

Dynamic Exploration and Editing of KEGG Pathway Diagrams

Christian Klukas & Falk Schreiber

Leibniz Institute of Plant Genetics
and Crop Plant Research Gatersleben
Germany

16. July 2007



IPK Gatersleben

Leibniz - Institut für Pflanzengenetik und Kulturpflanzenforschung Gatersleben

Leibniz-Institut für Pflanzengenetik
und Kulturpflanzenforschung (IPK)

Corrensstraße 3
D-06466 Gatersleben
Tel.: 039482-50 / Fax: 039482-5500



Mitglied der
**Leibniz
Gemeinschaft**



Outline

- 1 Motivation and Introduction
 - Visualisation
 - The KEGG Pathway library
 - Static and dynamic visualisation
- 2 Methods for a dynamic KEGG Pathway exploration
 - The KEGG Markup Language (KGML)
 - Overview and definitions
 - Top-down and bottom-up navigation, arranging and collapsing pathways
- 3 Implementation
 - Motivation
 - Processing of KGML structures
- 4 Summary and outlook
 - Results
 - Acknowledgments

What is Visualisation Good For?

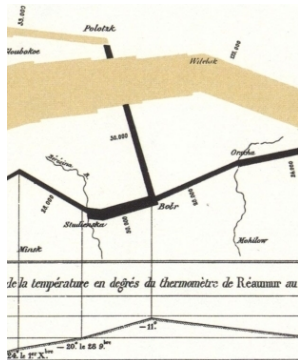
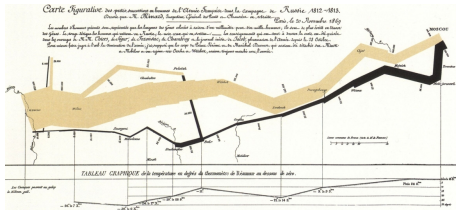
Visualisation categories

- Information transfer and presentation
- Visual data analysis/
Visual data mining

What is Visualisation Good For?

Visualisation categories

- Information transfer and presentation
- Visual data analysis/
Visual data mining



[Minard, 1869]

What is Visualisation Good For?

Visualisation categories

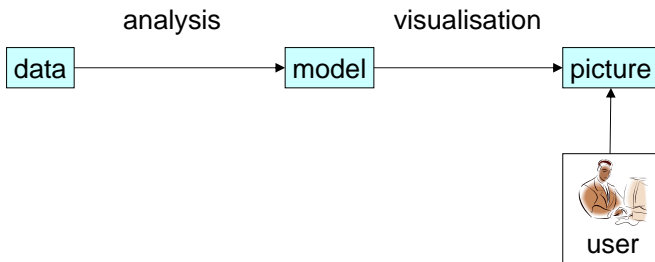
- Information transfer and presentation
- Visual data analysis/
Visual data mining



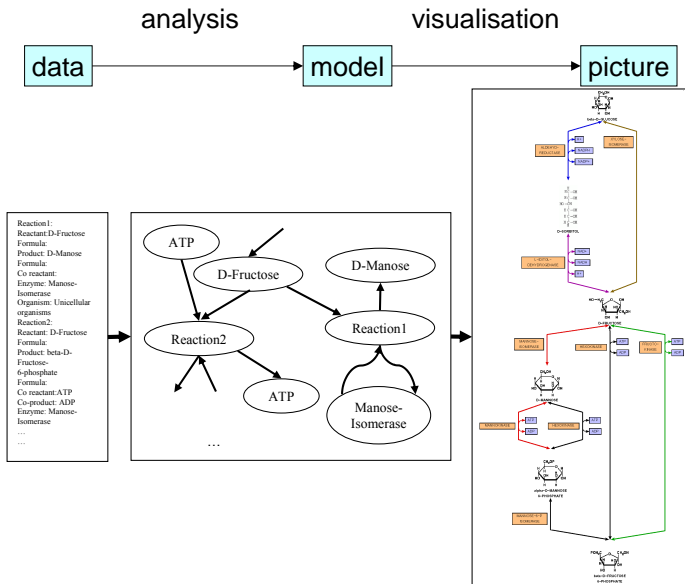
[Snow, 1854]

