# Index

#### а

active pharmaceutical ingredients (APIs) activity-specific subpockets 92 ADAPT program 107 adaptive biasing force (ABF) 387 adaptive molecular design 12 adenosine receptor (AR) 237 AllChem 350 AlleGrow 111, 115 AMBER program 103 amino acids 470-471 ant colony optimization 449-450 - and MHC-I binding peptides 450-456 antitarget activity reduction 296-298 artificial neural network 27, 443 associative networks 27 Astex "Rule of Three" 150 atom-based methods 101-103 atom property fields (APFs) 90-94 atomic force microscopic (AFM) 513 AutoGrow 107

### b

backbone cyclization 456-457 basic products 250 Bayes theorem 26 Bennett's acceptance ratio (BAR) 386 BICLAIM database 350-351 - application example 360-366 - focused libraries design 359 - fragment space 352-355 - FTrees fragment space search 356 - general workflow 351-352 - query 355-356 - scaffold selection 356-358 bioactive conformation 18 bioactivity estimation 207 biochemical assay screening 157-158

bioisosteres 417 422 – exemplar applications -- Drug Guru 429-431 – information-based bioisosteric replacement 427 – – SkelGen 431–433 - and isosterism, history of 418-420 - replacement methods 421-422 -- databases 422-424 -- molecular shape 426-427 - - molecular topology 426 – physicochemical properties 424–426 biolayer interferometry (BLI) 160-161 biophysical assay techniques 160-161 biophysical technique 155 BIOSTER database 422, 426 Boltzmann distribution 473 BOMB program 104 Bongard problem 10 Born–Oppenheimer approximation 5 BREED method 177 BRICS (breaking of retrosynthetically interesting chemical substructures) 333-334

## С

caDNAno 506 caDNAnoSQ 506 CAESA (computer assisted estimation of synthetic accessibility) program 33, 269 Cambridge Crystallographic Data Centre (CCDC) 422 Cambridge Structure Database (CSD) 422–423 CANADA program 503 cascaded neural network model 28 CASREACT (Chemical Abstracts Reaction Search) 270

De novo Molecular Design, First Edition. Edited by Gisbert Schneider.

© 2014 Wiley-VCH Verlag GmbH & Co. KGaA. Published 2014 by Wiley-VCH Verlag GmbH & Co. KGaA.

543

544 Index

CATS (chemically advanced template search) 338, 350, 426 CHARMm 104 ChEMBL database 289-293, 295, 301, 318, 423-424 Chemical Abstracts Service (CAS) 268, 328 Chemical Genesis program 107, 108 chemical similarity 9, 20 chemical space 3, 326-327 - assessment, and focused set of reactions 251-254 - complexity enumerating and sampling 61, 63-65 – models 4 - searching 105-108 - size estimation 327-328 - subspaces enumeration 328-330 ChemSpace technology 350 chromosomes 277 cloverleaf 529 COALESCE 240 cofactors and nonbiological protein assemblies 480-481 COLIBREE 338 CoLibri 335, 352 combinatorial space 325-326, 330-331 - chemical space 326-327 - - chemical subspaces enumeration 328-330 – size estimation 327–328 - generation 332 – computational space from chemical knowledge 334-335 -- from fragmentation 332-334 - manipulation 335-336 - querying 336 -- applications 340-341 -- fragment spaces 337-339 - - Markush structures 341-342 - - reaction-based combinatorial spaces 339-340 - visualization 342 CoMFA (comparative molecular field analysis) 218-222, 235, 341 complexity model, in molecular design 57 - 58- chemical space complexity enumerating and sampling 61, 63-65 - emergent behavior 74-75 - enhancements 60-61 - information content as factor in de novo design 69-72 - molecular interactions 58-60 - reductionism and drug design 67-69

- thermodynamic entropy and drug design 73-74 - validation 65-67 compound construction strategies 30-33 compound scoring strategies 33-35 - ligand-based scoring 37 - receptor-based scoring 35-37 computer-aided design (CAD) tools 506 computer docking 128 CONCEPTS program 103 CONCERTS program 108 Congreve's rule of threes 127, 131 Connolly algorithm 7 Connolly surface 7 consensus scoring 108-109 continuous spatial function 6 Contour algorithm 111, 115 CORINA 313, 428 Crooks equation 380 cumulant expansion 381

