Biochemical Cascade

A biochemical cascade is a series of chemical reactions in which the products of one reaction are consumed in the next reaction. These cascades facilitate the transformation or generation of complex molecules in small steps. At each step, various controlling factors are involved to regulate cellular reactions, responding effectively to cues about their changing internal and external environments. Chemical reactions are orchestrated by complex molecular networks consisting of entities such as proteins/enzymes or RNAs (second messengers) connected by activation or synthesis in biological processes.

In biochemistry, several important enzymatic cascades and signal transduction cascades participate in metabolic pathways or signalling networks, in which enzymes are usually involved to catalyze the reactions. For example, the tissue factor pathway in the coagulation cascade of secondary hemostasis is the primary pathway lead to fibrin formation, therefore, the initiation of blood coagulation. The pathways are a series of reactions, in which a zymogen (inactive enzyme precursor) of a serine protease and its glycoprotein co-factors are activated to become active components that then catalyze the next reaction in the cascade, ultimately resulting in cross-linked fibrin. Another example, sonic hedgehog signaling pathway is one of the key regulators of embryonic development and is present in all bilaterians. Different parts of the embryo have different concentrations of hedgehog signaling proteins, which give cells information to make the embryo develop properly and correctly into a head or a tail. When the pathway malfunctions, it can result in diseases like basal cell carcinoma. Recent studies point to the role of hedgehog signaling in regulating adult stem cells involved in maintenance and regeneration of adult tissues. The pathway has also been implicated in the development of some cancers. Drugs that specifically target hedgehog signaling to fight diseases are being actively developed by a number of pharmaceutical companies.

In the post-genomic age, high-throughput sequencing and gene/protein profiling techniques have transformed biological research by enabling comprehensive monitoring of a biological system, yielding a list of differentially expressed genes or proteins, which is useful in identifying genes that may have roles in a given phenomenon or phenotype. With DNA microarrays, RNA interference, and genome-wide gene engineering, it is possible to screen global gene expression profiles to contributed a wealth of genomic data to the public domain, and to distill the inferences contained in the experimental literature and primary databases into knowledge bases that consist of annotated representations of biological pathways, including biological processes, components, or structures in which individual genes and proteins are known to be involved in, as well as how and where gene products interact with each other. Pathway-oriented approaches for analyzing microarray data, by grouping long lists of individual genes, proteins, and/or other biological molecules

according to the pathways they are involved in into smaller sets of related genes or proteins, which reduces the complexity, have proven useful for connecting genomic data to specific biological processes and systems. Identifying active pathways that differ between two conditions can have more explanatory power than a simple list of different genes or proteins. Pathway analysis has been applied to the analysis of Gene Ontology (GO) terms (also referred to as a "gene set"), physical interaction networks (e.g., protein–protein interactions), kinetic simulation of pathways, steady-state pathway analysis (e.g., flux-balance analysis), and in the inference of pathways from expression and sequence data in public repositories such as GO or Kyoto Encyclopedia of Genes and Genomes (KEGG). The existing knowledge base–driven pathway analysis methods in each generation are summarised in Table 1.

The increasing amount of genomic and molecular information is the basis for understanding higher-order biological systems, such as the cell and the organism, and their interactions with the environment, as well as for medical, industrial and other practical applications. The KEGG resource (http://www.genome.jp/kegg/) provides a reference knowledge base for linking genomes to biological systems, categorized as building blocks in the genomic space (KEGG GENES), the chemical space (KEGG LIGAND), wiring diagrams of interaction networks and reaction networks (KEGG PATHWAY), and ontologies for pathway reconstruction (BRITE database), as illustrated in Figure 1. The KEGG PATHWAY database is a collection of manually drawn pathway maps for metabolism, genetic information processing, environmental information processing such as signal transduction, ligand–receptor interaction and cell communication, various other cellular processes and human diseases, all based on extensive survey of published literature.

Gene Map Annotator and Pathway Profiler (GenMAPP) as a free, open-source, stand-alone computer program is designed for organizing, analyzing, and sharing genome scale data in the context of biological pathways. GenMAPP database support multiple gene annotations and species as well as custom species database creation for a potentially unlimited number of species. Pathway resources are expanded by utilizing homology information to translate pathway content between species and extending existing pathways with data derived from conserved protein interactions and coexpression. A new mode of data visualization including time-course, single nucleotide polymorphism (SNP), and splicing, has been implemented with GenMAPP database to support analysis of complex data. GenMAPP also offers innovative ways to display and share data by incorporating HTML export of analyses for entire sets of pathways as organized web pages. In short, GenMAPP provides a means to rapidly interrogate complex experimental data for pathway-level changes in a diverse range of organisms.

