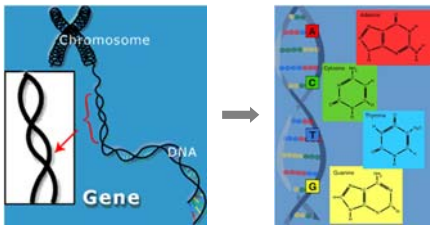
2. Biochemical Network Visualization

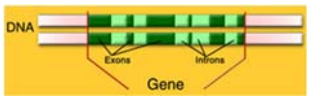
Information Visualization (186.141)
TU Vienna, Austria
May 24-25, 2011

- Correlations: Genome – Chromosome – Gene – DNA



- DNA: Double helix with nucleotides from sugar, phosphate and four bases (Adenin, Cytosin, Guanin, Thymin)

- Genome: only 5% genes consisting of Exons and Introns



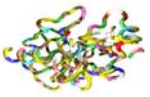
Organism	Base Pairs	Genes
yeast	12 mill.	6000
fruit fly	137 mill.	13.601
human	3 billions	~30.000

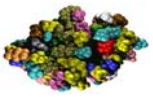
- RNA: Uracil instead of Thymin, single strands

- Proteins are built by amino acids [Video]
- 20 amino acids: coded by nucleotide triplets (Codons)
  - e.g. GCA, GCC, GCG → Alanin

Prime- Secondary- Tertiary Structure

ADMVIKAPAG
AKVTKAPVAF
SHKGHASMDC





- Protein synthesis
  - Transcription: DNA → mRNA
  - Translation: mRNA → amino acids → protein

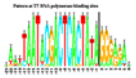
http://de.wikipedia.org/wiki/Codon

- Support of biologists within processing, analysis and interpretation of large data sets
- Approaches:
  - Biological data bases
  - Methods for comparison and function prediction of sequences
  - Function prediction of proteins
  - Locating of new correlations
  - Simulation of biological processes

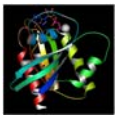
- Visualizations of networks and related structures
  - Networks
  - RNA-secondary structure
  - Phylogenetic trees

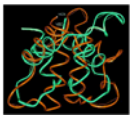
Here, we will focus on these issues!

- Visualizations of sequences
  - Sequence alignments



- Proteins
  - Structure prediction
  - Structure comparison
  - Protein dynamics





2.1.2 Visualizations

Nevertheless, a brief excurse on protein structures

- A main research area in molecular biology
- Research area is older than sequence analysis (at the beginning of the 70s)

Why is the structure so important?

- Folded protein has an irregularly shape
  - ⇒ Functionality of the protein depends on this shape
  - ⇒ Bonding capacity with other molecules or proteins (interaction)
- Gap between the number of known sequences and structures diverges extremely (ca. 100:1)

2.2 Biochemical Networks

2.2.1 Motivation

- **Network Biology**
  - Classical biology ⇒ Learning from the individual
  - Systems biology ⇒ Biological processes are integrated systems of many diverse interacting components
  - Starting with the human genome project, high-throughput methods generate a huge amount of data

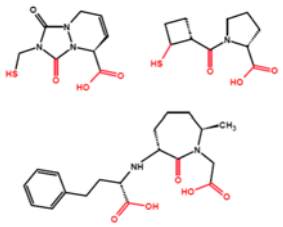
  - Networks are the key concept to structure and combine that data

[Inspired by O. Kohlbacher's talk at Dagstuhl 08]

2.2.1 Motivation

There is a hierarchy of such biological networks

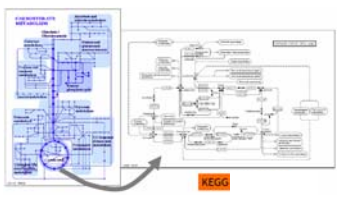
- (Molecular Graphs)
- Metabolic Networks
- Interaction Networks
- Regulatory Networks
- …



2.2.1 Motivation

There is a hierarchy of such biological networks

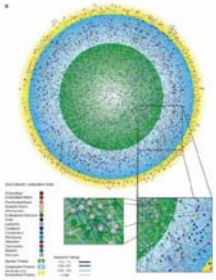
- (Molecular Graphs)
- Metabolic Networks
- Interaction Networks
- Regulatory Networks
- …



2.2.1 Motivation

There is a hierarchy of such biological networks

- (Molecular Graphs)
- Metabolic Networks
- Interaction Networks
- Regulatory Networks
- …



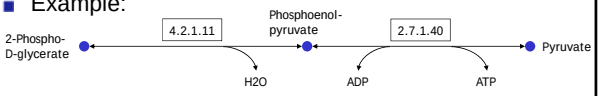
2.2.1 Motivation

Why is that important?