Visualisation Pipeline

For information transfer and presentation

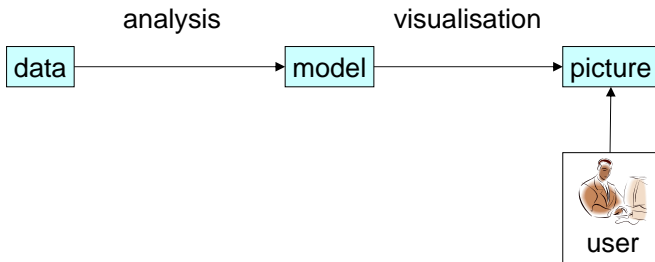


Visualisation Pipeline - Example



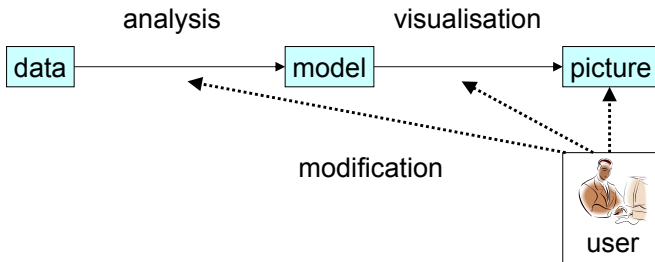
Visualisation Pipeline

For information transfer and presentation



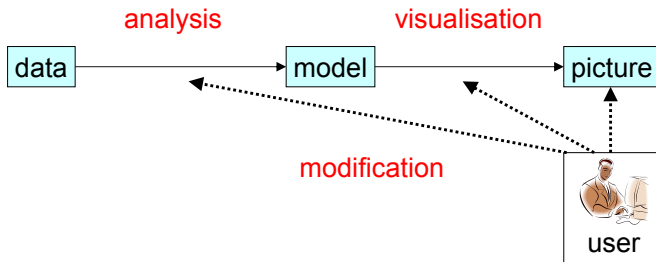
Visualisation Pipeline

For visual data analysis

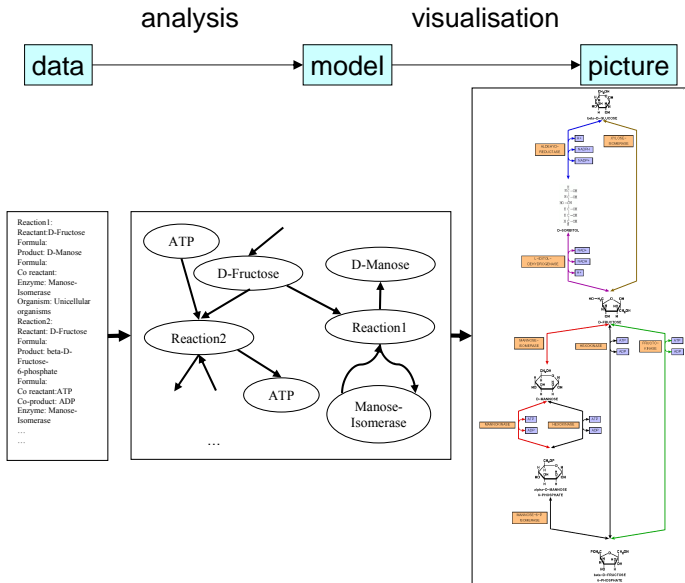


Visualisation Pipeline

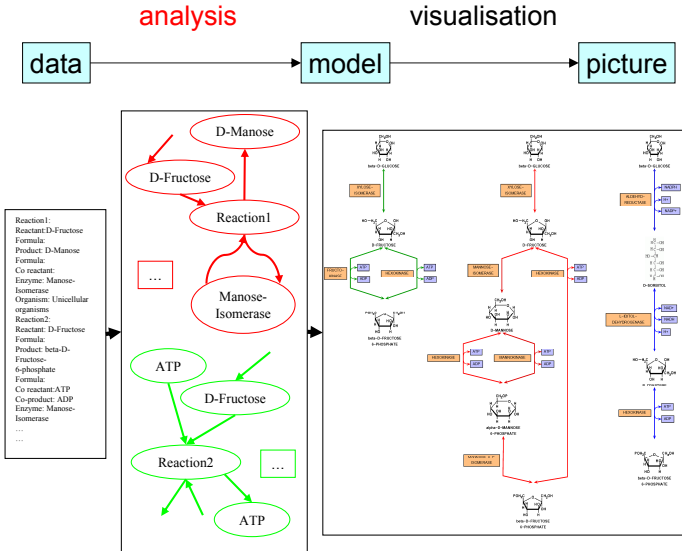
For visual data analysis



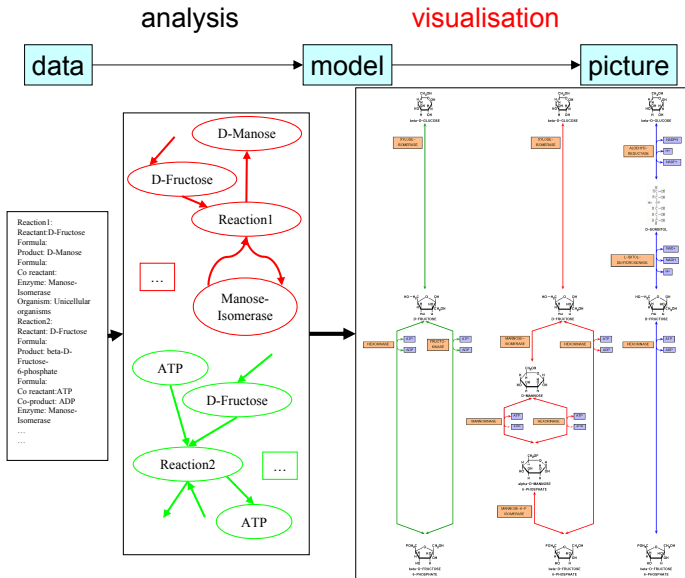