Given the genetic makeup of an organism, the complete set of possible reactions constitutes its reactome. Reactome, located at http://www.reactome.org is a curated, peer-reviewed resource of human biological processes/pathway data. The basic unit of the Reactome database is a reaction; reactions are then grouped into causal chains to form pathways. The Reactome data model allows us to represent many diverse processes in the human system, including the pathways of intermediary metabolism, regulatory pathways, and signal transduction, and high-level processes, such as the cell cycle. Reactome provides a qualitative framework, on which quantitative data can be superimposed. Tools have been developed to facilitate custom data entry and annotation by expert biologists, and to allow visualization and exploration of the finished dataset as an interactive process map. Although the primary curational domain is pathways from Homo sapiens, electronic projections of human pathways onto other organisms are regularly created via putative orthologs, thus making Reactome relevant to model organism research communities. The database is publicly available under open source terms, which allows both its content and its software infrastructure to be freely used and redistributed. Studying whole transcriptional profiles and cataloging protein–protein interactions has yielded much valuable biological information, from the genome or proteome to the physiology of an organism, an organ, a tissue or even a single cell. The Reactome database containing a framework of possible reactions which, when combined with expression and enzyme kinetic data, provides the infrastructure for quantitative models, therefore, an integrated view of biological processes, which links such gene products and can be systematically mined by using bioinformatics applications. Reactome data available in a variety of standard formats, including BioPAX, SBML and PSI-MI, and also enable data exchange with other pathway databases, such as the Cycs, KEGG and amaze, and molecular interaction databases, such as BIND and HPRD. The next data release will cover apoptosis, including the death receptor signaling pathways, and the Bcl2 pathways, as well as pathways involved in hemostasis. Other topics currently under development include several signaling pathways, mitosis, visual phototransduction and hematopoeisis. In summary, Reactome provides high-quality curated summaries of fundamental biological processes in humans in a form of biologist-friendly visualization of pathways data, and is an open-source project.

Pathway building has been performed by individual groups studying a network of interest (e.g., immune signaling pathway) as well as by large bioinformatics consortia (e.g., the Reactome Project) and commercial entities (e.g., Ingenuity Systems). Pathway building is the process of identifying and integrating the entities, interactions, and associated annotations, and populating the knowledge base. Pathway construction can have either a data-driven objective (DDO) or a knowledge-driven objective (KDO). Data-driven pathway construction is used to generate relationship information of genes or proteins identified in a specific experiment such as a microarray study. Knowledge-driven pathway construction entails development of a detailed pathway knowledge base for particular domains of

interest, such as a cell type, disease, or system. The curation process of a biological pathway entails identifying and structuring content, mining information manually and/or computationally, and assembling a knowledgebase using appropriate software tools. A schematic illustrating the major steps involved in the data-driven and knowledge-driven construction processes is shown in Figure 2. For either DDO or KDO pathway construction, the first step is to mine pertinent information from relevant information sources about the entities and interactions. The information retrieved is assembled using appropriate formats, information standards, and pathway building tools to obtain a pathway prototype. The pathway is further refined to include context-specific annotations such as species, cell/tissue type, or disease type. The pathway can then be verified by the domain experts and updated by the curators based on appropriate feedback. Recent attempts to improve knowledge integration have led to refined classifications of cellular entities, such as Gene Ontology (GO), and to the assembly of structured knowledge repositories. Data repositories, which contain information regarding sequence data, metabolism, signaling, reactions, and interactions are a major source of information for pathway building. A few useful databases are described in Table 2. A comprehensive list of resources can be found at http://www.pathguide.org.

