

Index

a

- absolute metabolomics and
 - thermodynamic analyses 272
- acetyl-coenzyme A (acetyl-CoA) 268
- additional reactions, draft to functional network
 - biomass reactions 7
 - demand reactions 6
 - exchange reactions 6
 - non-growth associated ATP maintenance reactions 7
 - spontaneous reactions 7
 - transport reactions 6
- AGORA 2 collection 61–62
- AGORA collection 59–60
- AI-agent for autonomous therapeutic peptide design 203–205
- Akaike Information Criterion (AICc) 393
- alcohol dehydrogenase (YqhD) 268
- AllerCatPro 2.0 198
- AllerTOP v.2 198
- AlphaFold 111, 113, 114, 145, 170, 202, 307
- alpha-ketoglutarate decarboxylase (KivD) 268
- amino acid auxotrophies 22
- amino acid composition (AAC) 134–136, 194
- AMPTrans-LSTM 201
- ANNA16 232–234
- anticancer peptide (ACP) 189, 195, 199
- antifungal peptide (AFP) classification model 194
- antimicrobial peptide classification models 192–193
- anti-parasitic peptides (APPs) 195
- antiviral peptide (AVP) classification model 193–194
- artificial intelligence (AI) 63, 69, 75, 116, 131–153, 161–178, 187, 202–206, 220–222, 226–228, 236, 240–243, 253–260, 264, 273–274, 302, 333, 334, 341, 345, 354, 356
 - conventional computational approaches for predicting enzyme function 132–133
 - microbial cultivation process optimization 238–240
 - prediction of enzyme functions
 - using deep learning 138–147
 - using machine learning 133–138
- artificial intelligence (AI)-based metabolomics data processing and analysis 273–274
- artificial intelligence (AI)/machine learning (ML) enabled retrobiosynthesis
 - computer-aided synthesis planning 161
 - data standardization and model interpretability 177–178
 - enzyme prediction and design capabilities 176–177
 - enzyme selection and optimization 168–169
 - de novo* enzyme design and discovery 172–174
 - enzyme catalytic efficiency 170–172
 - enzyme engineering 172
 - enzyme substrate specificity 169–170
 - integrating biocatalysis with chemocatalysis 175

- artificial intelligence (AI)/machine learning (ML) enabled retrobiosynthesis (*contd.*)
- next-generation AI for
 - retrobiosynthesis planning 175–176
 - ranking of synthesis routes 168
 - retrobiosynthesis 162
 - searching algorithm 167
 - template-based tools 162–166
 - template-free tools 166–167
- artificial metabolic network (AMN) 299
- artificial neural networks (ANN) 1, 138, 230, 234, 236, 304, 317, 320, 327, 331–332, 368–370, 377, 382–384, 389, 391, 394, 397
- ASKCOS 166
- atom mapping 178, 324, 326, 327, 329
- AuReMe (Automatic Reconstruction of Metabolic Models) 57–58
- AutoDock 170
- AutoKEGGRec 54
- automated biofoundries 341–342
- challenges for implementing 352–355
 - design-build-test-learn cycle 342
 - components 342–345
 - description and their operating parts 345–347
 - laboratory workflows with potential of automation 347–350
 - strain engineering 345
 - geographical distribution 350–351
 - perspectives 355–356
- automated GSM collections 59
- AGORA 59–60
 - AGORA 2 collection 61–62
- construction tools
- AuReMe 57–58
 - AutoKEGGRec 54
 - CarveME 52
 - CoReCo 50
 - flux analysis and modeling environment (FAME) 54–55
 - GEMSiRV 56
 - MEMOSys 58
 - Merlin 51
 - MetaDraft 56
 - metaGEM 50–51
 - MetExplore 58–59
 - MicrobesFlux 55
 - ModelSEED 46, 50
 - MS2/OMEGGA on KBase 51
 - PathwayTools 57
 - pyFBA 53
 - Raven 53
 - rBioNet 54
 - reconstructor 52
 - SuBliMinaL 56–57
 - EMBL GEM 60–61
 - future directions 62–63
 - general approach 45–46
 - MetaGEM 61
 - PATHGENN collection 62
- Automated Recommendation Tool (ART) 274, 301, 305
- autoregressive modeling (AR) 81–84
- b**
- Basenji 86
- basic local alignment search tool (BLAST) 45, 46, 50, 51, 56, 105–106, 109, 113, 114, 132–133, 136, 138, 169
- biclustering 296, 297, 304
- bidirectional encoder representation from transformers (BERT)-based MLM approach 106
- bidirectional encoder representations from transformers (BERT)-style models 83
- Big Data in biotechnology 287
- data multimodality and heterogeneity 287–288
 - handling of missing values and outliers 289–290
 - longitudinal studies 290
 - stratification and data transformations 288–289
- BioAutomata 301
- BioBricks 344
- bioinformatics 50, 133, 186, 187, 301, 317
- BioNavi-NP 162, 167
- bioprocess modeling and optimization
- challenges in 367–369
 - hybrid modeling
 - biosystem optimization 373, 377
 - discrepancy 384–385
 - dynamic optimization 385–386
 - dynamic optimization under uncertainty 386–387
 - greyness 384
 - literature review 372–373