Visualisation Pipeline - Modifications



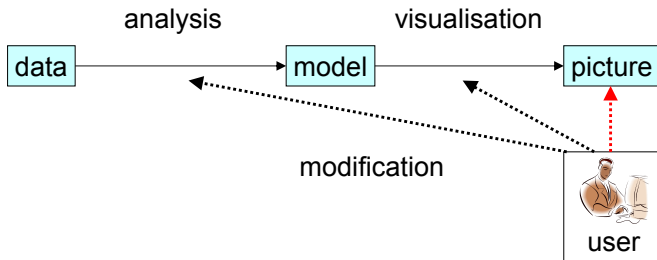
Visualisation Pipeline - Modifications



Visualisation Pipeline - Modifications



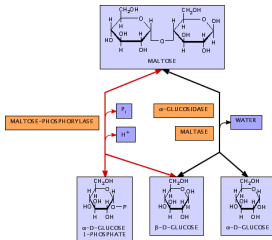
Visualisation Pipeline - Example



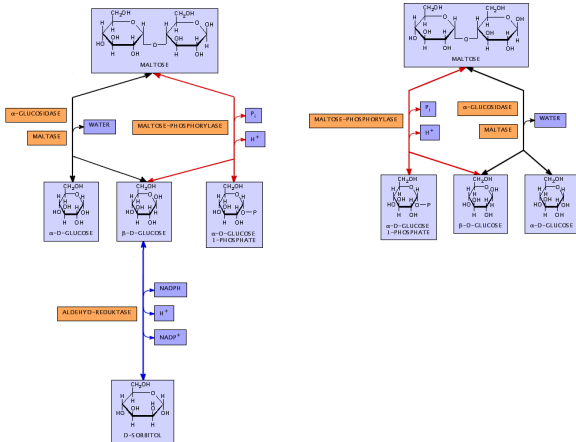
Modification of the picture

- Standard interaction methods (e.g. zooming, panning)
- Focus and context techniques (e.g. semantical zoom from overview to detail, fisheye views)
- Animations
- Important: preserving the mental map of the user

