networks 232–3
Pmodeller 154–5
pocket search approaches 165
position specific score matrix (PSSM) 14, 92–3, 168–9
position-specific iterated BLAST (PSI-BLAST)
de novo protein modeling 129, 135
enzymes 192
fold recognition 91–6
protein gene function 65, 69, 70, 72–4, 77–8
protein sequence analysis 14–16, 22, 29–30
protein structure prediction 44
protein–ligand interactions 164
post-translational modification 261–3
potential energy surfaces (PES) 121–2
probability density functions (PDFs) 102
PROCHECK 144, 145–6
profile-based methods 260
PROF_SIM 15–16
ProQ 146, 148, 150–5
ProsaII 145
Protein Databank (PDB) 160–3, 172–3, 200, 263
protein folding processes 122–3
protein function prediction (PFP) algorithm 70–4, 76–8
protein gene function 63–85, 259–79
1D sequence features 261–3
3D structures 263–5, 271
algorithms 236–50
assessment 272–4
automated prediction from sequence 63–85
comparative genomics 74–5
comparative performance 252–3
controlled vocabularies 66–8
cut-based methods 237, 238–42
de novo protein modeling 261–3, 264
distance-based hierarchical clustering 237, 248–9
enzyme commission numbers 66, 67
evaluation frameworks 250–2
flow-based methods 237, 242
function prediction competitions 76
functional catalogue 66, 67–8
functionally important residues 75–6
gene neighborhood methods 236–8, 239
homology-based prediction 63–5, 259–61
inference from sequence similarity 63–4
integration of prediction approaches 272–3
interaction reliability 235–6
kernel-based learning methods 249–50
large-scale experimental methods/networks 267–9
limitations and problems 66, 259–61
local graph clustering 237, 245
mathematical formulations 234
network alignments 238, 249
network visualization software 233
network-based hierarchical clustering 237, 245–8
notation 234–5
nuclear pore complexes 271–2
physical interaction networks 232–3
profile-based methods 260
protein function prediction algorithm 70–4, 76–8
protein–protein interactions 231–58, 267–9
quality assessment 250–3
quantifying functional similarity 68
recent/future developments 270–2
seeded module discovery 237–8, 245–6
subcellular localization 75
substrate specificity 260–1, 262
supervised learning 238
template based prediction 110–11
tertiary structure 46–50
tertiary structure 47, 48, 55
text mining 269–70
virtual screening 264
whole genome sequencing 259, 265–7
see also Gene Ontology
protein sequence analysis 1–38
database searches 3–4, 7–17
domains 4–6, 7–10, 18, 26–9
evolutionary mechanisms 1–4
homology-based prediction 259, 263, 265
multiple sequence alignment 17, 19–26
pairwise alignments 10–17
paralogs 2–3
phylogenetic analysis 17, 23–6
protein–DNA interactions 168–9
protein-encoding genes 1–2
protein–RNA interactions 174–6
sequence clustering 17–19
sequence motifs 6–12
workflow 30
protein structure prediction 3–4, 39–62
amino acid residues 51–3
benchmarks 54–5
disulfide bonds 51, 53
enzymes 187–209
intrinsically disordered proteins 39–40, 50–1, 55
long-range contacts 51–3
protein–DNA interactions 168, 169–70
protein–RNA interactions 171–4, 176–7
secondary structure 40–8, 55
solvent accessibility 44–7, 55
structure-sequence motifs 200–2
supersecondary structure 48–50
surface patch methods 202
tertiary structure 47, 48, 55
transmembrane proteins 47–8, 54
water-soluble proteins 44–7, 55
see also de novo protein modeling; template based prediction
protein–DNA interactions 161, 165–70
dNA-binding proteins 167–70
dNA-binding residues 168
docking methods 214, 215–18, 222
flexibility and adaptability 167
predictive methods 167–70
protein sequence analysis 168–9
protein structure prediction 168, 169–70
target sequences 168–70
thermodynamics and kinetics 166–7
protein-encoding genes 1–2
protein–ligand interactions 159–61, 164–5
docking approaches 164
docking methods 222–4
pocket search approaches 165
similarity search approaches 164–5
protein–nucleic acid interactions 50
protein–protein interactions 177–80, 231–58
algorithms 236–50
comparative performance 252–3
cut-based methods 237, 238–42
de novo protein modeling 117
disorder prediction 50
distance-based hierarchical clustering 237, 248–9
evaluation frameworks 250–2
flow-based methods 237, 242
gene neighborhood methods 236–8, 239
protein–protein interactions (Continued)