- machine learning component 385
- machine learning model
 - construction 381–383
- mechanistic model construction 378–380
- microalgal lutein production 387–397
- optimal feeding in fed-batch bioprocesses 386
- time-varying parameter estimation 380–381
- uncertainty estimation 383–384
- machine learning 369
 - artificial neural networks 370
 - ensemble learning 371
 - future prospects 371–372
 - Gaussian processes 370–371
 - principal component analysis 369
 - reinforcement learning 371
- blocked metabolites 8
- Borzoi 86–87
- bottom-up approach, for network reconstruction 3–4
- Braunschweig Enzyme Database (BRENDA) 3, 111, 162, 163, 169, 177, 342
- C**
- capillary electrophoresis (CE) 265–267
- Carbon-13 (¹³C) 7, 23, 29, 272, 275, 323, 324, 326, 334
- CarveME 5, 50, 52, 60
- CarveMe-generated EMBL GEMs 60
- cell line development and metabolic profiling 27–29
- Chai-1 113, 114
- clustering 227, 228, 290, 296, 297, 300
- CNN-BiLSTM hybrid model 194
- CodonTransformer 75–76
- comparative metabolic reconstruction (CoReCo) 50
- compound-protein interaction predictor (CPI-Pred) 108, 117–118
- computer-aided synthesis planning (CASP) 161
- concept bottleneck pLM (CB-pLM) 120
- conditional VAE (cVAE) 200
- constraint-based* methods 10
 - flux balance analysis 11–12
 - flux sampling 13–15
 - flux variability analysis 12–13
- constraint-based modeling (CBM) 26, 53, 55, 237, 286, 291, 298–300, 303–306
- contrastive learning 107, 115, 145, 147, 228
- convolutional neural networks (CNNs) 69–70, 89, 105–107, 139–141, 144, 171, 193–197, 199, 232
 - in regulatory sequence analysis 70–75
- CRISPR Associated Software for Pathway Engineering and Research (CASPER) 240
- CRISPR-based genome editing 221, 243, 267
- CRISPR technology 22
- critical process parameters (CPPs) 31
- critical quality attributes (CQAs) 31
- culturing 219–244, 265
- CYP7 133
- cytotoxicity 183, 195
- d**
- data multimodality and heterogeneity 287–288
- dead-end reactions 8, 9
- decision trees (DTs) and ensembles 293
- Deep-AFPpred 194
- Deep-AntiFP 194
- DeepEC 107, 139
- DeepECtransformer 107, 142, 144
- DeepGO 106, 107
- DeepGO-SE 106
- DeepGOGAT-SE 147
- DeepGOPlus 106, 139
- DeepGraphGO 106
- DeepImmuno 197
- deep learning (DL) 62, 69, 70, 72, 77, 85, 86, 89, 90, 93, 109, 117, 118, 148–153, 171, 173, 190, 192–198, 206, 228, 231–234, 236, 241, 242, 286, 293, 298, 303, 306, 307, 320, 331–333, 345
- convolutional neural networks 139–141
 - enzyme functions prediction 138–147
 - graph neural networks 145–147
 - model-guided sequence design 70–85
 - models 104, 106, 108, 116, 117, 120, 132, 138, 175, 178, 186, 202, 203, 229, 233, 302
 - recurrent neural networks 141–142

