
Index

- Abelson tyrosine kinase, 142
Ab initio method, 23, 28,
68–69
ABLE, 26
Accessible surface area, 290
ACE, *see* Atomic contact energy
Adenyl kinase, 30
All- α and all- β proteins, 213
AMBER force field, 246–247, 323–324
physical models for, 248
Aminoacyl-tRNA synthetase (aaRS), 136
Amyloid encephalopathy, 192
Anfinsen's thermodynamic hypothesis, 244
Antibody–antigen interactions, 283
Antibody/Antigen test, ZDOCK and, 296
Artificial evolution method, 22
ASA, *see* Accessible surface area
ASTRAL SCOP genetic domain sequence
subset list, 254
Atomic contact energy, 290, 293, 308
- Backbone-independent rotamer library, 24
Basic Local Alignment Search Tool, 20, 64, 70,
79, 185
Baum–Welch estimation procedure, 182,
185, 191
Bayesian probability formalism
knowledge-based scoring functions from,
252–254
neural network algorithms with, 258–259
Bayesian rules, 96
Bayes' theorem, 252
Bin-based construction method, 228
BLAST, *see* Basic Local Alignment Search Tool
BLOSUM62, 64, 139
Blue Gene project team, 68
BOINC software, 69
Boltzmann's principal, 10
Bond angles/length, 151
Boolean expressions, CAPRI scoring and, 287
Buckingham–Fowler potential, 250
- CABS, de novo method, 252
CAFASP, *see* Critical Assessment of Fully
Automated Structure Prediction
CAFASP-CONSENSUS, semiautomatic
meta-server, 29–30
cAMP-dependent protein kinase, 142
CAPRI, *see* Critical Assessment of Predicted
Interactions
 C_{α} RMSD, 287, 299
CASP, *see* Critical Assessment of Structure
Prediction
CATH Database, *see* Class, Architecture,
Topology, Homology
Center-star approach, multiple sequence alignment
and, 129
CHARMM program, 246–247, 293
Class, Architecture, Topology, Homology,
19, 71
Classification test, protein structure indexing and,
165–166
ClustalX program, 135
Clustering coefficients, 211
CM, *see* Comparative modeling
ColE1 repressor, 230
Collision free path, 221
Comparative modeling, 4, 20–25, 27–28, 45, 49,
52, 64–65
usefulness of, 54–55
Complex sequence searches, 80–81
COMPOSER, 22
Conformer clustering and high-resolution
refinement, decoy selection by,
268–270
Contact capacity score, RAPTOR
and, 103–104
CORNET method, 206
Critical Assessment of Fully Automated Structure
Prediction, 6, 26–31, 46, 74, 206
Critical Assessment of Predicted Interactions, 287
ZDOCK/RDOCK performance and, 288

- Critical Assessment of Structure Prediction, 5–6, 12–13, 16, 20, 26–31, 44, 50, 53, 62, 69, 74, 188, 206, 270
 docking algorithms performance and, 287
 experiments, 46–47, 50–52, 69
 measures, 47–48
- CTSS program, 149
- Cyclical search method, 24
- Database-based approach, loop prediction and, 23
- Dead-end elimination (DEE) algorithm, 24
- Decoy filters, design of, 267–268
- Decoy selection, by conformer clustering and high-resolution refinement, 268–270
- Decoy sets, knowledge-based scoring functions evaluation and, 261–265
- De novo protein structure prediction, scoring functions for, 243
 decoy filters, design of, 267–268
 decoy selection process, 268–270
 knowledge-based scoring functions, 251–257, 265–267
 neural network knowledge-based, 257–265
 physics-based energy functions, 246–251
- Desolvation free energy, ZDOCK and, 293
- DFT, *see* Discrete Fourier Transform
- Dihedral angle, 221, 225
- Dirichlet distribution, 234
- Discrete Fourier Transform, 291
- DISOPRED server, 79
- Distance-dependent energy, protein threading and, 11–12
- Distance matrix (DM), 199–200, 202
- Distance-scaled finite ideal gas reference state (DFIRE), 12
- Distributed Annotation System, 77
- Docking tools, 286
- Double-dynamic programing, 29, 67
- DSSP, 187
- EBI, *see* European Bioinformatics Institute
- EM, *see* Expectation-maximization
- ENCAD, 246–247
- Energy functions
 class I, 247–250
 protein structures, in aqueous environment, 250–251
 protein threading and, 9–14, 95–98
- e-Protein project, 77
- ESTs, *see* Expansive spaces trees
- Euclidean distance, 152, 199–200
