

Glossary

PFT	(Process Flow Template) is a concept that helps to enforce best practices, and increase the re-usability within a given scientific domain
VO	(Virtual Organizations), in Grid computing, refers to a dynamic set of individual and/or institutions defined around a set of resource-sharing rules and conditions
SWMS	Scientific Workflow Management System
VL-e	(Virtual Laboratory for e-Science) is a Dutch research project with the aim to bridge the gap between the technology push of the high performance networking and the Grid and the application pull of a wide range of scientific experimental applications
tele-science	merges advanced solutions for remote instrumentation (via Telemicroscopy), distributed data computation and storage, and transparent access to federated databases of cell structure
MACS	(Material Analysis of Complex Surface) experiments try to identify and determine the elements that compose complex surfaces, regardless of the nature of the sample
DAG	(Directed Acyclic Graphs) is a directed graph with no directed cycles
MDS	(Globus Monitoring and Discovery System) is the information services component of the Globus Toolkit and provides information about the available resources on the Grid and their status
HEFT	(Earliest Finish Time) is equal to the Early Start Time of a given task plus the duration of this task
OWT	(Olingo Web Tool) is a tool to automate the generation of the PFT data structure
SOAP	(Simple Object Access Protocol) is a protocol specification for exchanging structured information in the implementation of Web Services in computer networks
VRS	(Virtual Resource System) a Java API used which provide a single access point to the Grid resources

WSDL	(Web Service Description Language) is an XML format for describing network services as a set of endpoints operating on messages containing either document-oriented or procedure-oriented information
HPC	(High Performance Computing) uses supercomputers and computer clusters to solve advanced computation problems
CPU	(Central processing unit)the part of a computer (a microprocessor chip) that does most of the data processing
MoC	(Models of Computing) is the definition of the set of allowable operations used in computation and their respective costs
MIR	myGrid information repository
SWIG	(Simplified Wrapper and Interface Generator)s a software development tool that connects programs written in C and C++ with a variety of high-level programming language
RPC	(Remote Procedure Call) is an inter-process communication that allows a computer program to cause a subroutine or procedure to execute in another address space (commonly on another computer on a shared network) without the programmer explicitly coding the details for this remote interaction
PCA	(Principal component analysis) involves a mathematical procedure that transforms a number of possibly correlated variables into a number of uncorrelated variables called principal components, related to the original variables by an orthogonal transformation
XML	(Extensible Markup Language) is a set of rules for encoding documents in machine-readable form
GridFTP	(Grid-enabled File Transport Protocol) is an extension of the standard File Transfer Protocol (FTP) for use with Grid computing
LFC	(Logical File Catalog) has been developed by LCG (http://lcg.web.cern.ch/LCG/) to resolve problems with the EDG Replica Manager File Catalog
WSRF	(Web Service Reference Framework)is a family of OASIS-published specifications for web services. Major contributors include the Globus Alliance and IBM
ASKALON	is a Grid application development and computing environment
AWGL	(Abstract Grid Workflow Language) for describing Grid workflow applications at a high level of abstraction
JXPL	is an XML-based Scripting Language for Workflow Execution in a Grid Environment
GWorkowDL	(Generic Workflow Description Language) is a generic description language for workflows in distributed environments. This software package contains the XML Schema as well as Java tools for creating, parsing, and editing GWorkflowDL documents
Karajan	is a workflow specification language and execution engine, being developed within the Java CoG Kit

- Pegasus** project encompasses a set of technologies that help workflow-based applications execute in a number of different environments including desktops, campus clusters, grids, and now clouds
- Triana** is an open source problem solving environment developed at Cardiff University that combines an intuitive visual interface with powerful data analysis tools
- Kepler** is dedicated to furthering and supporting the capabilities, use, and awareness of the free and open source, scientific workflow application
- ICENI** (Imperial College e-Science Networked Infrastructure) is a collection of grid middleware used for providing and coordinating grid services for e-Science applications
- VBrowser** (Virtual Resource Browser) is intended as single frontend to the Grid. This is the main frontend from the VL-e Toolkit and most users from the VL-e Toolkit will only use this Graphical User Interface to access their Grid resources