- deep learning (DL) (*contd.*)
 sources of sequence–function data
 85–90
 synthetic biological parts evaluation
 using motif analysis 90–93
 transformer and protein language
 models 142–145
 deep neural networks (DNNs) 68, 69,
 104, 138, 139, 141, 152, 178,
 194–196, 370
 DeepSEED 89
 DeepSTARR 89, 90
 DeepVirFinder 231
 demand reactions 6
de novo enzyme design and discovery
 172–174
 design-build-test-learn cycle (DBTLc)
 267, 285, 286, 295, 301, 302, 306,
 341–352, 354–356
 DIAMOND 52, 105, 107, 144
 Diff-AMP 193
 diffusion models 69–71, 76, 79, 84–85,
 173, 183, 193, 200, 201
 directed evolution 68, 69, 71, 172, 302,
 346, 349
 discrepancy hybrid modeling 384–385
 DNA synthesis and assembly 347–348,
 354
 dynamic flux and isotope labeling analysis
 332–333
- e**
 elementary metabolite units (EMU) 323
 EMBL GEM (genome-scale models)
 collection 60–61
 EMOPEC 68
 Enformer 86
 ensemble learning (EL) 171, 193, 197,
 288, 328, 371
 ensemble modeling 320
 enzyme activity 107–108, 116–118, 120,
 133, 170, 172, 176, 178, 287, 318
 enzyme catalytic efficiency 118, 170–172
 enzyme-coding genes 131
 enzyme commission (EC) number 3, 51,
 54, 106–107, 115, 120, 134,
 137–139, 141–142, 144–145, 147,
 152, 162, 169, 174
 enzyme engineering 107, 116, 120, 152,
 169, 171, 172, 177
 EnzymeMiner 344
 enzyme substrate specificity 169–170
 Evolutionary Scale Modeling (ESM) 170,
 190
 Evo model 76, 84, 114
 eVOLVER platform 239
 exchange reactions 6, 10, 16, 22, 53, 326
 ExpressionGAN 80
 extreme gradient boosting (XGBoost)
 194, 195, 293
- f**
 FastGapFill 8
 Fastknock 344
 fast leak test function 10
 flow cytometry 221, 224
 flux analysis and modelling environment
 (FAME) 54–55
 flux balance analysis (FBA) 8–16, 22–26,
 30, 50, 52, 53, 55, 57, 59, 272, 286,
 295, 298–300, 303–306, 321–324,
 327
 flux sampling 13–15, 23, 31, 330
 flux variability analysis (FVA) 8, 9,
 12–13, 55, 56, 59
- g**
 Gaussian Processes (GPs) 290, 370–371
 GearNet 147
 GEM reconstruction
 bottom-up approach 3–4
 top-down approach 4–5
 GEMSiRV 56
 gene essentiality 22, 30, 50, 324, 327
 gene ontology (GO) 3, 106, 111, 114,
 138–139, 141, 142, 144, 145, 147,
 152, 275
 gene-protein-reaction (GPR) 3, 44, 46,
 52, 56, 59
 rules 43, 45, 51, 53, 54
 Generalized Pre-trained Transformers
 (GPTs) 70
 Generative Adversarial Network
 Drug-tArget Ligand Fructifier
 (GANDALF) 200, 201
 generative adversarial networks (GANs)
 4, 69–71, 76, 77, 80, 84, 174, 183,
 186, 193, 200–201, 289, 293, 307,
 320, 327
 generative artificial intelligence (GenAI)
 75–76, 93–94, 253
 limitations 259–260