Application

1. Colorectal cancer (CRC)

A program package MatchMiner was used to scan HUGO names for cloned genes of interest cloned are scanned, then are input into GoMiner (online at http://genomebiology.com/2003/4/4/R28), which leveraged the Gene Ontology (GO) to identify the biological processes, functions and components represented in the gene profile. Besides, Database for Annotation, Visualization, and Integrated Discovery (DAVID) (http://genomebiology.com/2003/4/9/R60) and KEGG (Kyoto Encyclopedia of Genes and Genomes) database (http://www.genome.ad.jp/kegg/) are used for the analysis of microarray expression data and the analysis of each GO biological process (P), cellular component (C), and molecular function (F) ontology. In addition, DAVID tools are used to analyze the roles of genes in metabolic pathways and show the biological relationships between genes or gene-products and may represent metabolic pathways. These two databases also provide bioinformatics tools online to combine specific biochemical information on a certain organism and facilitate the interpretation of biological meanings for experimental data. By using a combined approach of Microarray-Bioinformatic technologies, a potential metabolic mechanism contributing to colorectal cancer (CRC) have been demonstrated. Several environmental factors may be involved in a series of points along the genetic pathway to colorectal cancer. These include genes associated with bile acid metabolism, glycolysis metabolism and fatty acid metabolism pathways, supporting a hypothesis that some metabolic alternations observed in colon carcinoma may occur in the development of CRC.

2. Parkinson’s disease.

Cellular models are instrumental in dissecting a complex pathological process into simpler molecular events. Parkinson’s disease is multifactorial and clinically heterogeneous; the aetiology of the sporadic (and most common) form is still unclear and only a few molecular mechanisms have been clarified so far in the neurodegenerative cascade. In such a multifaceted picture, it is particularly important to identify experimental models that simplify the study of the different networks of proteins / genes involved. Cellular models that reproduce some of the features of the neurons that degenerate in Parkinson’s disease have contributed to many advances in our comprehension of the pathogenic flow of the disease. In particular, the pivotal biochemical pathways (i.e. apoptosis and oxidative stress, mitochondrial impairment and dysfunctional mitophagy, unfolded protein stress and improper removal of misfolded proteins) have been widely explored in cell lines, challenged with toxic insults or genetically modified. The central role of a-synuclein has generated many models aiming to elucidate its contribution to the dysregulation of various cellular processes. Classical cellular models appear to be the correct choice for preliminary studies on the molecular action of new drugs or potential toxins and for understanding the role of single genetic factors. Moreover, the availability of novel cellular systems, such as cybrids or induced pluripotent stem cells, offers the chance to exploit the advantages of an in vitro investigation, although mirroring more closely the cell population being affected.

3. Alzheimer's diseases.

Synaptic degeneration and death of nerve cells are defining features of Alzheimer’s disease (AD), the most prevalent age-related neurodegenerative disorders. In AD, neurons in the hippocampus and basal forebrain (brain regions that subserve learning and memory functions) are selectively vulnerable. Studies of postmortem brain tissue from AD people have provided evidence for increased levels of oxidative stress, mitochondrial dysfunction and impaired glucose uptake in vulnerable neuronal populations. Studies of animal and cell culture models of AD suggest that increased levels of oxidative stress (membrane lipid peroxidation, in particular) may disrupt neuronal energy metabolism and ion homeostasis, by impairing the function of membrane ion-motive ATPases and glucose and glutamate transporters. Such oxidative and metabolic compromise may thereby render neurons vulnerable to excitotoxicity and apoptosis. Recent studies suggest that AD can manifest systemic alterations in energy metabolism (e.g., increased insulin resistance and dysregulation of glucose metabolism). Emerging evidence that dietary restriction can forestall the development of AD is consistent with a major “metabolic” component to these disorders, and provides optimism that these devastating brain disorders of aging may be largely preventable.