- Euclidean 3D transformation, 132
- European Bioinformatics Institute, 79
- Evaluation of Automatic protein structure prediction (EVA), 44, 53
- E* value, of RAPTOR, 116–117
- Ewald convergence factor, 324
- Expansive spaces trees, 223
- Expectation-maximization, 182, 189, 191
- Extreme value distribution (EVD), 117
- Fast and Accurate Sidechain Topology and Energy Refinement (FASTER), 24
- FASTA, sequence-sequence alignment method, 20, 70, 79–80, 159
- Fast Fourier Transform, 285, 290–292, 299
- FFT, *see* Fast Fourier Transform
- Fold and Function Assignment System (FFAS), 70–71, 80
- Folding/unfolding transition, REMD simulations and, 316
- Fold recognition, 5–6, 9, 12, 20, 27–28, 45, 52, 55, 65–68, 100–101
 RAPTOR and, 107–116
 SVM approach to, 111–112
- Fourier component, 317
- FR, *see* Fold recognition
- FRAGFOLD method, 25, 69, 252
- Fragment assembly based methods, 25–26, 28, 55, 83
- Frozen approximation, 15, 29
- FR potential, 13
- FR/statistical significance, of threading alignments, 17–18
- FSSP, 19, 108
- FTDock, 285
- FUGUE, 15, 72, 81, 94
- Gap penalty, RAPTOR and, 103
- Gaussian distribution, 321, 324
- Gene3D, 74
- Generalized-Born (GB) approximations, 251
- Generalized-Born/surface-area (GBSA) models, 251
- Generalized suffix tree (GST) construction, 154–156
- 3D-GENOMICS, 74
- GenTHREADER method, 15, 71–72, 74, 81, 100
- Geometric hashing, 130, 148, 161, 163
- Gibbs sampling technique, 99
- Global Distance Test (GDT-TS), 48, 55, 77

- Global minimum energy conformation, 24
Glutathione S-Transferase family, 140
GMEC, *see* Global minimum energy conformation
gp39 mutations, 4–5
Gradient boosting algorithm, 107, 112
FR and, 113–114
GRAMM, 285
GRASP/Delphi, 250
Grid-based shape complementarity, 292–293
Grid technology, 77, 84
GROMACS, 324
GSC, *see* Grid-based shape complementarity
- Hash-based index, 149
Heuristic threading programs, 14–16
HGP, *see* Human Genome Project
Hidden Markov Models, 20, 29, 80, 95, 173, 206, 252
 bioinformatics and, 174
 local structure, 189–193
 profile, 193–195
 secondary structure, 187–189
 for transmembrane helices, 175–187
Hierarchical optimization strategy, 23
High-resolution refinement and conformer clustering, decoy selection by, 268–270
HitFinder, 299
HMMs, *see* Hidden Markov Models
HMMSTR, model for local structure, 190
 application of, 192–193
 training and topological modification of, 191–192
HMMTOP model, 184–186
Homology modeling, *see* Comparative modeling (CM)
HOMSTRAD database, multiple sequence alignments and, 138–140
Hooke's law, 247–248
HSP-60 protein fragment, contact maps and, 201
Human Genome Project, 4
Hybrid fold recognition searches, 81
Hybrid methods, 71
Hydration potential, 13; *see also* Energy function, protein threading and
Hydrogen-bonding potential, 13; *see also* Energy function, protein threading and
Hydrophobic–polar (H–P) energy function, 229
IFT, *see* Inverse Fourier Transform
INBGU, 72, 81
Insulin-like growth factors, 6
Interface RMSD (iRMSD), 285, 295
Intermediate Sequence Searches (ISS), 70
Interpolated RAPDF (IRAPDF), 256
Inverse Fourier Transform, 291
Isomorphic mapping, 16
- JACKAL Modeling Package, 22
3D-JIGSAW, 22, 65, 82
3DJury, 81
- Knowledge-based potential energies, in protein threading, 9–11, 97
- α -lactalbumin, wire skeletal model of, 6, 21
LFF profile vectors, 149
Ligand RMSD (lRMSD), 285
LikelihoodRatio, of amino acid, 134
Lindahl's benchmark, 108, 115
Linear and integer programming, RAPTOR and, 104–107
Linear integer programming (LIP) problem, 16
Livebench, 53–54
Local conformation (ψ , ϕ) potential, 13; *see also* Energy function, protein threading and
Local Feature Frequency (LFF) profile algorithm, 149
Local *versus* global contacts, 212–213
Lock-and-key mechanism, 284
Loop modeling method, 23, 92
Lysozyme, X-ray structure of, 6
- MALECON method, 130
Markov-chain theory, 230, 233