- LLM automated data extraction for
 - machine learning 258–259
 - text mining using knowledge graph tools 254
 - GraphRAG 256–258
 - NEKO 254–256
 - generative deep learning (DL) models 69
 - Generative Pre-trained Transformer (GPT) 80, 83
 - generative sequence modeling
 - data preparation for unsupervised learning 76
 - design of biological sequences 77–80
 - diffusion models 84–85
 - transformer-based DNA models 80–84
 - gene regulatory networks (GRNs) 2, 5, 24–25, 328
 - genetic algorithm (GA) 89, 192, 234, 320
 - genetic toggle switch 67
 - genome-scale metabolic models (GEMs) 2, 305, 342
 - genome-scale metabolic network models (GSMs) 43
 - automated development 45–59
 - manual creation 44–45
 - genomic and metagenomic data
 - processing 230–234
 - genomic data 2, 3, 72, 82, 230–234, 237, 241
 - genomic language models (GLM) 80, 114
 - glucose-6-phosphate dehydrogenase (G6PD) 27
 - GM-Pep 200
 - gradient boosting (GB) 230, 293, 371
 - graph-based biclustering algorithm 304
 - graph-based learning 30
 - Graph Isomorphism Networks (GINs) 196
 - graph neural networks (GNNs) 30, 145, 146, 171
 - GraphRAG 253, 256–258, 260
 - greyness 384
- h**
- HelixGAN 201
 - hemolysis evaluation 197–198
 - HHblits 105
 - hidden Markov models (HMMs) 53, 114, 133
 - high-content imaging and phenotyping 225–226
 - high-throughput (HTP) microbial technology
 - definition and scope of 221–222
 - flow cytometry 224
 - high-content imaging and phenotyping 225
 - matrix-assisted laser desorption/ionization-time of flight mass spectrometry 223
 - metagenomics 223
 - microfluidics and Lab-on-a-Chip (LOC) devices 224–225
 - next-generation sequencing (NGS) technologies 223
 - hybridisation of network-based and data-driven tools 26
 - hybrid ML and constraint-based models
 - CBM-FBA as input for ML 299
 - ML as input for CBM-FBA 299–300
 - hybrid modeling
 - discrepancy hybrid modeling 384–385
 - dynamic optimization 385
 - optimal feeding in fed-batch bioprocesses 386
 - under uncertainty 386–387
 - greyness 384
 - machine learning component 385
 - machine learning model construction 381–383
 - mechanistic model construction 378
 - microalgal lutein production 387–388
 - artificial neural network implementation for time-varying parameters 391–393
 - dynamic optimization 394–397
 - experimental setup and data availability 388–389
 - preliminary kinetic modeling 389–391
 - time-varying parameter estimation 380–381
 - uncertainty estimation 383–384
 - hydrophilic interaction liquid chromatography (HILIC) columns 266
 - HyenaDNA 82, 84, 90, 114

i

iGEM Registry of Standard Biological Parts 67
immunogenicity identification 197
INeo-Epp 197
infinite loops 9
information-rich isotope tracers 327–328
in silico predictions 15
 amino acid auxotrophies 22
 constraints 15–16
 gene essentiality analysis 22–23
 growth rate predictions 22
 intracellular predictive accuracy 23
 known host traits 23
 objective function 16
InterProScan 114
isopropyl-beta-d-1-thiogalactopyranoside (IPTG) 67
iterative model curation process 62, 63

k

key performance indicators (KPIs) 31
kinetic parameter estimation 30
k-nearest neighbor (KNN) 29, 105, 134, 229
Kyoto Encyclopedia of Genes and Genomes (KEGG) 3, 162, 342

l

large language models (LLMs) 175, 176, 202, 253
LeakGAN 201
leaks 10
library construction and cloning 348–349
LLM automated data extraction for machine learning 258–259

m

machine learning (ML) 133, 148, 220, 286
 algorithms for enzyme function prediction 136–138
 definition of 226
 extraction of enzyme features from amino acid sequences 134–136
 high-throughput microbial culturing AI in microbial cultivation process optimization 238–240

