Terminology

• **Metabolite: substance produced or used during metabolism such as lipids, sugars and amino acids**

• **Metabolome: the quantitative complement of all the low molecular weight molecules present in cells or biofluids in a particular physiological or developmental state**

• **Metabolomics: a comprehensive analysis of the whole metabolome under a given set of conditions**

Metabo*omics

• **Metabonomics:**

The quantitative measurement of the dynamic multiparametric response of living systems to pathophysiological stimuli or genetic modification (*Nicholson et al., Xenobiotica 1999*)

holistic analysis of biofluids and tissues in order to determine metabolic composition

deals with integrated, multicellular, biological systems including communicating extracellular environments in animal and human biochemistry

• **Metabolomics:**

Measurement of metabolite concentrations and fluxes in isolated cell systems (*Nicholson et al., Xenobiotica 1999*)

deals with simple cell systems and mainly intracellular metabolite concentrations in microbial and plant biochemistry

Metabolomics vs genomics and proteomics

Genomics and proteomics tell you what might happen, but metabonomics tells you what actually did happen
(*Bill Lasley, UC Davis*)

Although changes in the quantities of individual enzymes might be expected to have little effect on metabolic fluxes, they can and do have significant effects on the concentrations of numerous individual metabolites.

The metabolome is further down the line from gene to function and so reflects more closely the activities of the cell at a functional level. Thus, as the 'downstream' result of gene expression, changes in the metabolome are expected to be amplified relative to changes in the transcriptome and the proteome.

Metabolic fluxes (at least as exemplified by glycolysis in trypanosomes) are not regulated by gene expression alone.

General aplications

- Assessing gene function and relationships to phenotypes
- Understanding metabolism and predicting novel pathways
- To increase metabolite fluxes into valuable biochemical pathways using metabolic engineering
- To compare genetically modified organisms
- To assess the effect of environmental/stress/temperature changes that lead to changes in gene expression, flux pathways, and extent of carbon and electron flow through them

Who believes in metabolomics?

NIH Roadmap

http://nihroadmap.nih.gov/initiatives.asp

Building Blocks, Pathways, and Networks Implementation Group

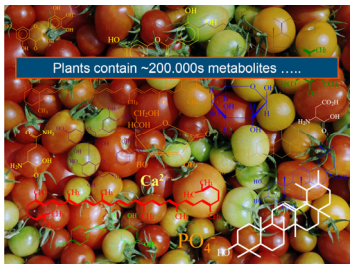
Metabolomics Technology Development. **This initiative will promote development of novel technologies to study cellular metabolites, such as lipids, carbohydrates, and amino acids. Knowledge gained from these studies will be used to understand more precisely the role of metabolites in the context of cellular pathways and networks.**

RFA for "Metabolomics Technology Development"

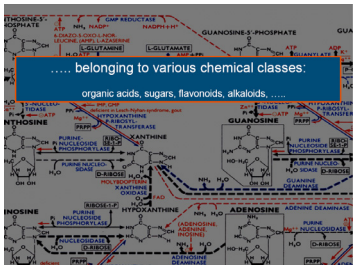
http://grants.nih.gov/grants/guide/rfa-files/RFA-RM-04-002.html

Characteristics of the metabolomes

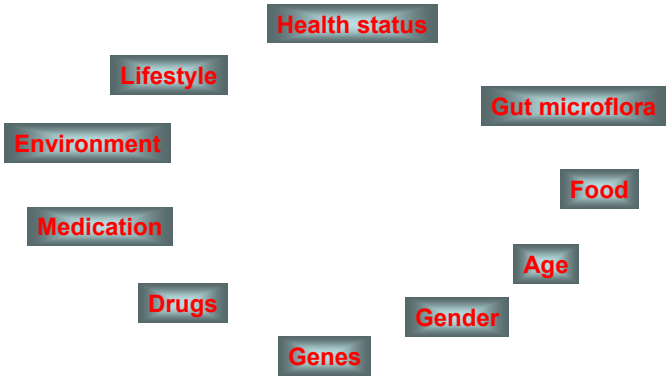
Metabolome size:
- **S. cerevisiae: about 600 metabolites**
- **Plants: estimated 200,000 primary and secondary metabolites**
- **Mammalians: ?**