Ind JWBK331-Bujnicki November 25, 2008 7:26 Printer: Yet to come

Gene Ontology 233–4
interaction reliability 235–6
kernel-based learning methods 249–50
local graph clustering 237, 245
mathematical formulations 234
network alignments 238, 249
network visualization software 233
network-based hierarchical clustering 237, 245–8
notation 234–5
physical interaction networks 232–3
protein gene function 66, 69, 231–58, 267–9
quality assessment 250–3
seeded module discovery 237–8, 245–6
supervised learning 238
template based prediction 111
webservers 162–3

protein–RNA interactions 162, 170–7
biological importance 170
empirical rules 171–4
interface residues 175–6
predictive methods 170–1, 174–7
protein sequence analysis 174–6
protein structure prediction 171–4, 176–7
RNA interface residues 175–6
RNA-binding proteins 170–1, 174–5
similarity 170–1
PSI-BLAST see position-specific iterated BLAST
PSSM see position specific score matrix
PT see Parallel Tempering
 QA-ModFOLD 147
quality assessment of protein models see model quality assessment programs
quantum mechanics 118, 123
reduced models 123, 124–30
reverse position-specific BLAST (RPS-BLAST) 15–16
rigid body assembly methods 100–2
ROBETTA model 154–5
RosettaDock 218–21
RPS-BLAST see reverse position-specific BLAST
satisfaction of spatial restraints methods 101–2
Schrödinger’s equation 118–19
SCOP classification 191
scoring functions 118–24
secondary structure 4, 40–8
definition 41–3
intrinsically disordered proteins 55

known protein structures 41–3
prediction 44–8
solvent accessibility 44–7, 55
transmembrane proteins 47–8
water-soluble proteins 44–7, 55
secondary structure elements (SSEs) 40–2, 44, 48
seeded module discovery 237–8, 245–6
segment matching methods 101–2
sequence analysis see protein sequence analysis
sequence clustering 17–19
sequence motifs 6–12
sequence-based fold recognition 91
sequence-based template identification 98
sequence/structure-based fold recognition 91–3
side-chain conformations 109–10
side-chain modeling 90, 104, 106–7
signal transduction proteins 5
similarity search approaches 164–5
similarity of trees methods 265–7
single linkage (SL) clustering 18
Smith-Waterman algorithms 13, 17, 64–5
SMs see structured motifs
solvent accessibility 44–7, 55
SP see sum-of-pairs
spatial restraints methods 101–2
SPC see super paramagnetic clustering
SSEARCH 64
SSEs see secondary structure elements
statistical force fields 118, 122–4, 130
stochastic flow-based clustering 246–7
structural domains 4
structural genomics 263
structure prediction see protein structure prediction
structure-based template identification 98
structure-sequence motifs 200–2
structured motifs (SMs) 6–7, 11–12, 29–30
subcellular localization 75
substrate specificity 260–1, 262
sum-of-pairs (SP) scores 19
super paramagnetic clustering (SPC) 19
supersecondary structure 48–50
supervised learning 238
support vector machines (SVMs)
de novo protein modeling 134, 136
fold recognition 93
kernel-based learning methods 249–50
protein gene function 69, 75
protein structure prediction 44
protein–DNA interactions 168, 174–5
protein–protein interactions 249–50
surface patch methods 202
SVMs see support vector machines
tandem affinity-purification/mass spectrometry 270–1

target-template alignment 88–90, 99–100
template based prediction 87–115
accuracy and errors 106–11
algorithms 193–4, 197–8
comparative modeling 88–110
components 193–9
convergent evolution 191–2
data structures 193–4, 197–8
de novo protein modeling 90, 95, 104
distant homologues 192–3
enzymes 190–200
evolutionary principles 88–9
features 193–7
fold recognition 89–99
fragment assembly 95
libraries 199–200
loop modeling 90, 103–5
meta-predictors 94–5, 97
model building 90, 100–2
model quality assessment 90, 98–9, 106
model refinement 90, 102–3
protein gene function 110–11, 191
protein sequence analysis 26–7
protein structure modeling 110–11
published methods 195–6
quality scoring 193–4, 198–9
side-chain modeling 90, 104, 106–7
stepwise procedure 89–90
target-template alignment 88–90, 99–100
template identification 90, 97–9
webservers 197
template-free predictions see de novo protein modeling
tertiary structure
protein gene function 65–6
protein structure prediction 47, 48, 55
text mining 269–70
threading methods 93, 97
TIM barrel domains 4, 263
TM see transmembrane
transmembrane domains 4
transmembrane (TM) proteins 4, 47–8, 54
van der Waals forces 129
Verify3D 144–5
Victor/FRST 146
virtual screening 264
water-soluble proteins 44–7, 55
WHATCHECK 144, 145–6
whole genome sequencing 259, 265–7
wrong fold errors 108
wrong side-chain conformations 109–10
yeast two-hybrid techniques 270–1
ZDOCK 218–21
Zhang-Server model 154–5