 challenges and limitations 241–242
 machine learning (ML)-driven microbial growth prediction 237–238
 synthetic biology and AI-guided strain engineering 240–241
high-throughput microbial identification 230
 genomic and metagenomic data processing 230–234
 imaging-based identification 236–237
 mass spectrometry-based identification 234–235
microbial identification 229–230
 supervised vs. unsupervised vs. reinforcement learning 226–229
machine learning (ML)-driven microbial growth prediction 237–238
machine learning for sequence-to-function approaches
 from BLAST to language models 105–106
 case studies 116–118
 challenges in sequence-to-function mapping 118–120
 enzyme activity 107–108
 enzyme commission (EC) number 106–107
 gene ontology 106
 protein solubility 109
 protein thermal stability 108
 protein toxicity 108–109
 tool-kits & benchmarks 111–115
MALDI Biotyper software 234
manual GSM creation 44–45
manual screening 8
masked language modeling (MLM) 82, 106, 114, 142, 144, 147
mass isotopomer distributions (MID) 324
mass spectrometry-based identification 234–235
matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS) 223
mechanism-agnostic, AI-driven approach 69
media and feed design 29–30

- MEMOSys (metabolic model research and development system) 58
- Merlin 51
- metabolic engineering 15, 26–27, 30, 240, 241, 267, 268, 285–287, 289, 290, 293, 295, 298, 306–308, 321, 345, 346
- ¹³C-metabolic flux analysis (MFA) 23
- metabolic flux analysis (MFA) 272, 287, 293, 321–333
- metabolic network 2
 - bottom-up approach 3–4
 - top-down approach 4–5
- metabolism 43
 - dynamic nature of 317–321
- metabolomics 263
 - artificial intelligence (AI)-based
 - metabolomics data processing and analysis 273–274
 - biotechnology 267
 - absolute metabolomics and thermodynamic analyses 272
 - evaluation and optimization of metabolic flux 272–273
 - identification of pathway bottlenecks 267–271
 - conceptual overview of future directions 274–277
 - detection and quantification 266–267
 - preparation of samples 264–266
 - metabolomics-guided strain engineering in various microorganisms 269
- MetaBoot 231, 232
- MetaCyc 5, 46, 51, 53, 57, 58, 162, 177, 342
- MetaCyc-based GSM 53
- MetaDraft 56
- MetaGEM 44, 50–51, 61
- metagenomics 219, 220, 223, 231
- meta learning 116, 120, 195
- MetExplore 58–59
- Michaelis-Menten kinetics 318
- MicrobesFlux 55
- microbial identification and culturing 219
- microfluidics and Lab-on-a-Chip (LOC) devices 224–225
- mixed-integer linear programming (MILP) 27
- ML-based approaches in microbial cell factory design
 - hybrid ML and constraint-based models 298–300
 - machine learning models 291–292
 - metabolic production 304–306
 - strain engineering and flux design 301–304
 - supervised machine learning 292–295
 - unsupervised machine learning 296–298
- model-agnostic meta-learning (MAML) 116
- model-guided sequence design using deep learning
 - CNNs in regulatory sequence analysis 70–75
 - generative sequence modeling 75–85
- ModelSEED 5, 46, 50–53, 58
- Modular Cloning (MoClo) technique 347
- monoclonal antibodies (mAbs) 306
- Monte Carlo Tree Search (MCTS) 167, 302
- motif analysis and deep learning 90–93
- MS2/OMEGGA on KBase 51
- multi-cellular models 25–26
- multi-layer modelling of Cellular Function 24
- multi-layer, multi-scale metabolic networks
 - integrating gene regulatory networks 24–25
 - integrating signalling networks 25
 - integrating transcription and translation 25
 - multi-cellular and multi-tissue models 25–26
 - multi-scale bioreactor models 26
- multiscale bioreactor models 26
- multitask learning (MTL) 307
- multitissue models 25–26
- n**
- native-context genome-derived datasets 86–87
- NEKO 253–256, 260
- NetSolP 109
- network validation functionality 6
 - additional reactions
 - biomass reactions 7
 - demand reactions 6
 - exchange reactions 6