Preserving the Mental Map

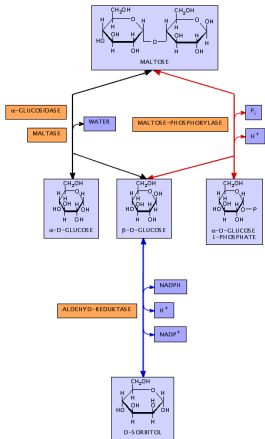


Preserving the Mental Map

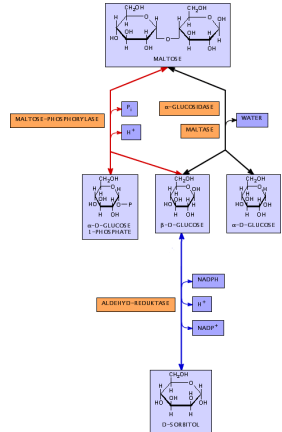
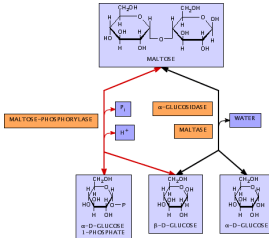


Not mental map
preserving

Preserving the Mental Map



Not mental map
preserving

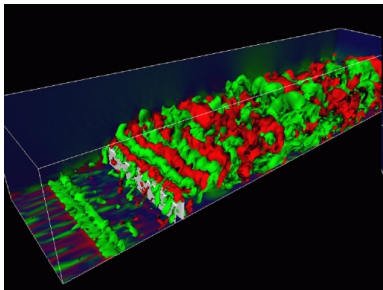


Mental map
preserving

Scientific versus Information Visualisation

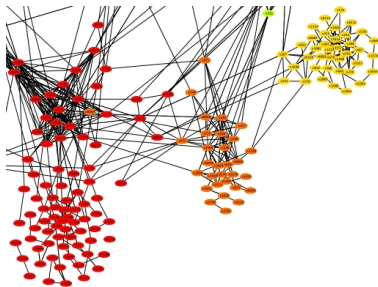
Scientific visualisation

- Data with given coordinates and geometry
- MRT pictures, flow over wings, ...



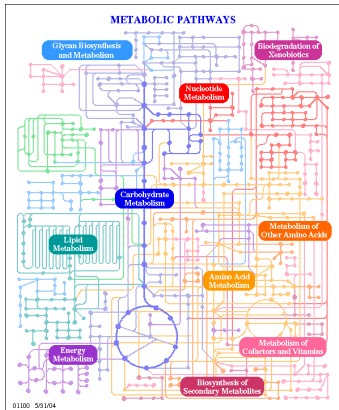
Information visualisation

- Data without given coordinates or geometry
- UML diagrams, biological networks, ...