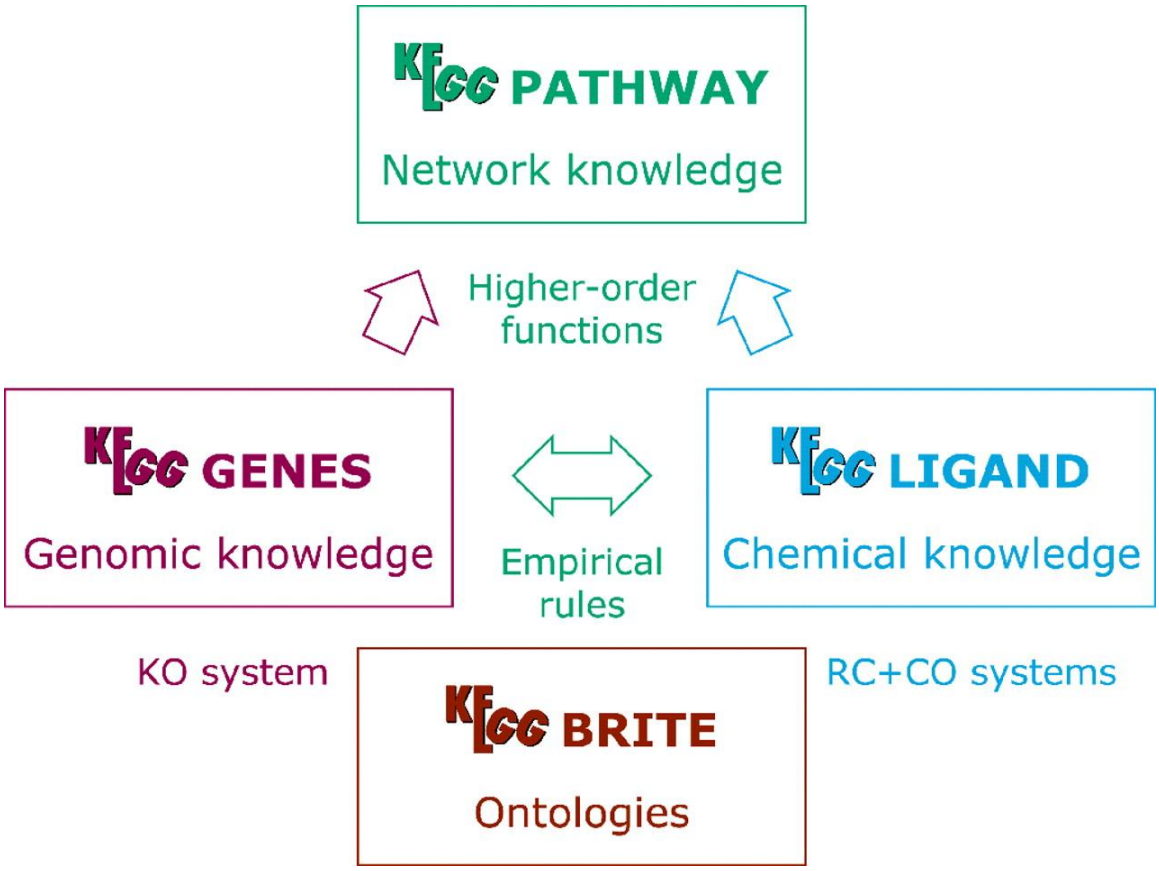

Figure 1. The overall architecture of KEGG consisting of four main components.

(Kanehisa M et al. Nucl. Acids Res. 2006;34:D354-D357)