## d

3D-QSAR approaches 215-216 - current methods 216-217 - Leapfrog 217-219 - recent advances 219-222 D4 receptor potency optimization 301 dead-end elimination (DEE) 474-475 decision space 452-453 diversity selection algorithms 153 DNA-branched structures 499 - alternative designs 511-514 - DNA junctions de novo design 499-504 - scaffolded DNA origami design 505-511 - tile-to-tile bonding 504-505 DNA origami 498, 499 - scaffolded design 505-506 -- multilayer 509-511 -- three-dimensional structures 509 -- two-dimensional structures 506-509 DOGS software 41, 42, 249, 255, 270, 340 donepezil 296, 301 DREAM++ 33 druggability 162-163 Drug Guru 429-431 drug-like compounds construction, by Markov chains 311-313 - applications 314 -- BACE-1 inhibitors design with FOG 316 - 319-- FOG compounds target class prediction 314-316 - FOG algorithm and library generation 313 drug-likeness assessment 207

Index 545

DrugScore 109 dummy products 340 DyoBlock 104

#### е

e-LEA3D 104 E-NoVo (Enumerated de NoVo design) 104 electrophoretic mobility shift assay-SELEX (EMSA-SELEX) 525 emergent behavior 74-75 EMIL (Example-Mediated Innovation for Lead Evolution) 422 empirical scoring functions 36, 109 end capping 458-459 energy functions 471-472 energy landscape theory 468-469, 472-473, 478, 485 enthalpy 16 entropy 16 entropy-based formalism 476 evolution strategy 13, 14 evolutionary algorithms 106-107, 276-277, 294-295, 444-446 - multiobjective approaches 108 EVOscreen<sup>™</sup> HTS 158 Evotec EVOsource database 153 exit vectors 341

# f

F-DycoBlock 104 feature trees fragment space (FTreesFS) 337-338, 349-352, 354-357, 359, 361, 366 feedforward networks 27 ferritin proteins 485 fidelity, of structure 502 FieldAlign 428 fit quality (FQ) 166 fitness function 440, 443-445, 450, 453, 456, fragment space 332, 335 460 fitness landscapes 9-10 - modeling 21-26 - - artificial neural network 27 -- Gaussian process 28-30 – – naïve Bayes classifier 26–27 – support vector machine (SVM) 27–28 FLARM (flexible atom receptor model) 231, 235-236 FlexNovo program 104, 340 FlexX molecular docking software 104, 108 fluorescence correlation spectroscopy (FCS) 155 Flux 149 FNY327 126

foldability criteria and negative design 472 - 473force-field based scoring functions 108 four-base code 495, 500 fragment-assisted drug discovery (FADD) 182 fragment-based approaches 31-32 fragment-based design, of focused compound libraries 349-351. See also BICLAIM database fragment-based drug design (FBDD) 125, 144-148.375-376 - detection methods 132-134 - final products 137-138 - high-throughput screening (HTS) 126, 129 -- optimization 130 - - small-molecule library design 126-129 -- target identification 126 - library generation 130-131 – computational techniques for library refinement 132 - optimization 135-137 - screening 134-135 fragment-based methods 101-105 fragment binding energies. See superadditivity fragment evolution 165 fragment library 332 fragment-linking principles 177-182 fragment merging 165, 175-177 fragment molecular orbital (FMO) 164, 168, 173 fragment optimized growth (FOG) 311-312 - BACE-1 inhibitors design 316-319 - compounds target class prediction 314-316 - and library generation 313 Fragment Shuffling 104 fragment trees 338 "Frankenstein" approach to drug discovery 175 free-energy methods, in ligand design 373 - best practices 396-397 - calculation and choice of method 382-385 - current issues 385-387 - experimental data 385 - in silico binding affinity methods 377 - - free energy perturbation (FEP) 378-379 - - linear interaction energy (LIE) 381 -- MM-PBSA and MM-GBSA 382 – nonequilibrium approaches 380–381 - - potential of mean force (PMF) calculations 379-380

