



KEGG BRITE: a renewed addition to KEGG for managing and using functional hierarchies of biological systems

Susumu Goto
Bioinformatics Center
Kyoto University





Contents

- Why BRITE?
- What is BRITE?
- How to use BRITE?
- What is next?






Why BRITE?

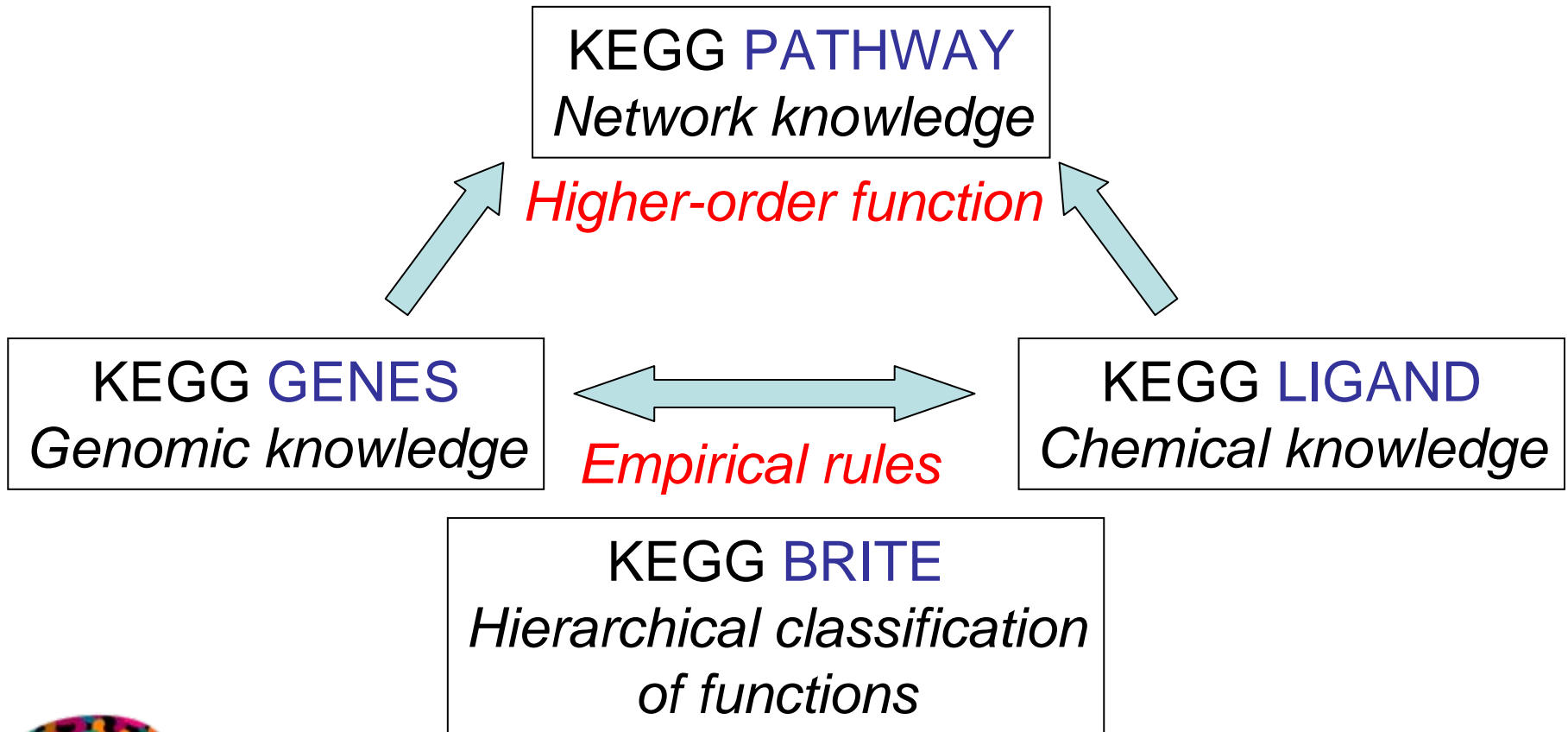
- Necessary for
 - Function annotation of genes/proteins
 - Classification and management of the function
 - Knowledge extraction
 - Integration of databases