Index

- A**
Affinity benchmark, 92, 99–100
AIRs. *See* Ambiguous Interaction Restraints (AIRs)
Ambiguous Interaction Restraints (AIRs), 109
Annotation, 21, 38
Antifreeze proteins, 51, 129
- B**
Betalactam, 15–19, 21
Binding affinity, 99
Blind prediction, 3, 20, 25–27, 30, 37–38, 88, 93–95, 110, 128
- C**
CAPRI. *See* Critical Assessment of PRedicted Interactions (CAPRI)
Carbon alpha, 4, 14
CASP. *See* Critical Assessment of Structural Predictions (CASP)
CASTp. *See* Computed Atlas of Surface Topography of proteins (CASTp)
Cavity, 25, 26, 30, 36–37, 45, 52, 56–57, 59, 60, 73, 74, 82, 108, 114, 155
Ceftazidime, 17, 19
Cephalosporin drug, 18, 19
CHARMM forcefields, 111
Community-wide experiment, 95–97
Comparative analysis, 5, 20, 55–85, 119–128
Computational biology, 89
Computed Atlas of Surface Topography of proteins (CASTp), 55, 60, 63–65, 67, 70–72, 75–78
Conservative residues, 56, 61–62
ConSurf, 28–29, 55, 56, 61–63, 67, 68, 71, 73, 75–77
Critical Assessment of PRedicted Interactions (CAPRI), 87–100, 106, 109–111
Critical Assessment of Structural Predictions (CASP), 93, 106
Crosslinking experiments, 109
Cube representation, 90–91
Cubic diagonals, 28, 60
- D**
Delta-minus, 4
Delta-plus, 4
3-Dimensional (3D) grid, 26–28
3-Dimensional (3D) sites, 2, 6
3-Dimensional (3D) structure, 1–21, 61, 63, 106
Dissemination, 136–138, 140–143, 145, 146
Divergence entropy, 50
DOCK, 89
Docking simulation, 88, 100
Downhill proteins, 51, 62
- E**
Easy proteins, 70, 72–74, 80, 81
Effective atom, 42–44, 107
Empirical hydrophobicity distribution, 44–46
Epitope mapping, 109
E-science, 135–156
Executable operations, 146
Experimentation process, 138, 140

F

- False negative (FN), 57–59, 63, 65, 66, 111, 112
- False positive (FP), 57–59, 63, 65, 66, 72, 89, 90, 96, 111, 112
- False positive rate (FPR), 59, 62, 83, 113, 115
- Fast-folding proteins, 51, 107, 129
- Fast Fourier Transform (FFT) algorithms, 91, 110
- Flexibility, 4, 6, 97, 143, 152
- Flexible docking, 92, 95–97, 100
- F-measure, 57–58, 63–75, 78, 79, 111, 112, 114, 117–121
- FN. *See* False negative (FN)
- FOD. *See* Fuzzy oil drop (FOD)
- FP. *See* False positive (FP)
- FPR. *See* False positive rate (FPR)
- Fuzzy oil drop (FOD), 42, 50–52, 55, 56, 59, 62, 63, 66–69, 72, 74–79, 81–85, 106–109, 111–115, 120–124, 126–129

G

- Gauss function, 42–44, 48, 49, 62
- Geographically distributed computing resources, 137–138
- Geometric alignment, 110
- Geometric analysis, 55–56
- Geometric hashing algorithm, 90–92
- Glycine polar, 4

H

- HADDOCK. *See* High Ambiguity Driven biomolecular DOCKing (HADDOCK)
- Hard proteins, 73, 74, 81, 82
- H-D exchange, 109
- High Ambiguity Driven biomolecular DOCKing (HADDOCK), 92, 97, 106, 109, 115–120, 122, 126, 128
- Homodimer, 50, 100, 108, 111, 113–117, 123–125
- Hydrophobic aliphatic, 4
- Hydrophobic aromatic, 4, 5
- Hydrophobic core, 41–52, 69, 72, 77, 78, 81, 84, 107, 120–122, 124, 126, 129
- Hydrophobicity deficiency, 45, 48, 59, 78, 79, 84, 108, 112–114, 116, 124
- Hydrophobicity excess, 45, 48, 59, 79, 82, 84, 108, 111, 112, 114–116
- Hyper-spectral data cube, 138–139

I

- Idealized hydrophobicity distribution, 42–44, 62, 107, 114
- Individual therapy, 155–156
- Inside protein, 26–28
- In solvent, 27–28
- Integrated application, 139, 149
- Interoperability between different workflows, 141
- Isomerases, 7, 9, 11