546 Index

free-energy methods, in ligand design (contd.) - - thermodynamic integration and alchemical transformations 377-378 - in lead optimization (LO) 373-374 – as emerging tool in industry 374–375 - - in fine-tuning ligand discovery 375-377 practical examples 387 – application to pharmaceutically relevant systems 389-396 - - model systems and studies 387-388 free energy perturbation (FEP) 378-379 - versus linear interaction energy (LIE) 383 versus potential of mean force (PMF) 383 - versus thermodynamic integration (TI) 384 functional/high-concentration screening 132 - 133g GA algorithm 106, 107 GANDI program 107-108 Gaussian process 28-30 Gedanken experiment 10 generated databases (GDBs) 268, 329-330 genetically evolved receptor models (GERMs) 232, 238 GENG algorithm 329 GenStar program 102-103

GeometryFit program 109 Gibbs energy 16 GIDEON program 503 GOLD docking software 154 G-protein-coupled receptor (GPCR) 84, 87, 91, 228, 240, 252, 254, 316 libraries 149 GRID program 69, 163 GROMACS 104 GroupBuild 103-104, 108 group efficiency (GE) 166 GrowMol program 102-103, 115

## h

halogen-enriched fragment libraries (HEFLibs) 152 Hartree-Fock method 5 high-throughput screening (HTS) 61, 63-65, 125-126, 129, 349-350, 360, 375, 376 - optimization 130 - small-molecule library design 126-129 – target identification 126 history, of de novo drug design 37-42

hit and lead identifications, from fragments 143 - fragment-assisted drug discovery (FADD) 182 - fragment-based drug design (FBDD) 144 - 148- fragment hit expansion and fragment evolution 170-175 - fragment library design incorporating computational methods 148 -- focused sets 154 – fragment optimization designing 154-155 - - molecular attributes and physicochemical properties 150-151 – screening methods influence on library selection 151 - - size of library and diversity 152-153 -- strategies 148-150 - - undesirable functionality removal 152 - fragment-linking principles 177-182 - fragment merging principles 175-177 - fragment prioritization for optimization 165 - - computational and thermodynamic methods for fragment selection and prioritization 167-170 -- efficiency metrics 165-167 - fragment screening 155 -- biochemical assay screening 157-158 - - biophysical assay techniques 160-161 – fragment library screening using computational methods 161-162 - - isothermal titration calorimetry (ITC) 160 - - ligandability screening using fragments 162-165 – membrane proteins assay techniques 161 -- NMR screening 156-157 -- SPR screening 157 -- thermal shift assays 158 - - X-ray crystallography screening 155-156 HSITE 101 Hückel theory of molecular orbitals 5 human pocketone 79 - compound activity prediction by ligand-pocket docking and scoring 87 - - optimizing pocket sets for reliable results 87-88 - - unusually large and multifunctional pockets 88-90 - diversity and redundancy 85-86

ligands clustering by functions and subpockets 92–94
pocketome-derived 3D chemical fields as activity prediction models 90–92
predicted pockets 79–83
validated compilation 83–85
hydrogen bonding 17–18
hydrophobic effect 18
hysteresis 379

## i

IADE (information automatic design) 427 icmPocketFinder 79–82 indole library 353 intelligence amplification (IA) 70 interligand nuclear Overhauser effect (ILOE) 156 ISOAK similarity measure 340 isothermal titration calorimetry (ITC) 160

j

Jarzynski's equality (JE) 380, 381

## k

Kemp elimination 478 kernel trick 27, 40 *k*-nearest neighbor technique 27 KNIME ((Konstanz Information Miner) tool 274–275, 280–281 knowledge-based scoring functions 36, 109 known drug space (KDS) 327, 328