Metabolite chemical diversity:
- **the metabolome extends over an estimated 7–9 magnitudes of concentration (pmol–mmol)**
- **wide variations in chemical (molecular weight, polarity, solubility) and physical (volatility) properties**



Factors affecting the human metabolome



Dietary contributions to the human metabolome

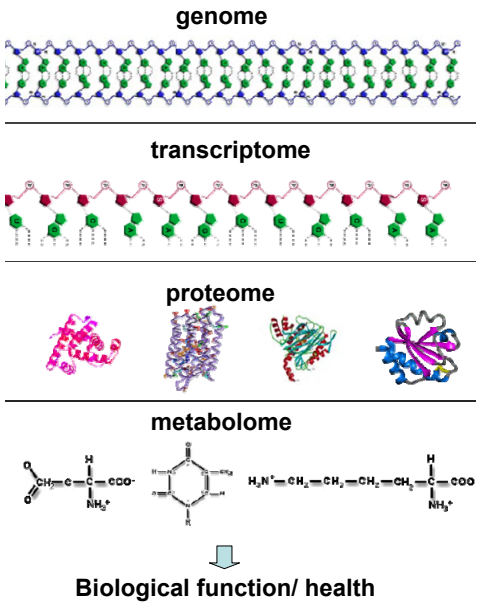
Macronutrient energy Sources

Essential micronutrients

Non-essential, beneficial dietary components

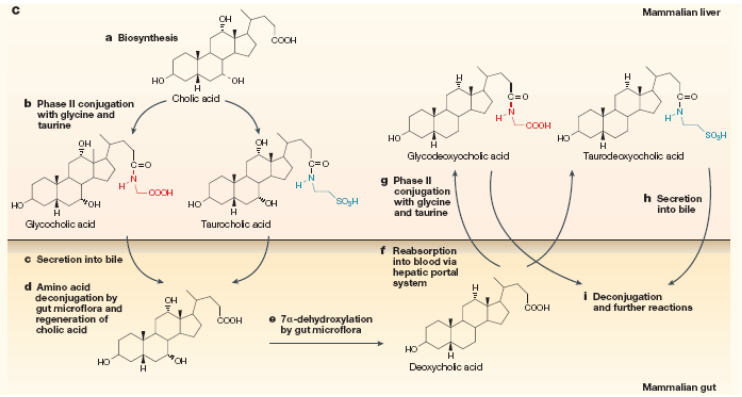
Metabolically neutral dietary components

Dietary toxins and toxicants



Biological function/ health

interacting metabolomes



Example of sym-xenobiotic metabolism occurring in mammals

Cholic acid and other bile acids biosynthesized in the liver undergo a series of conversions in both the host liver and inside the gut microflora; these compartments are connected by the entero-hepatic circulation

Classification of metabolomics approaches

Metabolomics is the study of metabolic changes. It encompasses metabolomics, metabolite target analysis, metabolite profiling, metabolic fingerprinting, metabolic profiling, and metabonomics – *the Metabolomics Society*

Metabolite target analysis: analysis restricted to metabolites of, for example, a particular enzyme that would be directly affected by abiotic or biotic perturbation

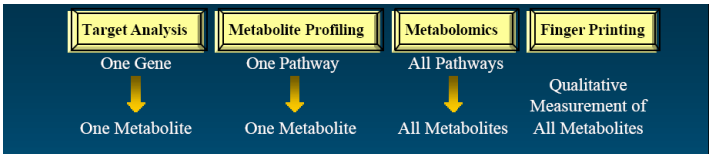
Metabolite profiling: analysis focused on a group of metabolites, for example, a class of compounds such as carbohydrates, amino acids or those associated with a specific pathway

Metabolomics: comprehensive analysis of the whole metabolome under a given set of conditions