- Diseases can be explained in context of networks
  - Infection: a foreign network starts operating in our own network
  - Genetic defects: incorrect connectivity of an element
- Applications
  - Drug design, metabolic engineering, …

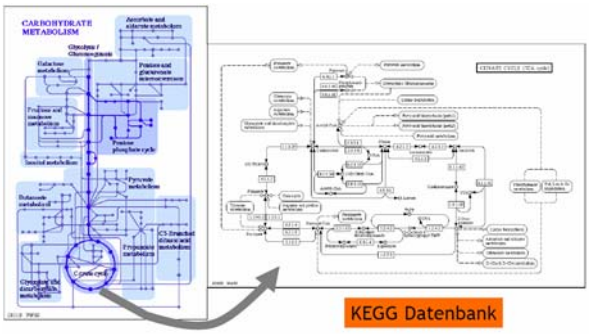
2.2.2 Metabolic Networks

- Paths: Sequence of reactions that provide all together for the translation of one substance into another one (pathways)
- One single reaction is defined by a catalyzing enzyme
- Example:



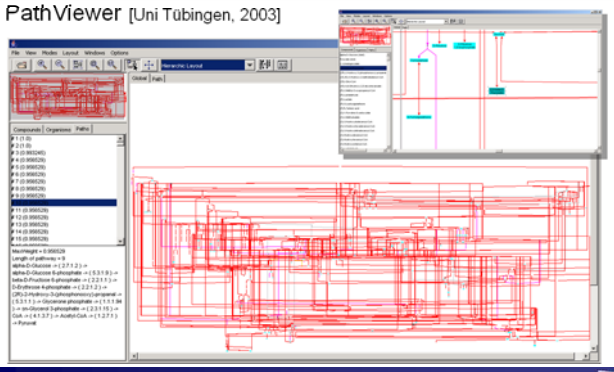
- As a whole, the paths build a so-called metabolic network (MN) [→ Graph]

2.2.2 Metabolic Networks



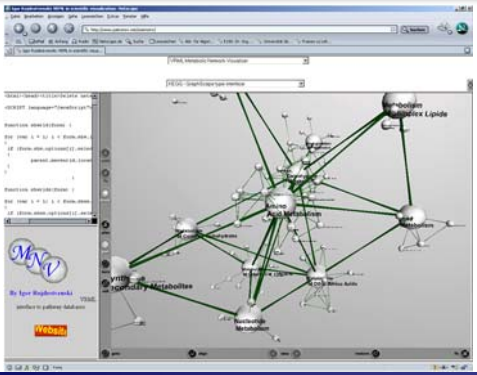
2.2.2 Metabolic Networks

PathViewer [Uni Tübingen, 2003]



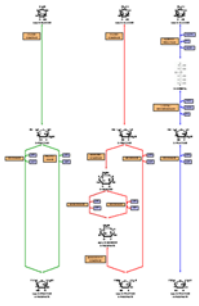
2.2.2 Metabolic Networks

MNV [I. Rojdestvenski, 2003]

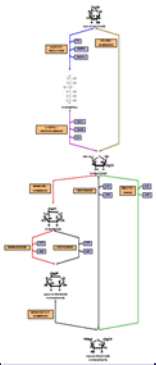


2.2.2 Metabolic Networks

- BioPath



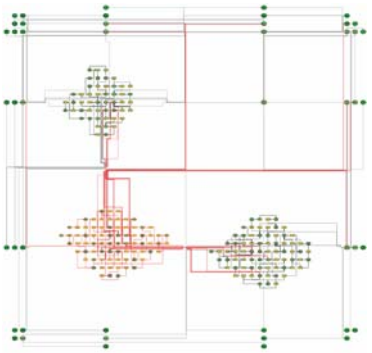
by F. Schreiber



2.2.2 Metabolic Networks

Grid-based Visualization

[Uni Leipzig & Växjö Univ., 2009]