## I

Laplacian-modified naive Bayesian (LMNB) model 291-292 LEA3D algorithm 104, 107-108, 115 Leapfrog 217-219 leave-one-out (LOO) cross-validation 233 LEGAND program 102 LEGEND program 108 Lennard-Jones potential 79 leucine zipper (bZIP) 479 libDOGS software 257-258 library encoding schemes 249-250 ligandability screening, using fragments 162 - 165ligand-based molecular design 227-231 – pseudoreceptor algorithms 231–232 - successful applications 232-240 ligand-based scoring 37 ligand-based topological pharmacophore (LBTP) 426 ligand-binding cavities 79 ligand efficiency (LE) 21, 135

ligand-efficiency-dependent lipophilicity (LELP) 167 ligand-induced structure-activity relationship display (LiSARD) 23 ligand lipophilicity efficiency (LLE/LiPE) 167 ligand lipophilicity efficiency AT (LLEAT) 167 LigBuilder 39, 104, 113-115 LigPrep 313 LigSite 163 linear interaction energy (LIE) 381 - versus FEP/TI 383 Lipinski's rule of fives 127, 131, 137, 207, 212. See also Rule of 5 lipocalins 480 lipophilic ligand efficiency 21 lipophilicity 57, 66, 72-73 local optimization 11 LoFT (library optimizer using feature trees) 335 - 336LUDI program 103, 109, 111-113 LY-517717 (Eli Lilly) 110-111

## m

Markush structures 331 mass spectrometry (MS) 133 matched molecular pairs (MMPs) 423 maximum common substructure (MCS) 290-291 MCDNLG program 106 mean field theory 475 MED-Hybridize 104 MED-Portions 104 MEGA program 108, 115, 280 melting curves 502 membrane proteins 481-483 - assay techniques 161 metadynamics 387 Metropolis Monte Carlo 106 mirror image chemistry 534 MM-PBSA and MM-GBSA versus FEP/TI 383 MOGA (multiobjective genetic algorithm) 278-279 molecular design cycle 9-14 molecular docking 154 molecular fingerprints 314, 316 molecular representation 1-9 molecular surfaces 6 MolTransformer node 275, 280 Monte Carlo-based methods 106 Monte Carlo methods 473-474, 529 Morgan algorithm 71 MoSELECT program 279

multidimensional scaling (MDS) 23 multilayer perceptrons (MLPs) 27 multiobjective de novo design 279-282 multiobjective de novo design, of synthetically accessible compounds 267-270 using evolutionary algorithms 276–277 - - multiobjective *de novo* design 279-282 – multiple objectives optimization 277-279 - using reaction vectors 270-275 multiobjective evolutionary algorithm (MOEA) 40 multiple copy simultaneous search (MCSS) 101.163 multitarget profiles and ligands de novo design 287 - 289- antitarget activity reduction 296-298 - D4 receptor potency optimization 301 - evolutionary algorithm 294-295 - experimental validation 295-297 - ligand design creativity automation 289-293 - novel ligands designing to defined profile 301.303-304

Muse program 218–223

#### n

naïve Bayes classifier 26-27 naïve Bayesian models 314 Nano-Engineer-1 program 503 NanoEngineering-1 506 NAUTY software 330 negative design 11, 23 NEWLEAD program 105 NSGA-II algorithm (nondominated sorting genetic algorithm) 279-280 nuclear magnetic resonance (NMR) 61, 73, 133-134, 156-157, 470, 477-479, 481-483 nucleic acid structures 495-499 alternative DNA designs 511–514 – DNA-branched structures 499 -- DNA junctions de novo design 499-504 -- tile-to-tile bonding 504-505 - scaffolded DNA origami design 505-506 -- multilayer DNA origami 509-511 -- three-dimensional structures 509 – two-dimensional structures 506–509

# 0

octet rule 418 OpenEye docking 87 open molecule generator (OMG) 330 orthogal screening approach 155