- network validation functionality (*contd.*)
 - non-growth associated ATP maintenance reactions 7
 - spontaneous reactions 7
 - transport reactions 6
 - networks (biology) 1
 - neural networks (NNs) 30, 68, 69, 73–75, 77, 104, 139, 141, 144, 170, 177, 200, 229–231, 292–293, 320
 - NEXT-FBA 30
 - next-generation AI for retrobiosynthesis planning 175–176
 - next-generation metabolic flux analysis
 - dynamic nature of metabolism 317–321
 - flux balance analysis and metabolic flux analysis 321–324
 - machine learning 324
 - challenges in applying 325–326
 - dynamic flux and isotope labeling analysis 332–333
 - general workflow 328–331
 - improved speed and accuracy 331–332
 - next-generation peptide design through multi-agent systems 201–203
 - next-generation sequencing (NGS) 223, 317
 - non-dominated sorting genetic algorithm-II (NSGA-II) 303
 - non-growth associated ATP maintenance (NGAM) reactions 7
- o**
- objective function 9, 11, 12, 15, 16, 83, 294, 299, 322, 326, 330, 379, 382, 383, 386, 387, 396
 - omics datasets 16, 17, 347
 - Operational Genomic Unit (OGU)
 - method 231
 - OptKnock 27, 28, 344
 - oracle 70, 71, 80, 90
- p**
- PandoraGAN 201
 - PaperQA 202
 - parsimonious version of flux balance analysis (pFBA) 20, 27, 52, 295, 299
 - partial convolutional neural network (PCNN) 16, 331, 332
 - Partial Least Squares (PLS) 369
 - PATHGENN collection 62
 - PathwayTools 49, 57
 - PDH 27
 - PepBank 187–189
 - PepCVAE 200
 - Pepstats 134, 135
 - peptides, definitions and main characteristics 184–185
 - Peptipedia v2.0 187–189
 - PepVAE 200
 - Pfam 112, 114, 135, 136
 - position-specific iterated BLAST (PSI-BLAST) 132, 133
 - position-specific score matrix (PSSM) 132, 136
 - position-specific scoring matrices (PSSMs) 194
 - primer on CNNs for supervised genomic sequence modeling 73–75
 - principal component analysis (PCA) 29, 62, 148, 149, 229, 235, 289, 297, 299, 300, 305, 306, 369, 385
 - process monitoring and forecasting 30–31
 - PROFEAT 134, 135
 - PromoGen 75
 - ProtBERT 106, 144
 - protein engineering 118, 119, 177, 253, 302, 342, 348, 349, 356
 - ProteInfer 141, 151
 - protein language models (pLMs) 106, 112, 142–145, 150–152, 165, 170, 171, 190, 192, 193, 196–198, 206
 - protein sequence-to-function paradigm 104, 106, 115, 120
 - protein solubility 109
 - protein thermal stability 108
 - protein toxicity 108–109
 - protein–protein Interaction (PPI) networks 2, 5, 110, 144, 147, 149–151
 - ProtFun 138
 - ProtParam 134, 135, 148, 199
 - ProtTrans 106, 110, 145
 - PyAMPA 199
 - pyFBA 48, 53
- q**
- QbD approach 31

r

Random Forest (RF) 193, 195, 197, 222, 229, 235, 293, 371
 ranking of synthesis routes 168
 Raven 48, 53
 rBioNet 48, 54
 RBS Calculator 68, 69, 344
 RBS Designer 68
 reaction directionality 4, 17
 reaction stoichiometry 3, 8
 Reaxys database 302
 Reconstructor 47, 52, 62
 Rectified Linear Unit (ReLU) 73, 74
 recurrent neural networks (RNNs) 89, 105, 107, 109, 110, 141–142, 190, 193, 197, 333, 370, 385
 reinforcement learning (RL) 31, 93, 193, 203–206, 226–229, 238, 239, 243, 292, 303, 328, 371, 374, 387
 Reinforcement Learning from Human Feedback (RLHF) 93
 REKINDLE 320, 321
 RENAISSANCE 320, 321
 representation learning and dimensionality reduction 297–298
 representative metabolomics-related databases 275–277
 Retro* 167
 RetrobioCat 162, 164, 166, 168
 Retropath2.0 344
 reverse diffusion process 85
 RiboDetector 231
 Rxn4chemistry 164, 166, 168