Inferring biological systems from genomes

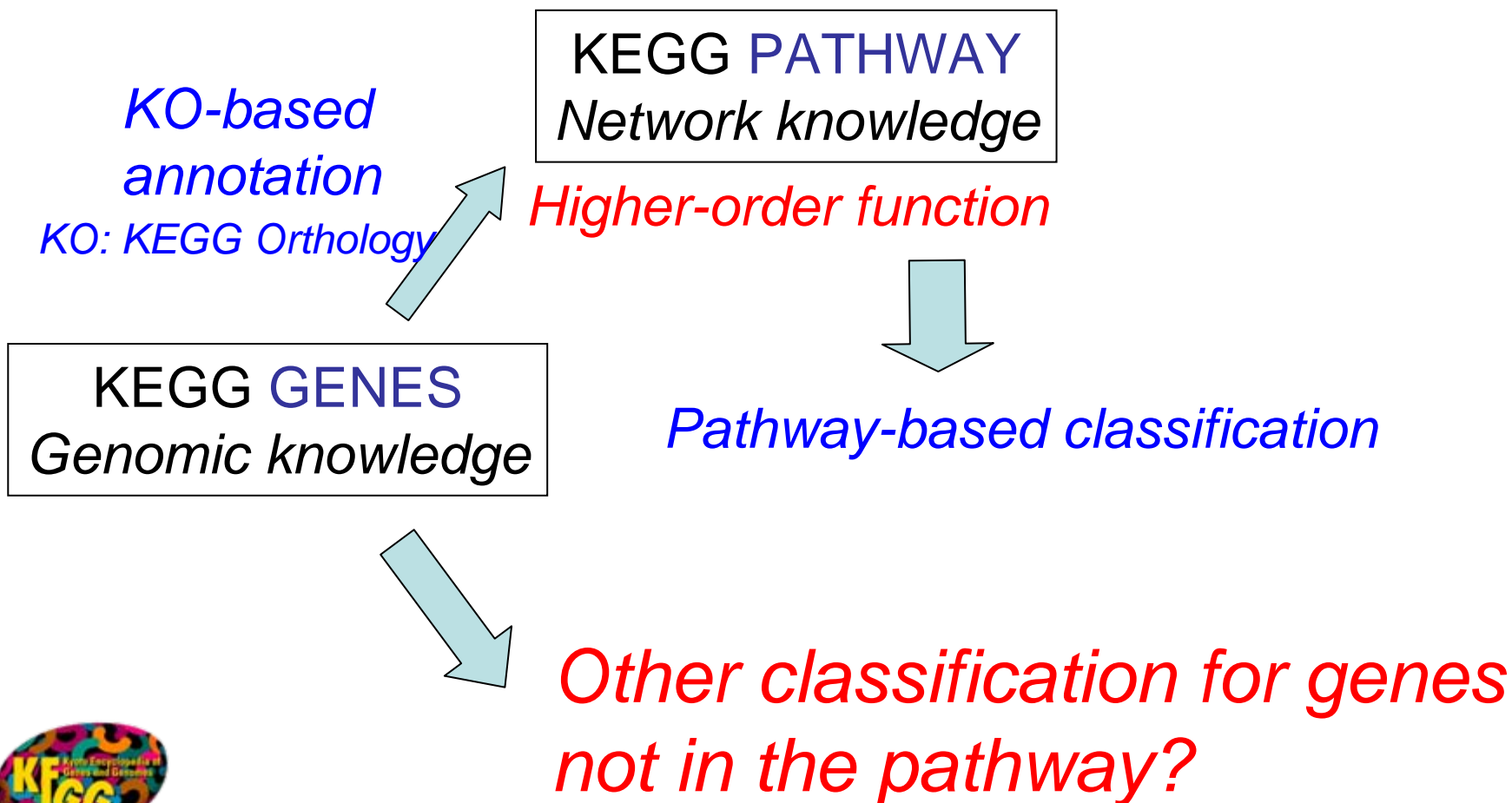
Kyoto Encyclopedia of Gene and Genomes





Function annotation

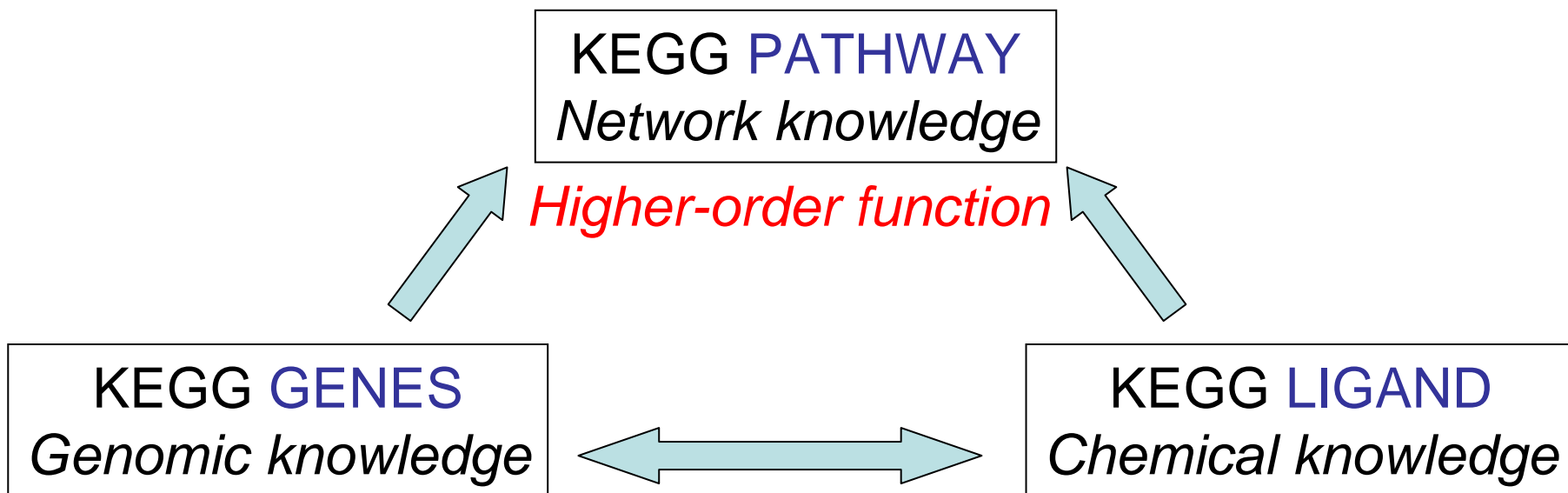
Kyoto Encyclopedia of Gene and Genomes





Classification and management of the function

Kyoto Encyclopedia of Gene and Genomes



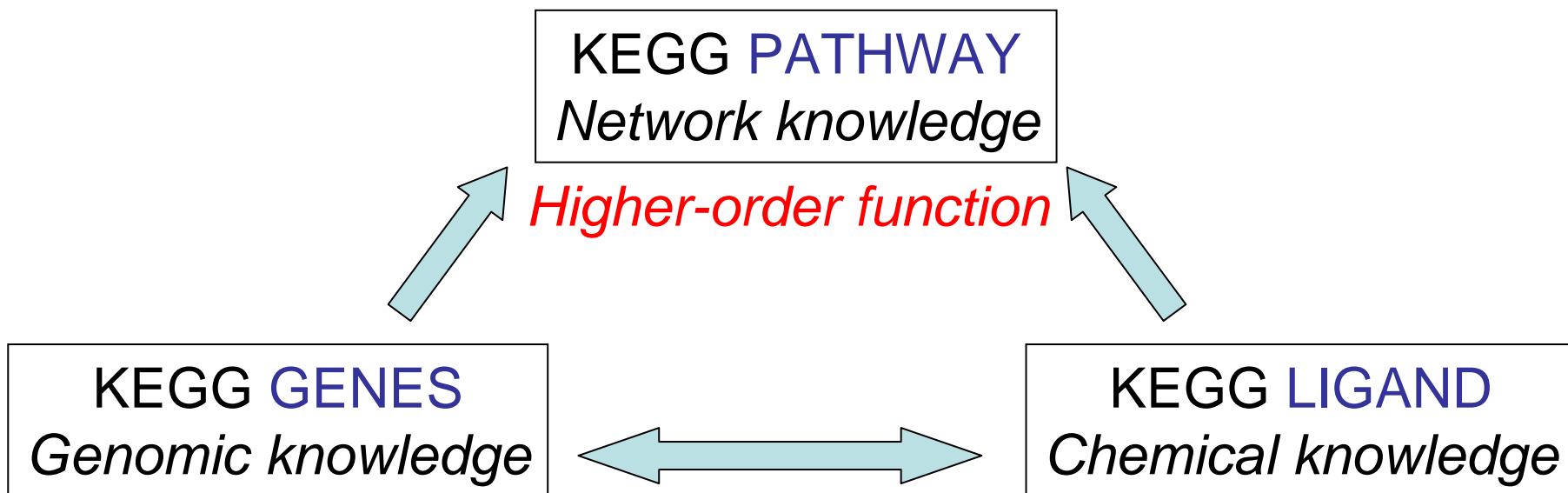
Protein-compound interaction
Relationship between molecules and diseases