## р

P450 CYP121 inhibitor 175, 176 pairwise interaction energies (PIEs) 173 pan assay interference compounds (PAINS) 152 Pareto frontier 278 Pareto ranking 278, 280, 293 PARM (pseudoatomic receptor model) 232, 233 partial least squares 215 particle swarm optimization 446, 448-449 peptide design by nature-inspired algorithms 437 - chemical modification 456 -- backbone cyclization 456-457 -- end capping 458-459 -- stapling 458 -- sugar coating 459-460 - MHC-I binding peptides, by ant colony optimization 450-456 - optimization 441-444 -- ant colony optimization 449-450 -- evolutionary algorithms 444-446 - - particle swarm optimization 446, 448-449 - template-based design 437-441 pharmacophore-based de novo design 201-202 - v1.0 -- basic scheme 202 - - fragment and linker databases 203-204 - - fragments connection by linkers 205 - 206-- fragments mapping 204 - - generated molecules assessments 206 - 208- v2.0 updated version modifications 208-209 -- designated fragment 209 – molecular construction process conformation optimization 209-210 - - two pharmacophore features sharing one fragment 210 - validation 210-212 pharmacophore-based methods 105 pharmacophore searching 129 PharmHand<sup>™</sup> software 129 PhDD program 108 phenotypic screening 264 phenprocoumon 146 pheromone matrix 452 photoaptamers 525 photo-SELEX approach 525 PI3Ky inhibitor 175, 176

plane of best fit (PBF) 150-151 PLANTS program 104 PLX4720 147 PNU-103017 146 PNU-96988 146-147 polymerase chain reaction (PCR) 522, 524, 534 polypharmacological fitness landscapes 23, 25 positive design 11, 23 potential of mean force (PMF) 109 - calculations 379-380 - versus FEP 383 - versus TI 383-384 PPARy 80-82, 88-90, 92-94 PrGen software 231, 232, 237, 238 principal components analysis (PCA) 22, 63 PRO SELECT 110 probabilistic approach 475-477 projection to latent structures (PLS) 22 PRO-LIGAND 104 promiscuity 66-67, 72, 84 prostate-specificmembrane antigen (PSMA) 522 Protein Data Bank (PDB) 83, 84, 97, 104, 531 protein design 467-469 - catalysis redox activity, and enzymes, 477 - 478 – cofactors and nonbiological protein assemblies 480-481 - - *de novo* design and redesign 478-479 – membrane proteins 481–483 – protein–protein interactions and protein assemblies 483-485 -- protein reengineering 479-480 - elements 470 -- amino acids 470-471 - - energy functions 471-472 - - foldability criteria and negative design 472 - 473 – sequence search and characterization 473-477 – – side-chain conformations 471 - – solvation 472 - - target structures 470 protein-ligand complex 103, 104, 109 protein-ligand crystal structures 59, 68, 69, 73 protein-ligand surface interaction analysis (PLSIA) program 240 protein-protein interaction inhibitor (PPII) 172 protein-protein interactions (PPIs) 163-165, 171-173

protein reengineering 479–480 pseudohalogens 419 pseudoreceptor algorithms 231–232 pseudoreceptor point similarity (PRPS) 239 PubChem 330

## q

quantitative structure–activity relationship (QSAR) 29, 69, 150 Quasar algorithms 238

# r

rank, of structure 500 rapid (elimination of swill) approach 22 RASSE program 103, 108 reaction-driven de novo design 245-247 - of chemical libraries addressing target families 256-261 - successful applications 254-256 - synthetic feasibility problem tackling 247-249 – electronic laboratory notebooks 249–251 – focused set of reactions chemical space assessment 251-254 reaction vectors 270-275, 280-282 reagent clipping 353-354 RECAP (retrosynthetic combinatorial analysis procedure) 33, 149, 154, 269, 333, 335, 338, 341, 350 receptor-based scoring 35-37 receptor-ligand interaction 14-21 ReCore 149, 340-341 recurrent networks 27 reductionism and drug design 67-69 REOS (rapid elimination of swill) 152 REPLACE method 172 retrosynthetic fragmentation 332 R-groups 341 riboswitches 520-521 RNA aptamer design 519 - functional analyses (in vitro and in vivo) 532-533 - future perspectives 535-536 - problems 533-535 - riboswitches 520-521 – SELEX 521–522 --method 522-525 – speeding up by computational methods 526-530 - - technical challenges and recent developments 526