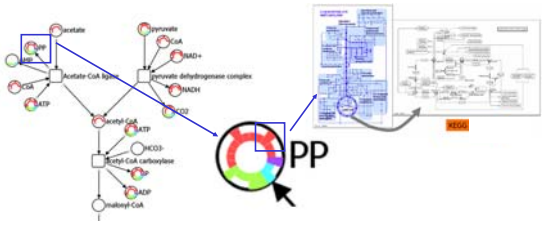


http://www.springerlink.com/content/1p020625151205 4w/

2.2.2 Metabolic Networks

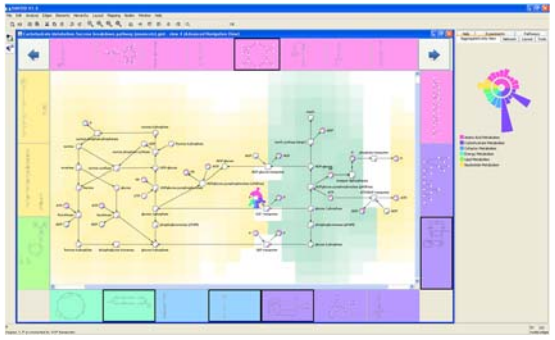
Interrelationships between network components

[Linnaeus Univ. & IPG Gatersleben, 2010]



2.2.2 Metabolic Networks

[Video]



http://sourceforge.net/projects/jlayl/

[I. Jusufi, C. Klukas, A. Kerren, and F. Schreiber. Guiding the Interactive Exploration of Metabolic Pathway Interconnections. IVS Journal, To appear.]

2.2.3 Protein-Protein Interactions

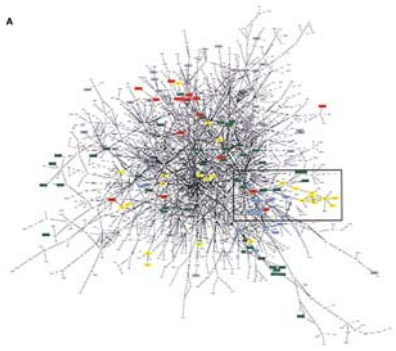
- Nodes are proteins and edges the interactions between them
- Interactions between proteins are important for many biological functions
- Example
  - Signal Transduction: signals from the exterior of a cell are mediated to the inside of that cell by protein-protein interactions of the signaling molecules
  - It plays a fundamental role in many biological processes and in many diseases

http://en.wikipedia.org/wiki/Protein-protein_interactions

2.2.3 Protein-Protein Interactions

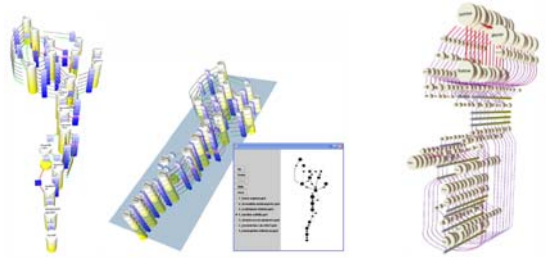
Cytoscape

[Schwikowski et al; 00]



2.2.4 Special Visualizations

- Comparison of biochemical networks
  [U. Brandes et al. Integrative Bioinformatics, 2004]



http://www.csse.monash.edu.au/~tdwyer/

2.2.4 Special Visualizations

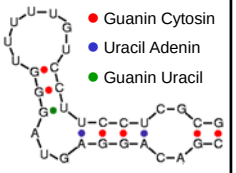
- Replacing nodes by diagrams
  [L. Broisjuk et al., Silico Biology, 2004]

Bar Charts

2.3 RNA Secondary Structure

- One-strand RNA folds itself to a double-strand at suitable positions
- Secondary structure depends on the function of the RNA (or vice versa)

CGACAGGAGUAGGGUUUUUGUCCUUCCUCGCG



2.3 RNA Secondary Structure

- Structure forms
  - E ... single strand
  - M ... multiple loop
  - I ... interior loop
  - H ... hairpin loop
  - B ... bulge loop
  - Helix

2.3 RNA Secondary Structure

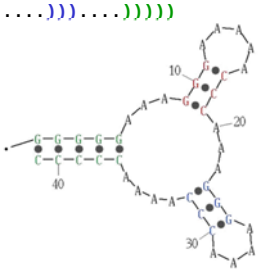
- Variants of visualization
  - Brace form
  - Matrix form
  - Circle graph
  - Structured graph
- Criteria that should be fulfilled by the visualization
  - Readable representation of the structure
  - No overlapping elements of the structure
  - Consistent representation of similar structures
  - Good aesthetics
  - Minimal user interaction (?)