Knowledge extraction

Kyoto Encyclopedia of Gene and Genomes



Reaction pattern and functions
Function prediction and pathway prediction





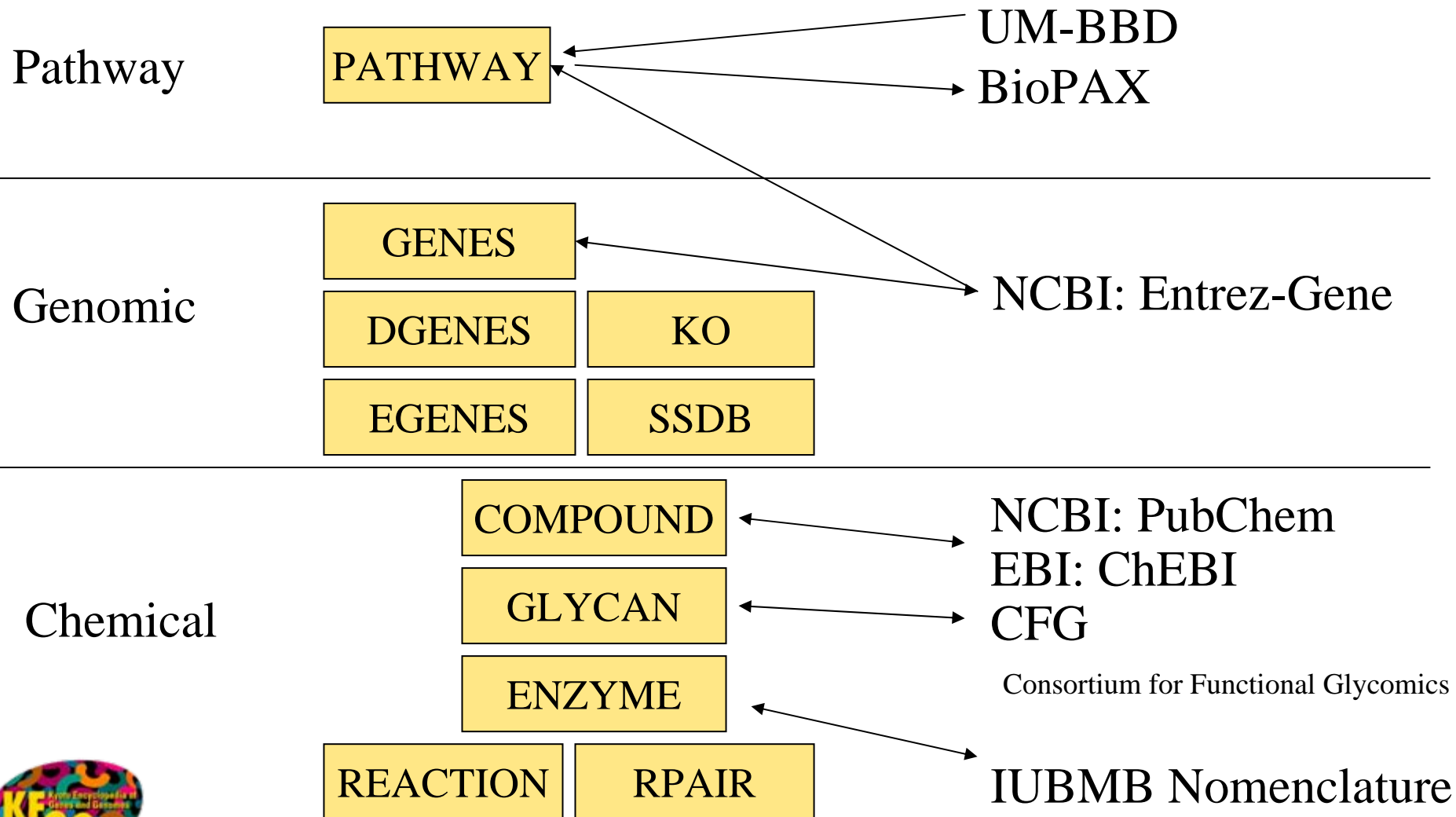
Integration of databases

- Mapping protein and chemical names among databases is important
 - Many databases for
 - Pathway information
 - Genomic information
 - Other omics information
 - How to integrate them?
 - Controlled vocabulary
 - Corresponding table between protein names





KEGG in the database collaboration





Contents

- Why BRITE?
- What is BRITE?
- How to use BRITE?
- What is next?