Markovian State Model, 232–235
MASS method, 130
MAX SNP-hard, 98
MD/MC simulations, 220–221, 225, 232, 234
Mean first passage time, 233–234
Meta-DP server, 79
Metropolis Monte Carlo (MC) simulation, 318
MFPT, *see* Mean first passage time
MODCHECK, 83
MODELLER, homology modeling program, 22, 25, 82

- Molecular dynamics (MD) simulations, of protein folding, 220
 REMD application and, 316–317, 323–327
 replica exchange algorithm, 318–323
 Monte Carlo (MC) methods, 24, 26, 209, 220
 Monte Carlo simulated annealing (MCSA), 244, 262
 MQAPs, 83
 MSM, *see* Markovian State Model
 Multiple linear regression approach, ZDOCK performance and
 high-quality predictions, 309–311
 medium quality predictions, 306–309
 Multiple sequence alignment, 125
 flexible alignment, 128
 HOMSTRAD database and, 138–140
 MultiProt, 131–133
 pairwise alignment cases, 135
 partial alignment, 126, 136–138
 sequence alignment, 128
 sequence order independent structure alignment, 135–136
 structural similarity and low sequence identity, 140–141
 structure-derived, 133–135
 subset alignment, 127–128, 136–138
 time efficiency, 128–131
 tyrosine kinase, loop movement in, 141–142
 MultiProt, 131–133, 135, 139
 MUMmer genome, 156
 MUSCLE, multiple sequence alignment method, 129
 MUSTA algorithm, 130
 MUSTANG method, 130
 M-ZDOCK, symmetric multimer docking with ZDOCK, 294
- National Center for Biotechnology Information (NCBI), 79–80
 NEST, 22, 25, 82
 Neural network algorithms
 with Bayesian probability formalism, 258–259
 decoy sets, scoring function evaluation and, 261–265
 local structures prediction and, 257–258
 training and post-processing of, 260–261
 Neural networks, 183, 187–188
 New fold (NF) methods, 28, 45, 68–69
 NMR, *see* Nuclear magnetic resonance
 NNs, *see* Neural networks
- Non-complementarity-determining region (non-CDR) loops, 295
 Non-HMM methods, 183
 Nose–Hoover temperature-coupling, 318, 324
 Nuclear magnetic resonance, 5, 22, 192, 244, 284, 323, 325, 327
- Optimal sequence alignment methods, 69
 Optimized Potential for Liquid Simulations (OPLS) all-atom force field, 23
- Pair-wise energy, 11
 Pairwise sequence alignment methods, 70, 135
 Parinello–Rahman pressure-coupling algorithms, 318
 Path-directed subdivision trees, 223–224
 PDB, *see* Protein Data Bank
 PDSTs, *see* Path-directed subdivision trees
 PISCES, 19
 Poisson–Boltzmann models, 250
 POSA method, 129
 Position-Specific Iterative BLAST, 20, 29, 64–65, 70–72, 74, 80, 93–94, 101
 3D-position-specific scoring matrix, 15, 72, 81, 94, 101
 Posterior probabilities, HMMs and, 178–180
 PriSM, 22
 Probabilistic Roadmap Methods (PRMs), 222–223, 225
 for protein-folding pathways, 225–228
 Probability density function (PDF), 116
 PROFcon, 206, 214
 PROFESY, 26
 Profile-based methods, 20–21
 Profile HMMs, 193–195
 Profile-profile methods, 13, 77
 ProGreSS, 149, 161, 163, 164, 166–167
 PROQ, 83
 PROSAIL, 82
 PROSPECTOR, 95
 PROSPECT, protein threading program, 16, 29, 81–82, 94–95, 99
 ProtDex, SSE-based method, 149
 Protein contact maps (CMs)
 HSP-60 protein fragment and, 201
 prediction, 205–207
 properties of, 203–204
 pros and cons of, 202
 protein structures and, 199–203

- reconstruction of protein structures and, 204–205
- small world and, 207–215
- Protein Data Bank, 5, 7, 19, 25, 30, 53, 65, 79, 91–92, 230, 262, 290
- in regression analysis data set, 303
- Protein domain swapping, 126
- Protein-folding mechanisms, 225
- Protein folding, molecular dynamics (MD)
 - simulations of, 315–318
 - REMD application and, 316–317, 323–327
 - replica exchange algorithm, 318–323
- Protein folding, roadmaps for, 219, 224
 - Markovian State models, 232–235
 - pathways, PRMs for, 225–228
 - Stochastic RoadMap Simulation (SRS), 229–232
- Protein–protein complexes, attributes of, 304
- Protein–protein docking, 283–285
 - benchmark for, 289–290
 - CAPRI experiment and, 287–289