s

SABIO-RK 162, 163, 177
 SAscore 164, 168
 SCScore 164, 168
 searching algorithm 162, 167
 SelenzymeRF 344
 self-supervision 293, 307
 SEVA plasmid system 344
 single-character tokenization 81
 siphons 10
 16S ribosomal RNA gene sequencing 219
 sources of sequence–function data for deep learning
 native-context genome-derived datasets 86–87

synthetic datasets 87–90
 Species METabolic ANALysis (SMETANA) framework 50
 spontaneous reactions 6, 7, 46, 318
 standardized benchmarks 114–115
 strain engineering 240–241, 253, 254, 256, 260, 267, 269, 275, 285–308, 342–346, 349, 351, 352
 StraPep 187–189
 stratification and data transformations 288–289
 structural validation 7–8, 205
 SuBliMinaL 49, 56–57
 supervised learning on genomic sequences 72–73
 supervised machine learning, 62 227, 292
 alternative predictive approaches 294
 decision trees (DTs) and ensembles 293
 neural networks (NNs) 292–293
 selected case studies 294–295
 supervised vs. unsupervised vs. reinforcement learning 226–229
 support vector machines (SVM) 29, 105, 107, 109, 111, 134, 136, 137, 148, 152, 170, 193–195, 227, 230, 232, 234, 235, 274, 275, 294, 299, 302, 304, 327, 376
 SynBio Q&A chatbot 253
 SynBioHub 67
 synthetic biology (SynBio) 67–94, 107, 115, 170, 221, 240–241, 243, 253–260, 263, 264, 267, 274, 341, 342, 346, 347, 350, 351, 355
 synthetic biology and AI-guided strain engineering 240
 synthetic datasets 86–90
 systems biology 1, 3, 5, 46, 51, 60, 171, 239

t

TATA box (“TATAAA”) 83, 90, 91
 template-based tools 161–166
 template-free tools 161, 166–167
 text mining using knowledge graph tools 254
 GraphRAG 256–258
 NEKO 254–256
t-distributed Stochastic Neighbor Embedding (t-SNE) 119, 298
 therapeutic peptides

- therapeutic peptides (*contd.*)
AI-agent for autonomous therapeutic peptide design 203–205
benefits and limitations of 185–186
computational design 186–187
data sources for peptide discovery 187–189
ML-based strategies 189–190
data-driven approaches 190–192
de novo design 199–201
peptide bioactivity classification 192–195
pharmacological profiles 198–199
toxicity classification models 195–198
next-generation peptide design through multi-agent systems 201–203
thermodynamically constrained Flux Balance Analysis (tfBA) 9
TLImmuno2 197
top-down approach, for network reconstruction 2, 4–5, 52
ToxinPred 109, 111, 196
transfer learning 166, 193, 194, 196, 197, 203, 233, 236, 242, 293, 307, 328
transformer and protein language models 142–145
transformer-based DNA models 80–84
transformer-based language models 69, 70, 76, 81, 200, 201
transport reactions 6, 52, 56
tumor-homing peptides (THPs) 195
- U**
UDSMProt 142, 149
uncertainty estimation 383
Uniform Manifold Approx. Projection (UMAP) 289, 298
UniProt 57, 103, 111, 112, 169, 177, 187
unsupervised learning, data preparation 76
unsupervised machine learning 62, 227
biclustering 297
clustering 296
representation learning and dimensionality reduction 297–298
USPTO database 161, 166, 167
UTR Designer 68
- V**
variance-weighted (σ^2) sum of squared residuals (VarSSR) 324, 325, 331
variational autoencoder (VAE)-based approaches 200
variational autoencoders (VAEs) 70, 174, 183, 186, 200, 289, 307
- W**
whole-genome sequencing technologies 131
WikiCrow 202
wound-healing peptides (WHPs) 195
- X**
Xpresso 86