BRITE

- Hierarchical classification of functions
 - Proteins
 - Compounds
 - Others
- Database cross-references
 - KEGG GENES ID \leftrightarrow NCBI-GI, GeneID

<http://www.genome.jp/kegg/brite/>





Hierarchical classifications of functions

- Genomic knowledge (genes & proteins)
 - Network hierarchy
 - Protein families
- Chemical knowledge (compounds & reactions)
 - Compounds
 - Compound interactions
- Drugs and diseases
 - Drugs
 - Diseases
- Cells & organisms
 - Organisms





Classification of genomic knowledge

Functional Hierarchies

● Network hierarchy

KO - Pathway-based classification of orthologs

● Protein families

Enzymes
Cytochrome P450
Transcription factors
Ribosome
Translation factors
ABC transporters
G-protein coupled receptors
Ion channels
Cytokines
Cytokine receptors
Cell adhesion molecules (CAMs)
CAM ligands
CD molecules
Bacterial motility proteins

- **KO: KEGG Orthology**
 - Unique representation (controlled vocabulary) of gene function for all species
 - ~ 7,800 entries
 - Used for pathway reconstruction





Pathway-based ortholog classification

KEGG Orthology (KO)

[[1st Level](#) | [2nd Level](#) | [3rd Level](#) | [4th Level](#) | [Text Search](#)]

01100 Metabolism

01110 Carbohydrate Metabolism

00010 Glycolysis / Gluconeogenesis [PATH:ot00010 tab00010] [GO:0006096 0006094]

K00845 E2.7.1.2, glk; glucokinase [EC:2.7.1.2] [COG:COG0837] [GO:0004340]

K00844 E2.7.1.1; hexokinase [EC:2.7.1.1] [GO:0004396]

K01084 E3.1.3.9, G6PC; glucose-6-phosphatase [EC:3.1.3.9] [GO:0004346]

K01810 E5.3.1.9, pgi; glucose-6-phosphate isomerase [EC:5.3.1.9] [COG:COG0166] [GO:0004347]

K00850 E2.7.1.11, pfk; 6-phosphofructokinase [EC:2.7.1.11] [COG:COG0205 COG1105] [GO:0003872]

K01086 E3.1.3.11; fructose-1,6-bisphosphatase [EC:3.1.3.11] [GO:0042132]

K03841 FBP1, fbp; fructose-1,6-bisphosphatase I [EC:3.1.3.11] [COG:COG0158] [GO:0042132]

K02446 GLPX; fructose-1,6-bisphosphatase II [EC:3.1.3.11] [COG:COG1494] [GO:0042132]

K04041 FBP3, fbp; fructose-1,6-bisphosphatase III [EC:3.1.3.11] [GO:0042132]

K01622 E4.1.2.13; fructose-bisphosphate aldolase [EC:4.1.2.13] [GO:0004332]

K01624 E4.1.2.13B, fbaA; fructose-bisphosphate aldolase, class II [EC:4.1.2.13] [COG:COG0191] [GO:0004332]

K01623 E4.1.2.13A, fbaB; fructose-bisphosphate aldolase, class I [EC:4.1.2.13] [COG:COG1830 COG3588] [GO:0004332]

K01803 E5.3.1.1, tpiA; triosephosphate isomerase (TIM) [EC:5.3.1.1] [COG:COG0149] [GO:0004807]

K00134 E1.2.1.12, GAPD, gapA; glyceraldehyde 3-phosphate dehydrogenase [EC:1.2.1.12] [COG:COG0057] [GO:0004365]

K00927 E2.7.2.3, pgk; phosphoglycerate kinase [EC:2.7.2.3] [COG:COG0126] [GO:0004612]

K01834 E5.4.2.1, gpm; phosphoglycerate mutase [EC:5.4.2.1] [COG:COG0406 COG0588 COG1102]

K01689 E4.2.1.11, eno; enolase [EC:4.2.1.11] [COG:COG0148] [GO:0004634]

K00873 E2.7.1.40, pyk; pyruvate kinase [EC:2.7.1.40] [COG:COG0469] [GO:0004743]

K00160 E1.2.4.1, pdh; pyruvate dehydrogenase [EC:1.2.4.1] [GO:0004739]

K00163 E1.2.4.1C, aceE; pyruvate dehydrogenase E1 component [EC:1.2.4.1] [COG:COG2609] [GO:0004739]

K00161 E1.2.4.1A, pdhA; pyruvate dehydrogenase E1 component, alpha subunit [EC:1.2.4.1] [COG:COG1071] [GO:0004739]

K00162 E1.2.4.1B, pdhB; pyruvate dehydrogenase E1 component, beta subunit [EC:1.2.4.1] [COG:COG0022] [GO:0004739]

K00627 E2.3.1.12, pdhC; pyruvate dehydrogenase E2 component (dihydrolipoamide acetyltransferase) [EC:2.3.1.12] [GO:0004739]

K00382 E1.8.1.4, pdhD; dihydrolipoamide dehydrogenase [EC:1.8.1.4] [COG:COG1249] [GO:0004148]

K00016 E1.1.1.27, ldh; L-lactate dehydrogenase [EC:1.1.1.27] [COG:COG0039] [GO:0004459]

Links to GO

Links to COG





KO for function annotation in KEGG

- Function assignment by annotators
 - Based on sequence similarities in SSDB, motifs, genetic location on the genome
 - KO assignment to each gene
- Function assignment by experts
 - Based on literature and experiments
 - Using community databases
 - BSORF for *Bacillus*, CYORF for cyanobacteria
 - Feedback to the KO development





Protein families

- Enzymes
- Cytochrome P450
- Transcription factors
- Ribosome
- Translation factors
- ABC transporters
- G-protein coupled receptors
- GTP-binding proteins
- Ion channels
- Cytokines
- Cytokine receptors
- Cell adhesion molecules (CAMs)
- CAM ligands
- CD molecules
- Bacterial motility proteins





Classification of chemical knowledge

- Compounds
 - Compounds with biological roles
 - Lipids
 - Phytochemical compounds

Functions

- Compound interactions
 - Ion channel agonists/antagonists
 - Cytochrome P450 substrates