 - predicted complexes, accuracy measurement of, 285–287
 - protein complex characters, ZDOCK
 - performance and, 298–311
 - ZDOCK/RDOCK/M-ZDOCK approach, 290–298
- Protein representation, roadmap methods and, 220–221
- Protein secondary structure prediction, methods for, 187–189
- Protein structure indexing, suffix tree and, 147, 149
 - classification test, 165–166
 - database indexing, 148
 - GST construction, 154–156
 - local feature extraction, 150–153
 - multiple structure alignment approaches, 148
 - normalization, 153–154
 - pair-wise structure alignment methods, 148
 - performance test, 166–167
 - PSIST, 150
 - querying, 156–160
 - retrieval test, 161–164
- Protein Structure Indexing using Suffix Trees, 150, 156, 161–164, 166–167
- Protein Structure Initiative, 92
- Protein structure prediction, template-based
 - Ab initio methods, 5
 - assessment of, 43–55
 - CASP/CAFASP, 26–31
 - comparative/homology model building, 21–25
 - de novo methods, 4
 - fold recognition (FR) (*see* Fold recognition)
 - fragment assembly, 25–26
 - future of, 83–84
 - key steps for, 7
 - major milestones in, 8
 - physics-based methods, 4
 - sequence-based alignment methods, 19–21
 - threading (*see* Threading, protein structure prediction by)
- Protein structures
 - in aqueous environment, 250–251
 - reconstruction, from contact maps, 204–205
 - small world and, 209–211
- Proteomes, databases serving structural annotations for, 74–77
- PROTINFO, de novo method, 252
- PSI-BLAST, *see* Position-Specific Iterative BLAST
- PSIPRED server, 72, 75, 94
- PSIST, *see* Protein Structure Indexing using Suffix Trees
- PSSM, *see* 3D-position-specific scoring matrix
- RAPDF scoring function, 255–256, 262, 265
- Rapidly exploring random trees, 223
- RAPTOR, optimal protein threading, 16, 18, 20, 29–30, 94–95
 - E* value of, 116–117
 - fold recognition, 107–116
 - integer programming formulation, 104–107
 - scoring function, 101–104
 - threading assumptions and model, 104
- RDOCK, refining ZDOCK predictions, 294
- Relaxins, 6
- REMD algorithm, *see* Replica exchange molecular dynamics algorithm
- Replica dynamics (RD) methods, 316–317
- Replica exchange molecular dynamics algorithm, 315–316, 320
 - three-helix bundle protein and, 323–327
- Replica exchange (RE) algorithm, 318–323
- Retrieval test, protein structure indexing and, 161–164
- Rigid-body method, 21–22, 65, 284, 299
- Rigidity-based sampling, 228
- Rigorous threading algorithms, 16
- RMSD, *see* Root mean square distance
- Roadmaps
 - construction method, 227
 - for protein folding, 224–235
 - protein representation, 220–221
 - robot motion planning, algorithms for, 221–224

- Robot motion planning, roadmap algorithms for, 221–224
- Root mean square distance, 24, 204–205, 262, 268–269, 285, 295, 299, 315, 323–325
- Rop protein, alignment of, 135
- ROSETTA method, 25–26, 69, 252
- RRTs, *see* Rapidly exploring random trees
- SAG1–antibody complex structure, 289
- SAM method, 80
- SAM-T98, 70–71
- Scale-free networks and contact maps, 213–214
- SCATD, 24
- SCOP, *see* Structural classification of proteins
- Scoring functions, knowledge-based, 251, 265–267
 - from Bayesian probability formalism, 252–254
 - compilation of probabilities, 254–255
 - pairwise distance, 255–257
- SCWRL method, 24, 82
- Secondary structure elements, 148–149
- Segment-matching method, 22
- seg* method, 79
- SegMod/ENCAD, 22, 25, 82
- Sequence alignment to known structure, steps for
 - models building, 82–83
 - sequence preparation, 77–79
 - structural annotation databases for model, 77
 - template structure, 79–81
- Sequence-based alignment methods, protein
 - structure prediction and, 19–21, 128
- Sequence-based search methods, 70
- Sequence Derived Properties (SDP), 72
- Sequence searching improvements, 70–71
- Sequence-structure alignment algorithms, 14–17
- SETTLE, 324
- SF sequences, 154, 156, 158
- SHAKE, 324
- 3D-SHOTGUN, 74