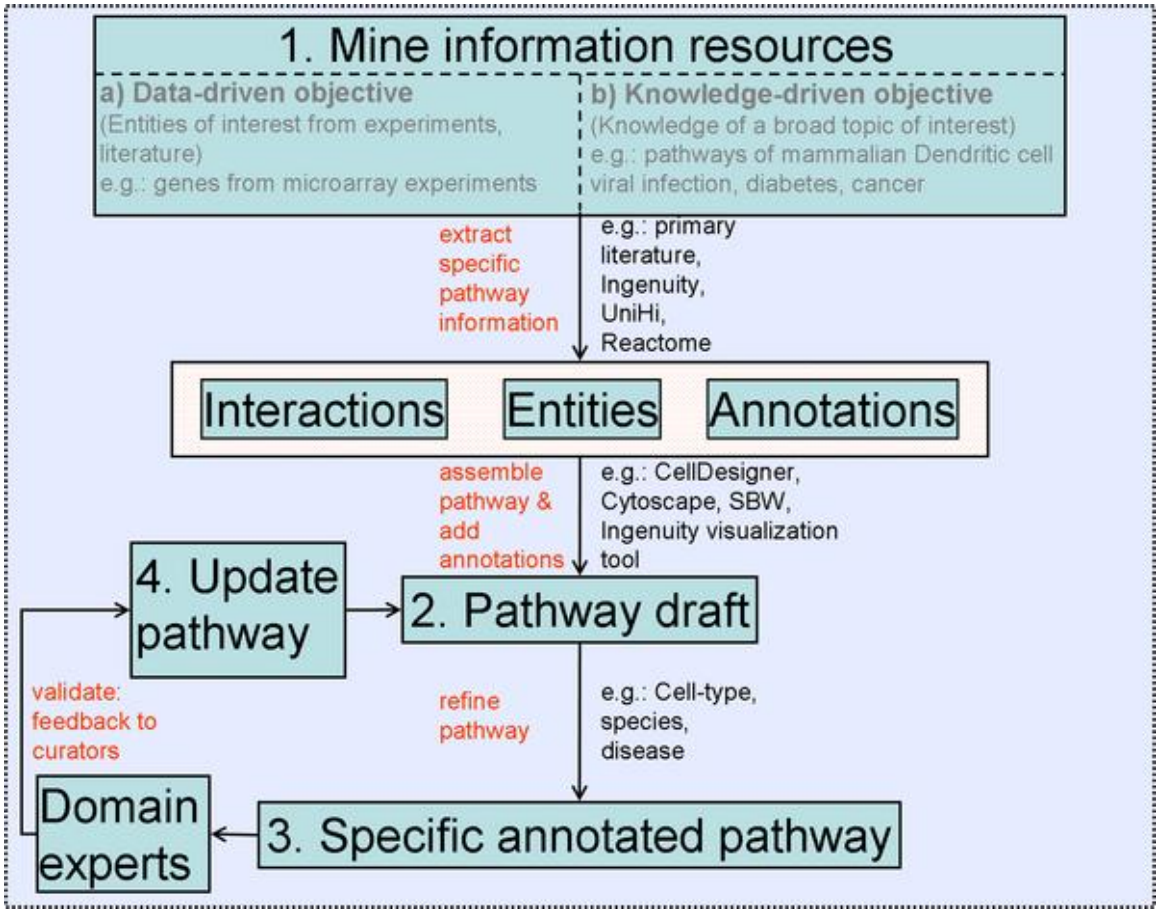

Figure 2. Schematic Illustrating the Biological Pathway Building Process

Pathway curators initially mine information (Step 1). The mining process can be initiated by two broad pathway building objectives: (a) DDO wherein a list of genes and/or proteins are obtained by high-throughput experiments such as microarray, mass spectrometry or (b) KDO wherein a broad topic of interest is chosen and then the knowledge concerning this topic is mined from resources such as the primary literature and knowledgebases. Information from the mining process is assembled (Step 2), using pathway building tools, into a pathway, which, following many iterations of feedback from domain experts (Step 3) and refinement (Step 4), leads to the desired specific annotated pathway.

(Viswanathan et al. PLoS Comput Biol 2008; 4(2): e16)

Table 1. Examples of pathway analysis tools in each generation.

Name	Availability
ORA tools	
Onto-Express	Web (http://vortex.cs.wayne.edu)
GenMAPP	Standalone (http://www.genmapp.org)
GoMiner	Standalone, Web (http://discover.nci.nih.gov/gominer)
FatiGO	Web (http://babelomics.bioinfo.cipf.es)
GOstat	Web (http://gostat.wehi.edu.au)
FuncAssociate	Web (http://llama.mshri.on.ca/funcassociate/)
GOToolBox	Web (http://genome.crg.es/GOToolBox/)
GeneMerge	Standalone, Web (http://genemerge.cbcb.umd.edu/)
GOEAST	Web (http://omicslab.genetics.ac.cn/GOEAST/)
ClueGO	Standalone (http://www.ici.upmc.fr/cluego/)
FunSpec	Web (http://funspec.med.utoronto.ca/)
GARBAN	Web
GO:TermFinder	Standalone (http://search.cpan.org/dist/GO-TermFinder/)
WebGestalt	Web (http://bioinfo.vanderbilt.edu/webgestalt/)
agriGO	Web (http://bioinfo.cau.edu.cn/agriGO/)
GOFFA	Standalone, Web (http://edkb.fda.gov/webstart/arraytrack/)
WEGO	Web (http://wego.genomics.org.cn/cgi-bin/wego/index.pl)
FCS tools	
GSEA	Standalone (http://www.broadinstitute.org/gsea/)
sigPathway	Standalone (BioConductor)
Category	Standalone (BioConductor)
SAFE	Standalone (BioConductor)
GlobalTest	Standalone (BioConductor)
PCOT2	Standalone (BioConductor)
SAM-GS	Standalone (http://www.ualberta.ca/~yyasui/software.html)
Catmap	Standalone (http://bioinfo.thep.lu.se/catmap.html)
T-profiler	Web (http://www.t-profiler.org)
FunCluster	Standalone (http://corneliu.henegar.info/FunCluster.htm)
GeneTrail	Web (http://genetrail.bioinf.uni-sb.de)
GAzer	Web
PT-based tools	
ScorePAGE	No implementation available
Pathway-Express	Web (http://vortex.cs.wayne.edu)
SPIA	Standalone (BioConductor)
NetGSA	No implementation available

doi:10.1371/journal.pcbi.1002375.t001

(Khatri et al. PLoS Comput. Biol. 2012; 8(2): e1002375.)