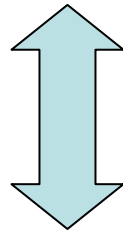
Functions with interacting proteins





Classification of drugs and diseases

- Drugs
 - Therapeutic category of drugs
 - Drug classification



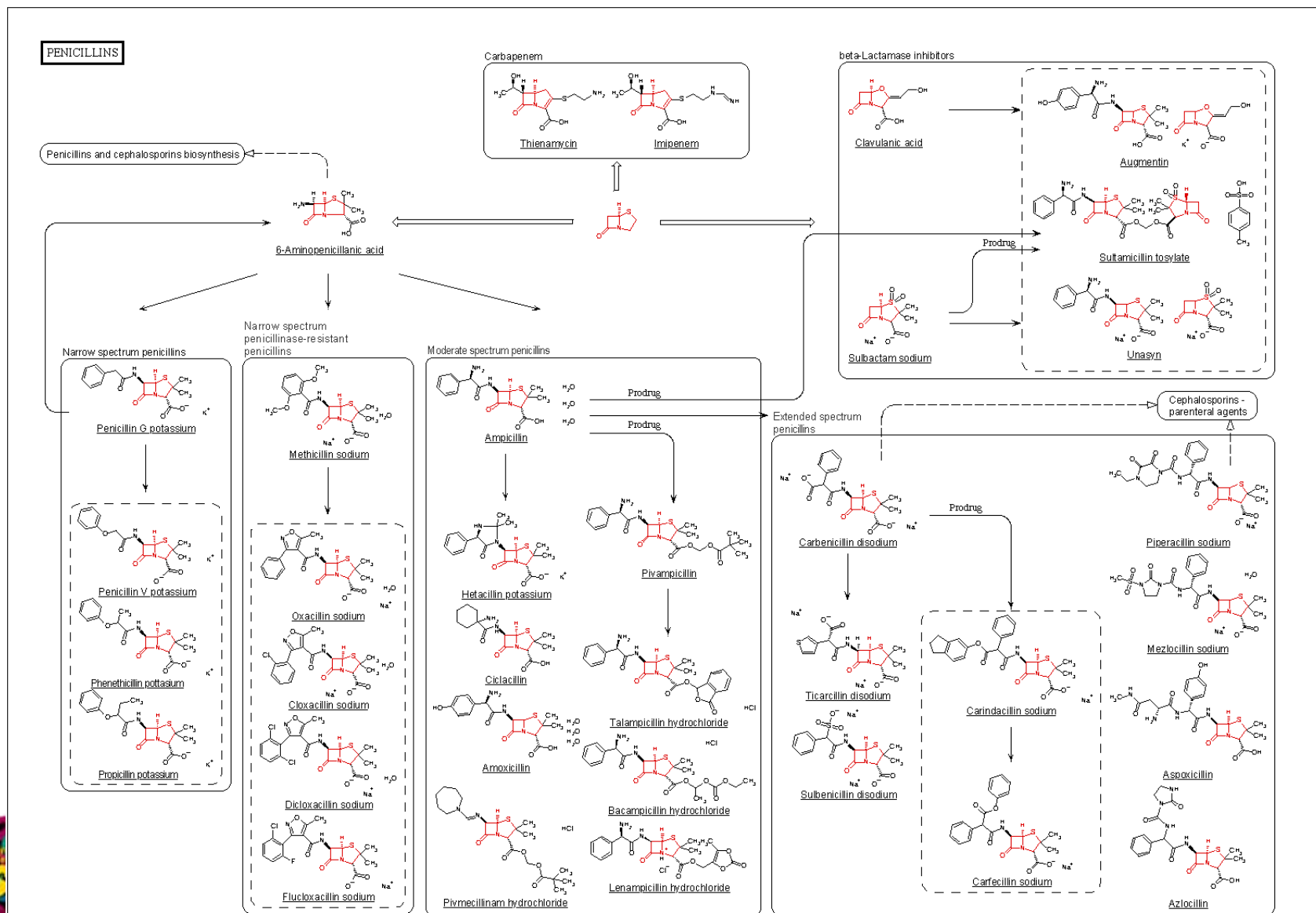
Target bacteria & viruses
Target diseases