K

- Knowledge mining, 55
- Knowledge sharing, 135–156
- Kullback-Leibler entropy, 47

L

- Lectins, 14–15
- Life-cycle, 136, 137, 139–142, 146, 154
- LIGSITE^{csc}, 27–29, 33–34, 37
- Lock-key, 114

M

- MACS. *See* Material Analysis of Complex Surface (MACS)
- Material Analysis of Complex Surface (MACS), 138, 139, 142, 146–148, 150, 151, 154
- Mathematical morphology, 26
- Matthews correlation coefficient (MCC), 57, 58, 63–79, 111, 112, 117–122
- MCC. *See* Matthews correlation coefficient (MCC)
- MetaPocket, 27, 30–38
- Molecular dynamics, 41, 92, 96
- Monte-Carlo, 91, 92, 96
- Monte Carlo-based modeling package, 110
- Multivariate data analysis, 150
- Mutagenesis, 38, 109, 110

N

- Near surface, 26–28

O

- Objects, 3–5, 7–11, 14, 15, 19, 20, 56–57, 146, 147, 149
- Observed hydrophobicity distribution, 44–47, 49–51, 59, 62, 66, 82, 107
- Oil drop, 42, 48, 107

P

Pair-wise interaction, 44, 114, 116
Pegasus, 143–145, 148, 163
Penicillin drug, 19
PFT. *See* Process flow template (PFT)
Physico-chemical groups, 3
Pocket, 2, 25–38, 48, 51, 55, 60, 61, 69, 70, 77, 78, 81, 82, 84, 85, 121, 123
Pocket-finder, 60–61, 64, 65, 68, 71, 75–78
Pocket identification, 25–38, 70
Precision, 57–58
Preprocessing phase, 138, 139
Process flow template (PFT), 137, 145–151, 155, 161
Proteases, 7–14, 20, 89, 90, 93, 99
Protein-ligand binding sites, 2, 6, 7, 25–38, 60, 62–85
Protein-protein (P-P) interaction, 21, 45, 48, 49, 87–100, 106, 109, 125, 128
Protein-protein recognition, 88, 95–96, 156
Proteins, 1–21, 25–38, 41–52, 55–85, 87–100, 105–129, 155–156
 engineering, 98
 flexibility, 97
Proxy-service, 151–153

Q

Q-SiteFinder, 26, 27, 30, 34, 35, 37, 63, 65, 66, 68, 72, 75–78

R

Random distribution, 46, 47, 52
Recall, 57–58
Receiver operating characteristic (ROC), 57–59, 62, 78, 83, 112, 113, 115
Reproducibility of results, 141
Residues conservation, 27–29
Rigid body docking, 90–91, 94, 96, 106
ROC. *See* Receiver operating characteristic (ROC)
ROC curve, 58–59, 62, 78, 112, 113, 115
RosettaDock, 91, 92, 106–107, 109–110, 117–118, 120–122, 128
Rotamer packing, 109

S

Searching workflows, 141
Sensitivity, 58, 59

Sequence conservation, 62, 106, 109
Serine proteases, 7, 8, 10, 11, 14, 20
Shared repository, 146
Side-chain rotamer probabilities, 110
Single nucleotide polymorphism (SNP), 155
SNP. *See* Single nucleotide polymorphism (SNP)
Solvated docking, 109
Storage resources, 137, 152
Structural bioinformatics, 2
Structural discordance, 47
SuMo. *See* Surfing the Molecules (SuMo)
Surfing the Molecules (SuMo), 1–21, 55, 56, 63, 66, 69, 72, 75, 77, 78

T

Target protein, 60, 70–75, 90, 93–96, 98, 100, 109, 110, 112
Target structure, 6
Template-based docking, 92–93
Tertiary structure stabilization, 41
Theoretical hydrophobicity distribution, 42–44, 49
TN. *See* True negative (TN)
TP. *See* True positive (TP)
TPR. *See* True positive rate (TPR)
Triangle form, 3
True negative (TN), 57–59, 111, 112
True positive (TP), 57–59, 63, 111, 112
True positive rate (TPR), 59, 62, 83, 113, 115

U

Uniform interface, 137

V

Virtual data system, 145
Virtual organization (VO), 136, 138, 147, 161

W

Workflow, 29–31, 106, 135–156
Workflow support, 135–156

Z

ZDOCK, 92, 106, 110–111, 118–120, 122, 128