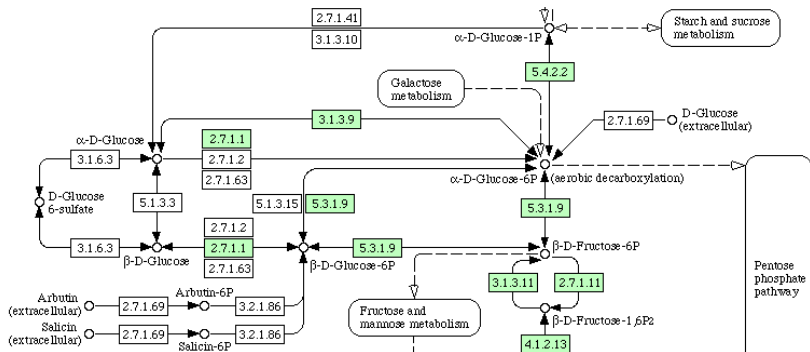
Uses of the KEGG Pathway library

- The Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway database is a valuable information resource for researchers in the fields of life sciences
- It contains metabolic and regulatory processes in the form of wiring diagrams used for browsing and information retrieval
- May be used as a base for modelling and simulation



Overview of KEGG metabolic pathways (source: KEGG website)

Uses of the KEGG Pathway library



KEGG pathway (source: KEGG website)

Dynamic and static visualisation

STATIC Visualisation

- Uses pictures which were created a longer time before they are used
- Pre-defined view of the data
- Limited support for navigation
- Editing is not possible for the end-user

DYNAMIC Visualisation

- Pictures are created by the end-user based on up-to-date data
- Details of the view may be modified
- Support for navigation and extension of the drawing
- Editing of the structure of the drawing is often possible

Use of semi-static pathway visualisation

- The KEGG website uses semi-static visualisations for the presentation and navigation of its pathway information. Parts of the drawing (e.g. colouring) are flexible, but the structure stays fixed
- KEGG's hand curated visualisation approach offers a good pathway presentation and navigation
- *But* it does not allow the creation and visualisation of user-specific pathways
- Solution to this problem: Development of methods for the dynamic visualisation, interactive navigation and editing of KEGG pathway diagrams

The KGML file format as the basis of dynamic KEGG pathway visualisation

- Foundation for the dynamic KEGG Pathway visualisation is the XML based KEGG Markup Language
- KGML is widely used in life science research (e. g. VisANT, kegg2sbml, Biopathways Workbench, BioUML, VANTED, GenMAPP, PathwayExpert and BioRuby)
- KGML contains layout information which enables us to combine the advantages of static and dynamic visualisation

Graph representation

Graph

A *graph* $G = (V, E)$ consists of a finite set of vertices V and a set of edges E , where each edge connects two different vertices.

Graph representation

Graph

A *graph* $G = (V, E)$ consists of a finite set of vertices V and a set of edges E , where each edge connects two different vertices.

Pathway Overview Graph

Let $G_O = (V_O, E_O)$ be the *pathway overview graph* where each *map link vertex* $v \in V_O$ represents a KEGG pathway and each edge $e \in E_O$ represents the connection between pathways. In G_O each vertex $v_i \in V_O$ represents a pathway graph G_i .

Graph representation

Graph

A *graph* $G = (V, E)$ consists of a finite set of vertices V and a set of edges E , where each edge connects two different vertices.

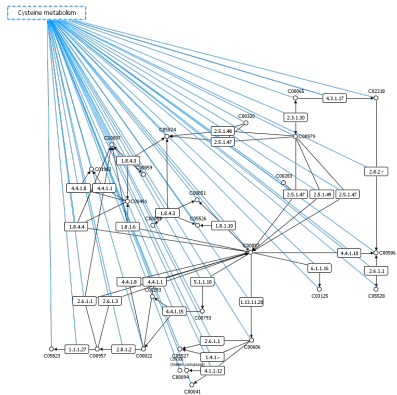
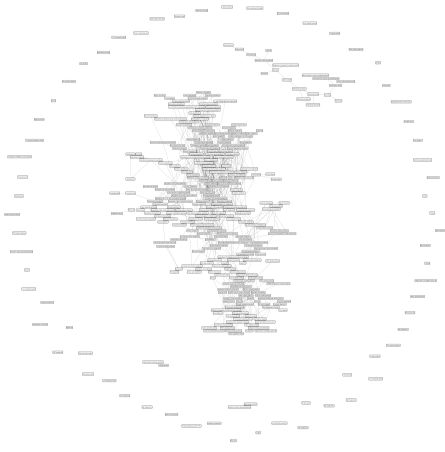
Pathway Overview Graph

Let $G_O = (V_O, E_O)$ be the *pathway overview graph* where each *map link vertex* $v \in V_O$ represents a KEGG pathway and each edge $e \in E_O$ represents the connection between pathways. In G_O each vertex $v_i \in V_O$ represents a pathway graph G_i .