- Disease
 - Infectious diseases
 - Metabolic diseases



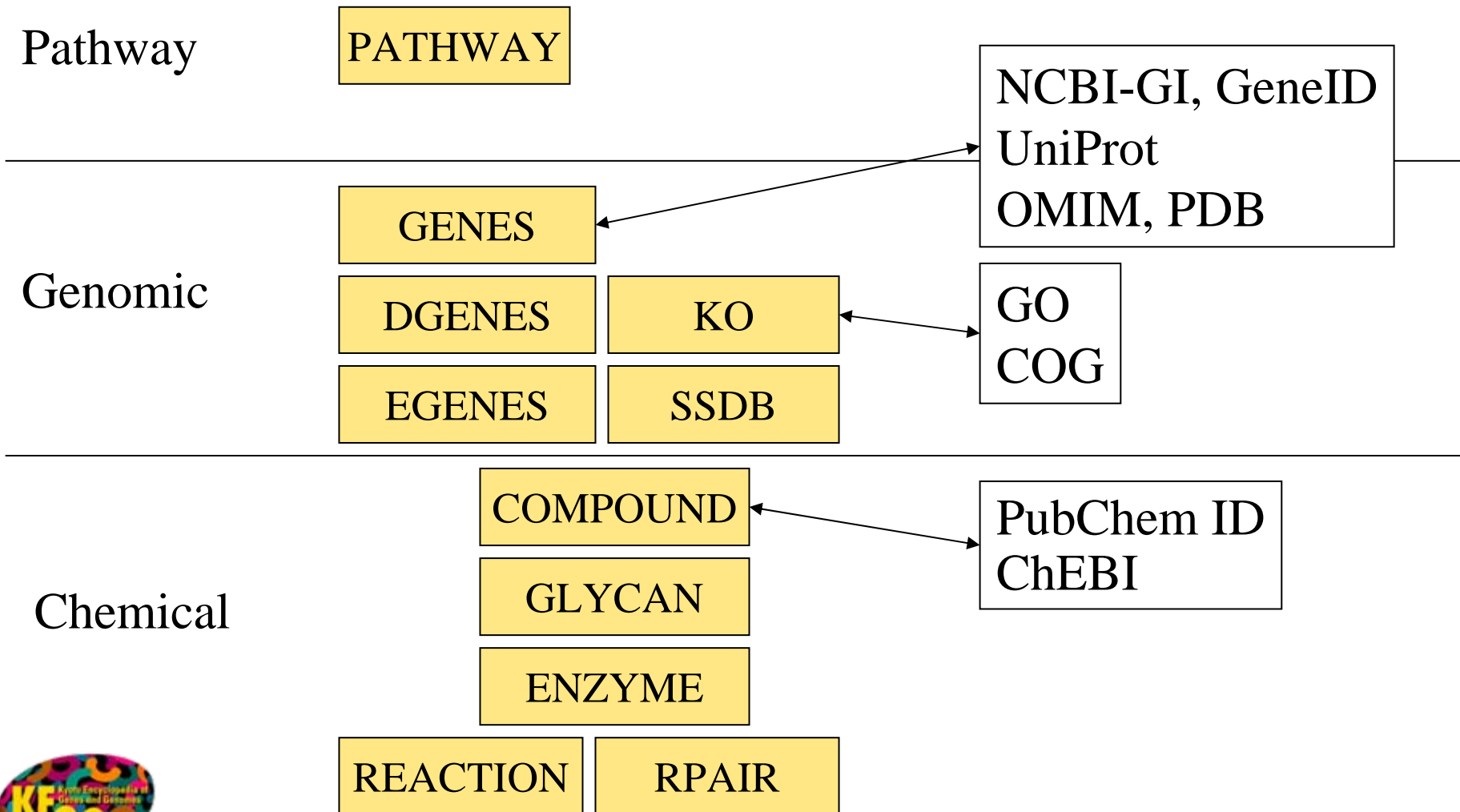


Drug Structure Map





Database cross-references





Contents

- Why BRITE?
- What is BRITE?
- How to use BRITE?
- What is next?





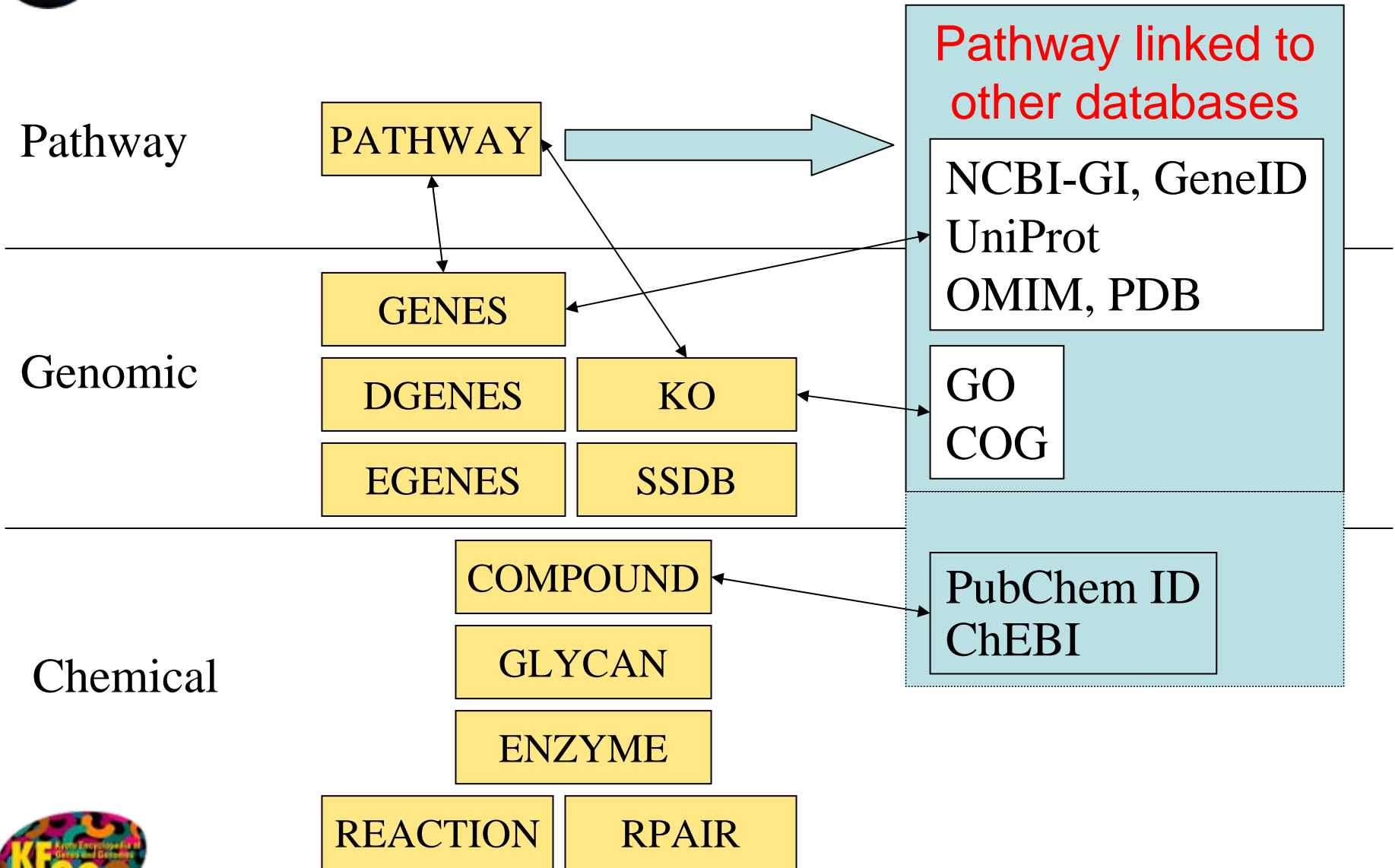
Application

- Database cross-references
 - *bconv* tool for ID conversion between KEGG GENES and outside DBs
 - Pathway maps linked to outside
 - KegArray for outside DB IDs
- Hierarchical classification
 - Query relaxation for pathway computation
 - Classification of reactions
 - Evolutionary analysis of protein domains