- structures and probing methods 530-532

RNA interference (RNAi) 520 ROCS (rapid overlay of chemical structures) 161–162, 356, 357, 362, 363 rotatable bond 207 roulette-wheel selection algorithm 206 Route Designer 33 rubredoxin 478 Rule of 5 326–327. *See also* Lipinski's rule of fives

# S

SARSE 506 scaffold 252-254, 259-260, 329, 499 - selection 356-358 Schrödinger equation 4, 5 Scientific Vector Language (SVL) 170 scoring methods 108-109 selective optimization of side activities (SOSAs) 295 self-organizing map (SOM) 23, 456 sequence symmetry minimization principle 501-503 SEQUIN program 499-503 Shannon entropy 70-71, 73, 441 SHAPE chemistry 532 SHAPES library 149, 154 side-chain conformations 471 similar property principle 424 simulated molecular evolution (SME) 445, 448 single-stranded DNA (ssDNA) 495, 499, 505, 511-513 site identification by ligand competitive saturation (SILCS) method 163 size-independent ligand efficiency (SILE) 166 SkelGen program 105, 431-433 small molecule growth (SMoG) algorithm 311-312 SMARTS Reaction 251 SMILES 313, 356, 360, 429 SMIRKS 270, 271, 275, 429 SMoG program 102, 109 social memory 446 solvation 472 solvent-accessible surface (SAS) 6-7 spherical harmonics 8-9 spiegelmers 534 SPLICE program 105 SPROUT 110, 269 staple strands 505 stapling 458 StarDrop 292, 296

statistical potential based scoring functions. See knowledge-based scoring functions steered molecular dynamics (SMD) 380-381 stochastic algorithm 13-14 stochastic proximity embedding (SPE) 23 strong causality 9, 445 β-structures 478 structure-activity relationship (SAR) 1, 3, 4, 9-10, 23, 245, 256, 358-359. See also ligand-based molecular design structure-based de novo drug design (SBDND) 97-99 - current progress in methodologies 99 – binding site identification 100–101 -- chemical space searching 105-108 -- molecular design 101-105 -- scoring methods 108-109 -- synthetic accessibility 110 - recent applications 110-114 structure-based drug design (SBDD) 145, 227-229, 241 structure kinetic relationship (SKR) 157 structure-thermodynamic relationships (STRs) 160 sub-substructures (SSSs) 338 sugar coating 459-460 superadditivity 180 support vector machine (SVM) 27-28 surface plasmon resonance (SPR) 61, 133, 151, 157 - SPR-SELEX 524 SYBL 217 SYLVIA program 33, 110 SYNOPSIS 33, 110, 247, 249, 255, 270, 339 - 340synthetic accessibility 110 - assessment 207-208 Systematic Evolution of Ligands by EXponential Enrichment (SELEX) 521-522 - method 522-525 - speeding up by computational methods 526-529 – design of structures 529–530 - technical challenges and recent developments 526 t tailored-SELEX 525

tailored-SELEX 525 target-immobilized NMR screening (TINS) 151 tenascin 479 theoretical fragment linker factor 180 thermal shift assay 158 - (fluorescence-based) 133 thermodynamic entropy and drug design 73-74 thermodynamic integration (TI) - and alchemical transformations 377-378 - versus FEP 384 thioredoxin 477 TIBCO Spotfire 259, 262–263 tile-to-tile bonding 504-505 Tipranavir 145–147 toggle-SELEX 525 TOPAS software 39, 247, 350 TopClass algorithm 317-318 topomers 221-223, 341 triose phosphate ismerase (TIM) barrel family 470, 478

### и

umbrella sampling 379 Uniquimer 3D program 503–504 untranslated region (UTR) 520 Urbild 3–5

## V

Vemurafenib 147 virtual screening 375–376, 382, 383, 396, 397 VTP-27999 (Vitae Pharmaceuticals) 111–113

## w

WaterMap 261
Watson-Crick base pairing rule 495, 496, 499
weave tiles 513
weighted histogram analysis method (WHAM) 379, 387
WhichDock™ 128

## X

X-Cscore 109 X-ray crystallography 61, 134, 470, 485, 531 – screening 155–156

## z

ZINC database 131