2.3.1 Brace Form

- Typical representation of a brace form:

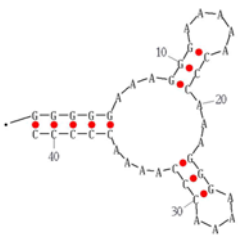
```
GGGGGAAAGGGAAAAACCCAAAGGGAAAACCCAAAACCCCC
(((((...(((.....)))...(((....)))....)))))
```

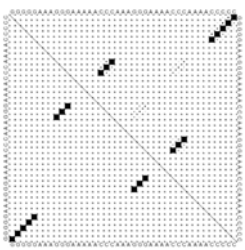
- For each pair:
  an open and a closed brace
- For each unbonded base:
  a dot



2.3.2 Matrix Form

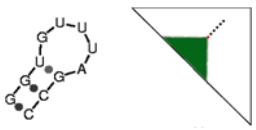
- We add a dot to each position (i, j) in the matrix, if base i and j are paired
- Dot size: Probability for this pairing



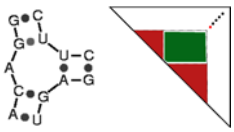


2.3.2 Matrix Form

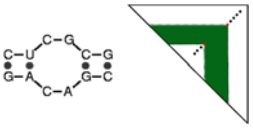
Hairpin loop



Multiple loop



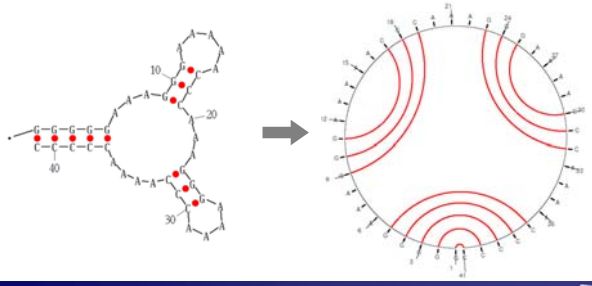
Interior loop



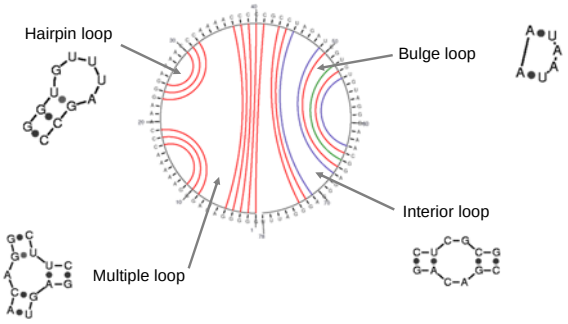
red: further pairings
green: no base pairing

2.3.3 Circle Graph

- Spread bases on a circle
- Bondings between two bases are pairings

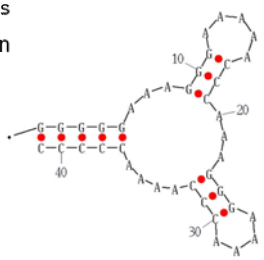


2.3.3 Circle Graph

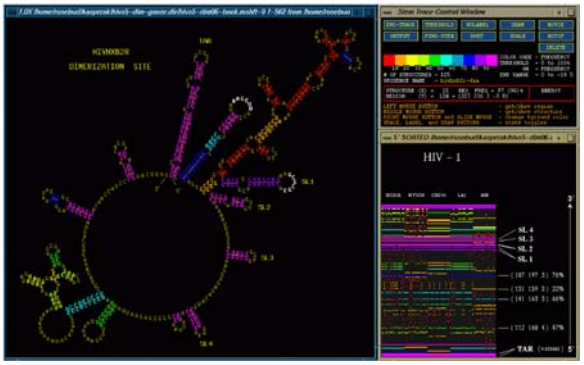


2.3.4 Structured Graph

- Undirected graph
  - Nodes: bases
  - Edges: Bondings between bases
- Representation of the hydrogen bondings (dots, lines, …)
- Numbers
  - Number of the bases
- Start symbol
  -