Database cross-references





Reaction data



REACTION: R02401

Help

Entry	R02401	Reaction
Name	Glutarate-semialdehyde:NAD ⁺ oxidoreductase	
Definition	5-Oxopentanoate + NAD ⁺ + H ₂ O <=> Glutarate + NADH	
Equation	C03273 + C00003 + C00001 <=> C00489 + C00004	
	<p>The diagram illustrates the chemical reaction where 5-oxopentanoate (C03273) is converted to glutarate (C00489). This process involves the reduction of NAD⁺ (C00003) to NADH (C00004). The structures shown include the reactants and products, with a red arrow indicating the reaction direction.</p>	
RPair	RP: A02182 C00489_C03273 main	
Pathway	PATH: rn00310 Lysine degradation	
Enzyme	1.2.1.20	
LinkDB	All DBs	





From reaction to reactant pair



RPAIR: A02182

Help

Entry	A02182	ReactantPair
Name	C00489_C03273	
Compound	C00489 C03273	
Type		
RDM		
Relation		
Reaction		
KCF data		

Extracting reaction center (RC)
by compound alignment

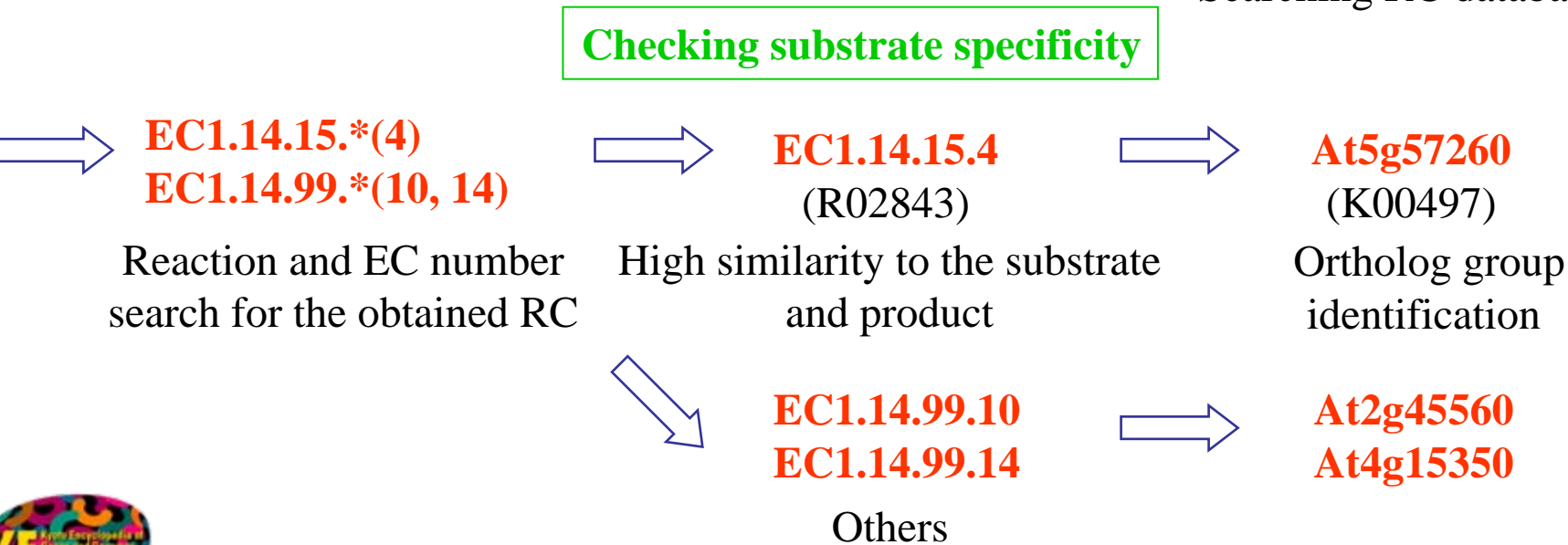
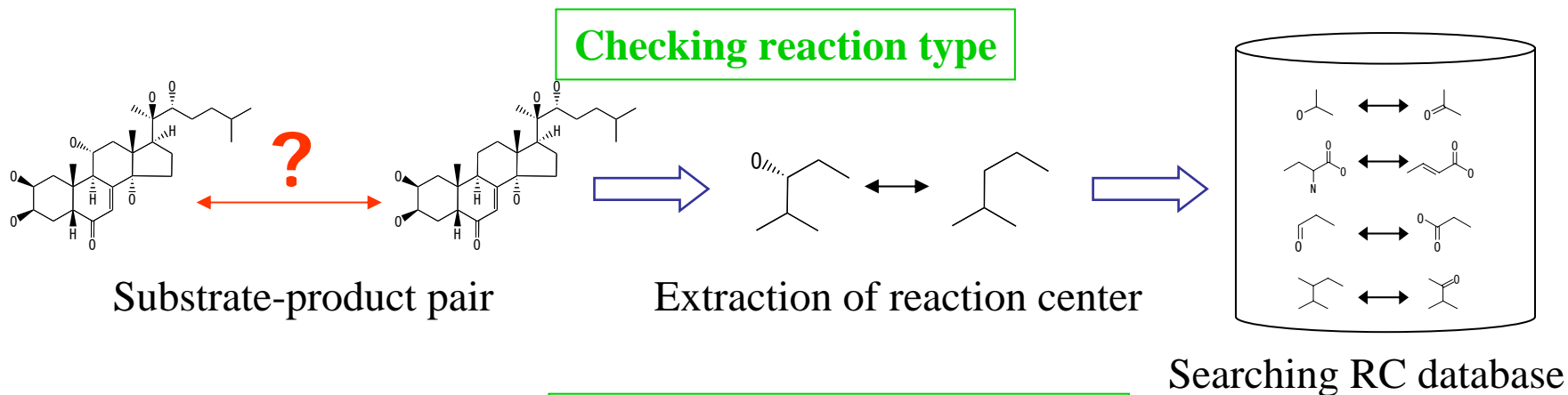
C6a-C4a : O6a-* : C1b+O6a-C1b+O4a