Pathway Graph

A KEGG *pathway graph* $G_i = (V_i, E_i)$ is a graph where each vertex $v \in V_i$ represents a compound, enzyme, reference to another pathway or another object from KEGG and each edge $e \in E_i$ represents a relation or reaction.

Graph representation



KEGG pathways - overview and hierarchical structure

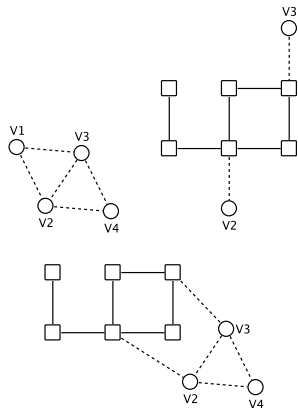
Approaches for the dynamic visual exploration of KEGG pathways

- The following methods have been developed:
 - ① **Extending the overview**
 - ② **Stepwise pathway extension**
 - ③ **Arranging pathways**
 - ④ **Collapsing pathways**

Extending the pathway overview (top-down navigation)

Overview graph G_O is extended by replacing a vertex v_i with its pathway graph G_i :

- 1 The positions of V_i are modified so that their average position equals the position of v_i in the graphical representation of G_O
- 2 v_i is removed
- 3 Elements of G_i are added to G_O
- 4 Map link vertices which are part of G_i pointing to a pathway already linked from G_O are removed, new edges pointing from their neighbour vertices to the remaining matching map link vertex (present in G_i) are created

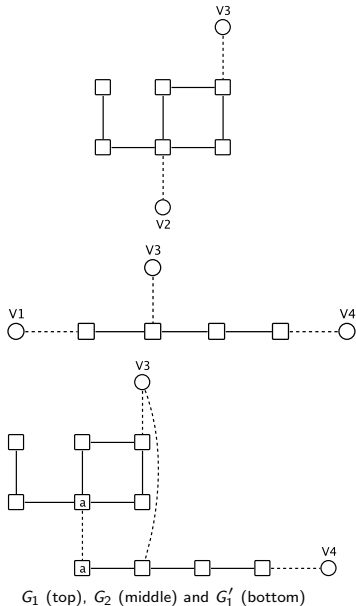


Overview graph G_O (top-left), pathway graph G_1 (top-right) and the integration of G_1 in G_O

Stepwise pathway extension (bottom-up navigation)

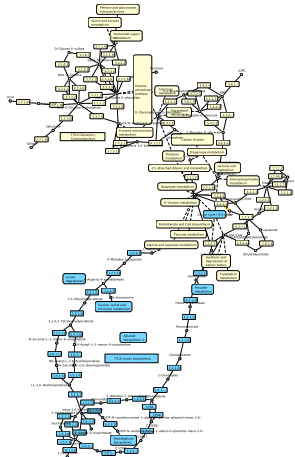
Pathway graph G_i is extended with G_j :

- 1 The positions of V_j are modified so that their average position equals the position of v_i in the graphical representation of G_i
- 2 The graph elements of G_i and G_j are combined
- 3 The neighbour vertices of the map link vertices v_i and v_j are pairwise connected to each other, v_i and v_j are removed from G'_i
- 4 Map link vertices which are part of G_j pointing to a pathway already linked from G_i are removed from G'_i , new edges are created (as before)



Arranging pathways (motivation)