Table 2: A list of frequently used databases, classified based on the type of information represented, during a biological pathway construction, their properties and URL.

Database	Curation type	GO Annotation (Y/N)	Description	URL
1. Protein-protein interactions databases				
BIND	M	N	200,000 documented biomolecular interactions and complexes	http://www.bind.ca/
MINT	M	N	Experimentally verified interactions	http://mint.bio.uniroma2.it/mint/
HPRD	M	N	Elegant and comprehensive presentation of the interactions, entities and evidences	http://www.hprd.org/
MPact	B	N	Yeast interactions. A part of MIPS	http://mips.gsf.de/genre/proj/mpact/
DIP	B	Y	Experimentally determined interactions	http://dip.doe-mci.ucla.edu/
IntAct	M	Y	Database and analysis system of binary and multi-protein interactions	http://www.ebi.ac.uk/intact
PDZBase	M	N	PDZ Domain containing proteins	http://icb.med.cornell.edu/services/pdz/start/
GNPV	B	Y	Based on specific experiments and literature	http://genomenetwork/nig.ac.jp/
BioGrid	M	Y	Physical and genetic interactions	http://thebiogrid.org/
UniHi	B	Y	Comprehensive human protein interactions	http://theoderich.fb3.mdc-berlin.de:8080/unihi/
OPHID	B	Y	Combines PPI from BIND,	http://ophid.utoronto.ca/ophid/

			HPRD, and MINT	
2. Metabolic Pathway databases				
EcoCyc	B	Y	Entire genome and biochemical machinery of E. Coli	http://ecocyc.org
MetaCyc	M	N	Pathways of over 165 species	http://metacyc.org
HumanCyc	B	N	Human metabolic pathways and the human genome	http://humancyc.org
BioCyc	B	N	Collection of databases for several organism	http://biocyc.org
3. Signaling Pathway databases				
KEGG	M	Y	Comprehensive collection of pathways such as human disease, signaling, genetic information processing pathways. Links to several useful databases.	http://www.genome.ad.jp/kegg/
PANTHER	M	N	Compendium of metabolic and signaling pathways built using CellDesigner. Pathways can be downloaded in SBML format.	http://panther.appliedbiosystems.com/
Reactome	M	Y	Hierarchical layout. Extensive links to relevant databases such as NCBI, ENSEMBL, UNIPROT, HAPMAP, KEGG, CHEBI, PubMed, GO. Follows PSI-MI standards.	http://www.reactome.org/
Biomodels	M	Y	Domain experts curated biological connection maps and associated mathematical models	http://www.ebi.ac.uk/biomodels/
STKE	M	N	Repository of canonical pathways.	http://stke.sciencemag.org/cm/

Ingenuity Systems	M	Y	Commercial mammalian biological knowledgebase about genes, drugs, chemical, cellular and disease processes, and signaling and metabolic pathways.	http://ingenuity.com/
PID	M	Y	Compendium of several highly structured, assembled signaling pathways	http://pid.nic.nih.gov/PID/
BioPP	B	Y	Repository of biological pathways built using CellDesigner	http://tsb.mssm.edu/pathwayPublisher/boradcast/

Legend: M – Manual curation, A – Automated curation, B- Both manual and automated curation; Y – Yes, N – No; BIND – Biomolecular Interaction Network Database, DIP – Database of Interacting Proteins, GNPV – Genome Network Platform Viewer, HPRD = Human Protein Reference Database, MINT – Molecular INTeration database, MIPS – Munich Information center for Protein Sequences, UNIHI – Unified Human Interactome, OPHID – Online Predicted Human Interaction Database, EcoCyc – Encyclopaedia of E. Coli Genes and Metabolism, MetaCyc – aMetabolic Pathway database, KEGG – Kyoto Encyclopedia of Genes and Genomes, PANTHER – Protein Analysis Through Evolutionary Relationship database, STKE – Signal Transduction Knowledge Environment, PID – The Pathway Interaction Database, BioPP – Biological Pathway Publisher.