RC: diff: match





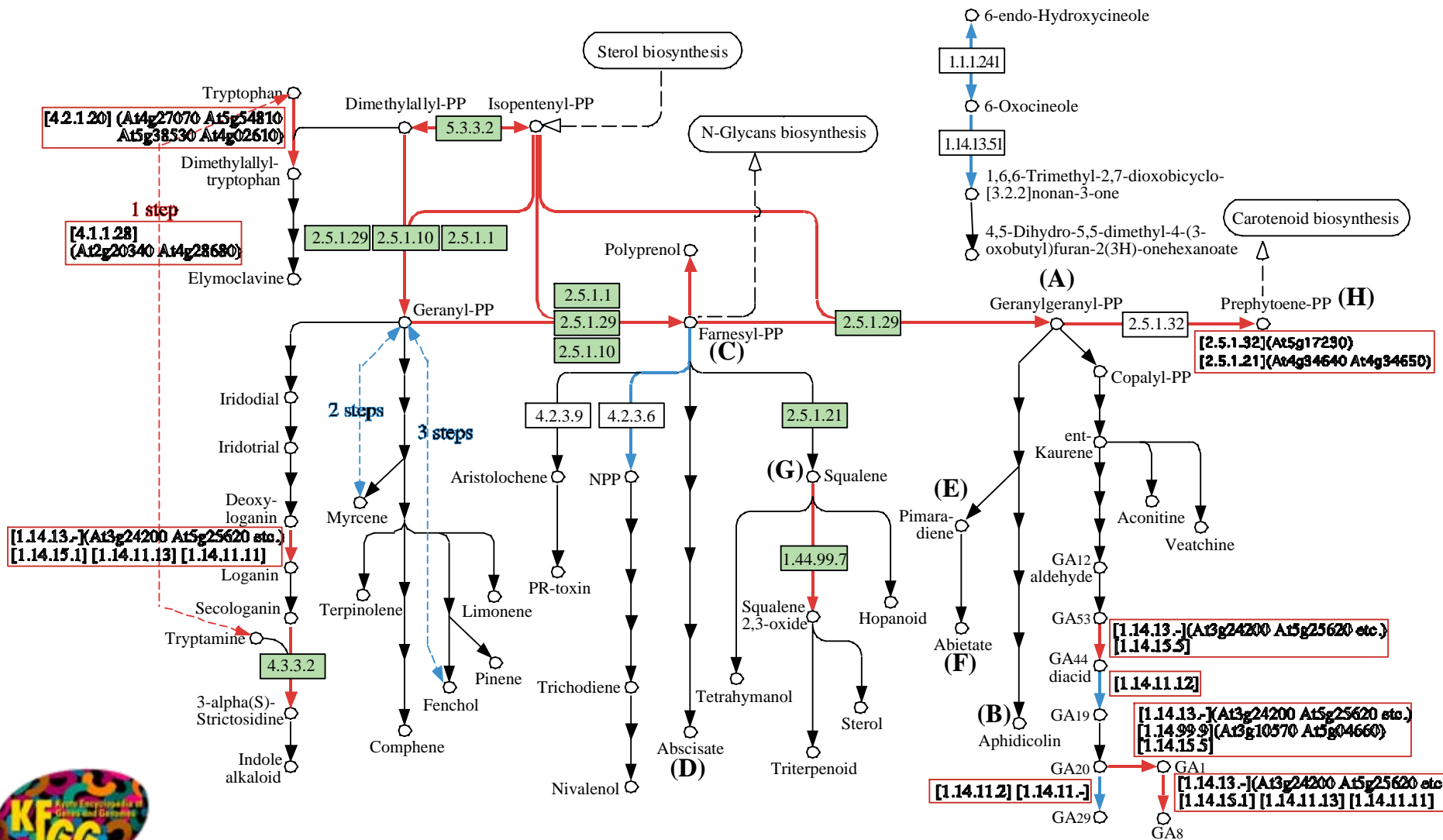
Prediction of novel reaction pathways





TERPENOID BIOSYNTHESIS

Terpenoid biosynthesis





Contents

- Why BRITE?
- What is BRITE?
- How to use BRITE?
- What is next?





Current problems & future work

- Classification and handling of generic form
 - Glucose 6-phosphate
 - D-Glucose 6-phosphate
 - alpha-D-Glucose 6-phosphate
 - beta-D-Glucose 6-phosphate
 - Amino acid
 - L-Amino acid
 - L-Alanine
 - L-Lysine
 - D-Amino acid
 - D-Alanine
 - D-Lysine

L-Amino acid <--> D-Amino acid

--> L-Alanine

? -->





Current problems & future work

- Naming problem
 - IUBMB Enzyme nomenclature -> KEGG LIGAND
 - $\text{pai} \leftrightarrow \text{pi}$
 - $\text{protein} \leftrightarrow [\text{protein}]$
 - Need a heuristics or structure-based comparison
 - No names for many glycans
- More relationships between classification and pathways
 - Drug structure maps
- Application to pathway prediction and drug design





Summary

- BRITE provides
 - Functional hierarchies for function annotation of genes and knowledge management
 - Database cross-references for integrating existing databases and KEGG
- TM and IR will be useful for
 - Extraction of new knowledge such as drug development and protein-ligand interaction
 - Extraction of knowledge for verifying the predicted novel pathways





KEGG Project Team

Project Leader

Minoru Kanehisa

System Development Team

Susumu Goto, Kotaro Shiraishi (FQS), Yoko Sato (FQS),
Masayuki Kawasima (FQS), Satoshi Miyazaki (FNS),
Miho Yamaguchi (FNS), Yukiko Shimata (FNS), Koichi Ohkubo (SGI),
Junji Fukumoto (SGI), Hideya Uehara (SGI), Koichiro Tonomura

PATHWAY Database Development

Masami Hamajima, Yuriko Matsuura, Junko Yabuzaki, Toshi Nakatani,
Junko Nishida, Atsuko Yano, Etsuko Okabe, Sachiyo Ito, Mao Tanabe,
Toshiaki Katayama, Shuichi Kawashima, Masumi Itoh,
Vachiranee Limvipuvadh, Mika Hirakawa

KO/GENES Database Development

Miho Furumichi, Kayo Okamoto, Megumi Yamaguchi, Ayumi Kirioka,
Tomoko Hattori, Tomomi Kamiya

LIGAND Database Development

Rumiko Yamamoto, Tomoko Komeno, Kana Matsumoto, Ami Tanaka,
Mayumi Takashio, Kosuke Hashimoto, Masahiro Hattori, Masaaki Kotera