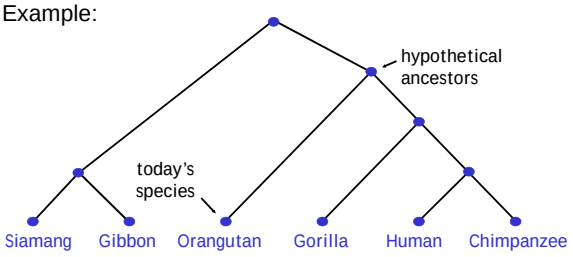
2.3.4 Structured Graph



http://www-lmmb.ncifcrf.gov/~bshapiro/index.html

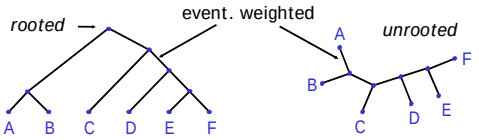
2.4 Phylogenetic Trees

- Problem
  - Analysis of the development history of today's species and their relations
- Example:



2.4 Phylogenetic Trees

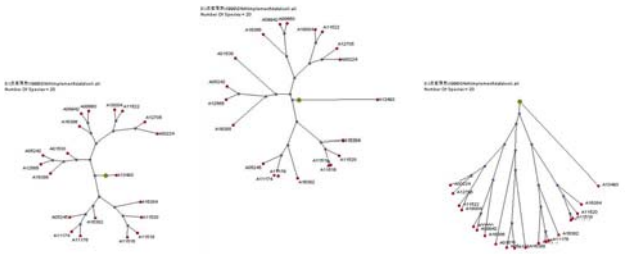
- Objects
  - Species, populations, DNA, proteins
- Tree



- There are two methods for the determination of the root
  - Outgroup method: with not-related objects
  - Hypothesis of a molecular clock: root has the same distance to all objects

2.4.1 Visualizations

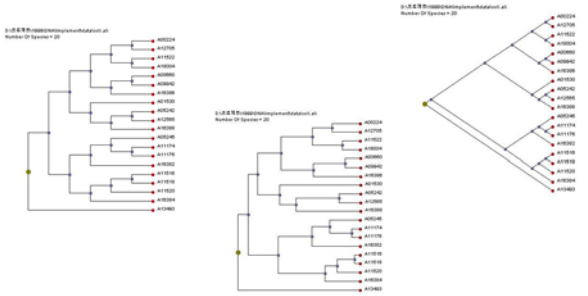
- Example tool: PHYLODRAW
  - Presented by Choi et al., `00



http://pearl.cs.pusan.ac.kr/phylodraw

2.4.1 Visualizations

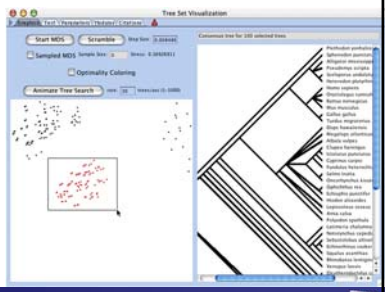
- Example tool: PHYLODRAW (cont.)



http://pearl.cs.pusan.ac.kr/phylodraw

2.4.1 Visualizations

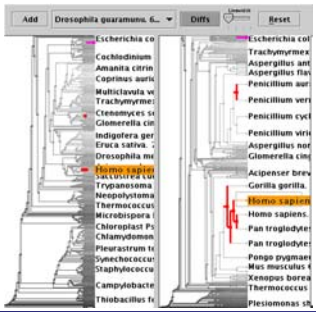
- Example tool: **Tree Sets**
  - Presented by Amenta and Klinger, InfoVis `02



- 2 views
- Points = trees, MDS-Layout
- Consensus tree
- Interaction: Selection of specific points, point sets, (multiple selection possible)

2.4.1 Visualizations

- Example tool: **TreeJuxtaposer**
  - Presented by T, Munzner et al, SIGGRAPH `03



- Focus & Context
- “Guaranteed visibility”
- Computing structural differences

[Video]