- Side-chain prediction methods, 23–24
- Side-chain–side-chain interaction potential, 13; *see also* Energy functions, protein threading and
- SignalP, 186
- Signal peptides prediction, HMMs and, 186–187
- SIMFOLD, 26
- Simple sequence search, 79–80
- Singleton energy, 10–11, 18
- Small world, contact maps (CMs) and, 207–208
 - all- α *versus* all- β contacts, 213
 - local *versus* global contacts, 212–213
 - properties, 214–215
 - and protein structures, 209–211
 - scale-free networks and, 213–214
- Smith-Waterman algorithm, 20, 64, 69, 160
- SNAPP potential, 265
- SPARKS, 20, 81
- SPratt2, 130
- SRS, *see* Stochastic RoadMap Simulation
- SSEARCH, 70, 80
- SSEs, *see* Secondary structure elements
- STACCATO, 139–140
- Stochastic RoadMap Simulation, 229–232
- STRIDE, 187
- Structural classification of proteins, 19, 29, 70–71, 93, 150, 160, 166, 213, 262–263, 290
- Structural superposition methods, 15
- Structure prediction from sequence, techniques
 - for, 62
 - ab initio and new fold methods, 68–69
 - comparative modeling, 64–65
 - fold recognition, 65–68
- Support vector machine, 18, 107, 109, 252
 - linear SVM regression, 110–111
 - nonlinear SVM regression, 111
- Surface Generalized Born implicit solvent
 - model, 23
- SVM, *see* Support vector machine
- SVM regression, 115–116
 - FR and, 111–112
 - linear, 110–111
 - nonlinear, 111
- SWISS-MODEL, 22, 65, 82
- TASSER method, 30, 252
- T-COFFEE method, 82, 130
- Template structure library, 19
- THREADER method, 67, 69, 71, 81
- Threading, protein structure prediction by, 8, 91
 - branch-and-bound algorithm, 99
 - divide-and-conquer algorithm, 99
 - dynamic programming algorithm, 98
 - energy function, 9–14, 95–98
 - fold recognition (FR) (*see* fold recognition (FR))
 - FR/statistical significance of threading
 - alignments, 17–18
 - RAPTOR (*see* RAPTOR, optimal protein threading)
 - sequence-structure alignment algorithms, 14–17
 - sequence-template alignment problem,
 - computational complexity of, 98
 - sequence to structure alignments and, 69–70

- targets and templates, representation
 - of, 93–95
- template structure library, 19
- tree decomposition-based algorithms, 99
- Threading scoring function, RAPTOR and, 101
 - contact capacity score, 103–104
 - environmental fitness score, 102
 - gap penalty, 103
 - mutation score, 102
 - pairwise contact score, 103
 - secondary structure score, 103
- TMHMM model, 183–186
- Torsion angles, 151–152
- Transition state ensemble (TSE), 231
- Transmembrane helices, HMMs for,
 - 175–177
 - more complex models, 180–181
 - parameter estimation, 181–183
 - posterior probabilities, 178–180
 - signal peptides prediction, 186–187
 - topology of, 183–186
 - Viterbi algorithm, 177–178
- Tree decomposition-based algorithm, 16, 24, 99
 - for threading problem, 16–17
- Tree-progressive* approach, multiple sequence alignment and, 129–130
- Triose-phosphate isomerase (TIM) barrels, 7
- Two-dimensional robotic configuration space, 221–222
- Type-1 beta hairpin, 190
- Tyrosine kinase, loop movement in, 141–142
- UNDERTAKER, 26
- Uniform distribution model, 12
- ϕ -values, 231–232
- van der Waals (vdW) potential, 292
- VERIFY3D, 82
- Viterbi algorithm, 174, 177–178, 181
- Wire skeletal model, of α -lactalbumin, 6, 21
- WU-BLAST, 70
- X-ray crystallography, 5, 48, 192, 284
- ZDOCK, 285
 - FFT-based initial stage docking algorithm, 290–294
 - M-ZDOCK and, 294
 - performance and protein complex characteristics, 298–312
 - performance on benchmark 2.0, 295–298
 - RDOCK and, 294
 - ZDOCK performance
 - on benchmark 2.0, 295–298
 - versus* interface curvature, 305
 - multiple linear regression approach, 306–311
 - and protein complex characteristics, 298–312
 - regression analysis of, 302–303
 - simple linear regression approach, 303–306
 - ZDOCK/RDOCK performance, CAPRI experiment and, 288
 - Z-score scheme, of threading, 18, 100–101, 115