- A manual layout of networks combining several pathways becomes complicated, standard automatic layouts produce pictures which are difficult to understand
- The combination of dynamic visualisation and semi-static visualisation (using manually layouted single pathway graphs) can help

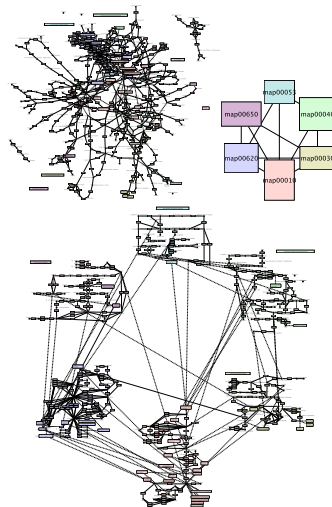


Force directed layout for a graph combining two pathways

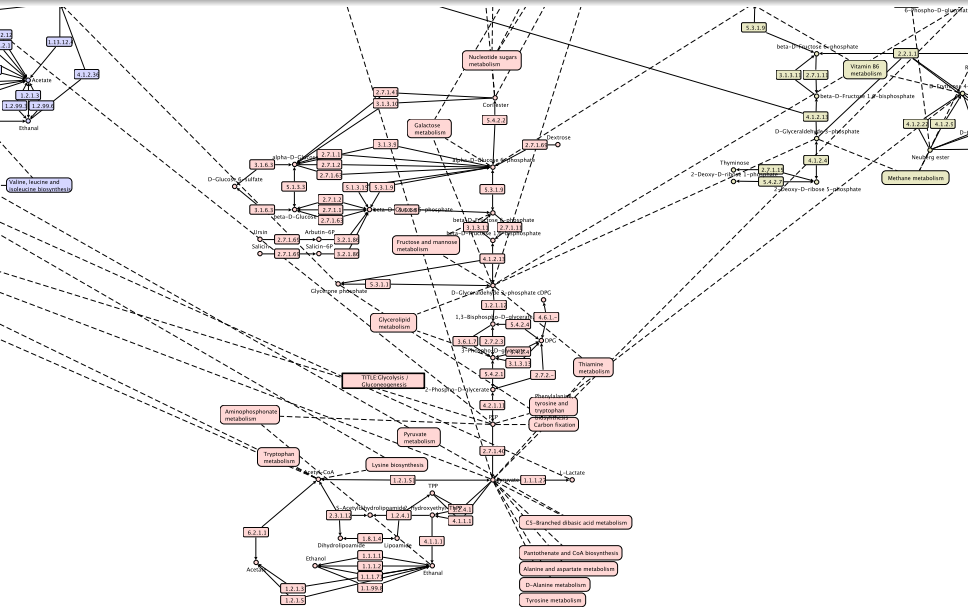
Arranging pathways

Improvement of the layout of integrated pathway graphs:

- 1 To create a specific overview graph G for each relevant pathway graph G_i ; a vertex v_i is included in G and edges are created, representing references between pathway graphs
- 2 The specific overview graph is layouted (top-right figure)
- 3 The vertices v_i of the overview graph are replaced by the pathway graphs G_i ; as explained before (top-down navigation), to obtain a layout of all combined pathways (bottom figure)



Arranging pathways

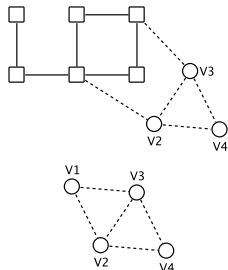


Collapsing pathways

After a detailed investigation a pathway may be collapsed into a map link vertex:

A pathway G_i is part of a graph G

- 1 A new vertex v_i representing the pathway graph G_i is created in G and placed at the centre of the drawing of pathway G_i
- 2 All edges connecting vertices of G which are not part of G_i with vertices of G_i are reconnected to v_i
- 3 All vertices of G_i are removed from G



Graph G (top), pathway graph G_i (part of G) is collapsed and replaced by $V1$ (bottom)

Implementation of a interactive KEGG pathway browser and editing system

Requirements

- Interactive visualisation and exploration of pathways is a requirement to study biological processes
- Scientists often would like to change the pathway structure, e. g. to design species-specific pathways

Solution

- Based on Gravisto (developed at the University of Passau) we implemented the presented visual exploration methods in the Java Web Start application **“KGML-ED”**
- KGML-ED is a graph editor and visualisation system with KGML file im- and export; it supports editing of pathway structures and attributes

Transformation between KGML and graphs

```

<?xml version="1.0"?>
<!DOCTYPE pathway SYSTEM "http://www.genome.jp/kegg/xml/
    KGML_v0.6.1_.dtd">
<pathway name="path:map00010" org="map" number="00010"
    title="Glycolysis / Gluconeogenesis">
  <entry id="18" name="ec:1.2.4.1" type="enzyme" reaction="rn:R03270"
    link="http://www.genome.jp/dbget-bin/www_bget?enzyme+1.2.4.1">
    <graphics name="1.2.4.1" fgcolor="#000000" bgcolor="#FFFFFF"
      type="rectangle" x="362" y="885" width="45" height="17"/>
  </entry>
  <entry id="96" name="cpd:C00248" type="compound"
    link="http://www.genome.jp/dbget-bin/www_bget?compound+C00248">
    <graphics name="C00248" fgcolor="#000000" bgcolor="#FFFFFF"
      type="circle" x="358" y="927" width="8" height="8"/>
  </entry>
  <entry id="112" name="ec:1.1.1.140" type="enzyme" map="37"
    link="http://www.genome.jp/dbget-bin/show_pathway?
      map00051+1.1.1.140"/>
  ...
  <relation entry1="54" entry2="10" type="maplink">
    <subtype name="compound" value="82"/>
  </relation>
  ...
  <reaction name="rn:R03270" type="irreversible">
    <substrate name="cpd:C00248"/>
    <substrate name="cpd:C05125"/>
    <product name="cpd:C00068"/>
    <product name="cpd:C01136"/>
  </reaction>
</pathway>

```

Parts of the glycolysis pathway KGML definition: pathway information, entries, relations and reactions

Transformation between KGML and graphs

General pathway information ↔ Graph Attributes

Attributes such as the pathway title are mapped to graph attributes

Entries ↔ Graph Vertices/Attributes

Entries are modelled as graph vertices; entries connected with map link entries are modelled as graph vertex attributes

Relations ↔ Graph Edges

Relations in KGML are transformed to and from graph edges

Reactions ↔ Graph Edges

Substrate and product vertices specified for a reaction are connected to the corresponding enzyme vertices (in case of multiple occurring substrates or products the nearest placed entry vertex is used)

Summary

- Novel methods for the dynamic exploration of KEGG pathway diagrams
- Methods use a combine aspects of static and dynamic visualisation
- KGML-ED tool may help researchers to gain further access to the comprehensive KEGG Pathway information. It is available at <http://kgml-ed.ipk-gatersleben.de>

Further Details

C. Klukas and F. Schreiber (2007): Dynamic exploration and editing of KEGG pathway diagrams. *Bioinformatics* 23, 344-350, 2007

Thank you for your attention!

Acknowledgements

- Christian Klukas
- Gravisto development team at University of Passau - <http://gravisto.fmi.uni-passau.de/>
- Kanehisa Teams at Kyoto and Tokyo - <http://www.genome.jp/kegg/xml/>

Dynamic Exploration and Editing of KEGG Pathway Diagrams

Christian Klukas & Falk Schreiber

Leibniz Institute of Plant Genetics
and Crop Plant Research Gatersleben
Germany

16. July 2007

<http://nwg.bic-gh.de>
<http://kgml-ed.ipk-gatersleben.de>

Interactive visual network analysis is very useful to observe the expected and discover the unexpected, and will help in exploring biological systems.