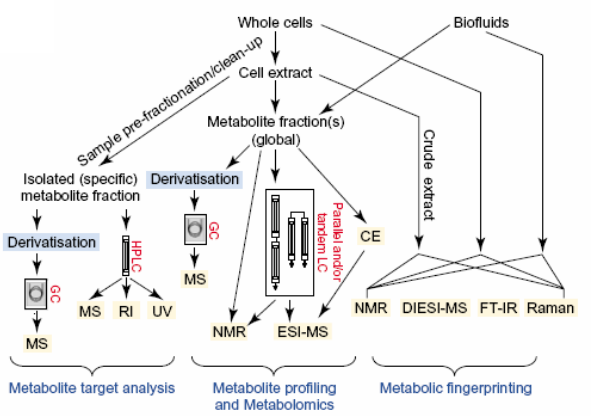
Metabolic fingerprinting: classification of samples on the basis of provenance of either their biological relevance or origin

Metabolic profiling: often used interchangeably with 'metabolite profiling'; m.p. is commonly used in clinical and pharmaceutical analysis to trace the fate of a drug or metabolite

Metabonomics: measure the fingerprint of biochemical perturbations caused by disease, drugs and toxins



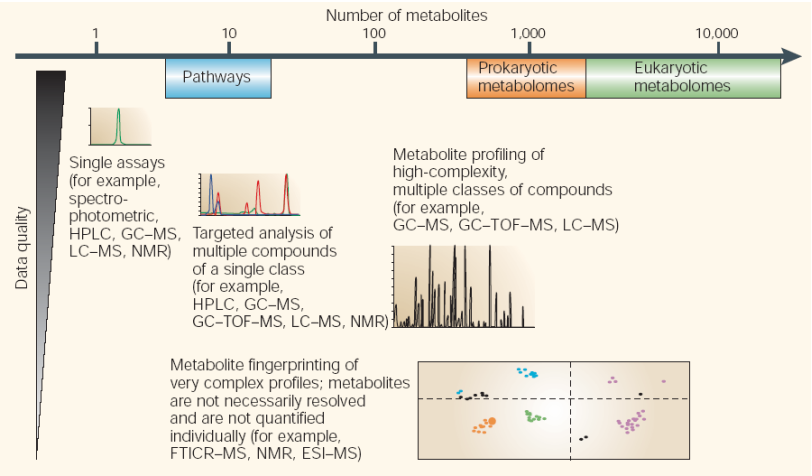
Technologies for metabolome analysis



General strategies for metabolome analysis.

CE, capillary electrophoresis; DIESI, direct-infusion ESI, which can be linked to Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS); NMR, nuclear magnetic resonance; RI, refractive index detection; UV, ultraviolet detection

Quality vs metabolic coverage



The trade-off between metabolic coverage and the quality of metabolic analysis

Gas chromatography-mass spectrometry (GCMS)

Ionization of the molecules in GCMS can be done in different ways:
- Electron ionization (positive and negative)
- Chemical ionization (positive and negative)

The fragment ions are detected by time-of-flight (TOF) or by quadruple mass spectrometry (MS). Often derivatisation methods are used to make metabolites more volatile.

GCMS mainly separates metabolites that are smaller than 500 Dalton. Separation is based on boiling point and binding to the column. Many metabolites can be identified in the GCMS, such as sugars, fatty acids, organic acids and amino acids. GC-MS is poor for the analysis of substances, which are non-volatile due to their high molecular weight and/or polarity. GCMS is suitable as a broad metabolic profiling technique.

Liquid chromatography-mass spectrometry (LCMS)

Different types of LCMS approaches:
- lipid LCMS
- ion pair LCMS
- polar (derivatised) LCMS

LCMS is the better choice for (semi) polar and non-volatile compounds. It can also be applied to profiling of polar compounds, but special (ion pair) agents need to be used or derivatisation in order to retain polar compounds on the column. In LCMS, more combinations of LC (Normal Phase, Reversed Phase, Ion Pair, Hilic,..) and MS (TOF, Ion trap, Quadrupole, FTMS instruments…) parts are available for different applications However, the identification of metabolites is more difficult than with GCMS. Often derivatisation methods are used to make metabolites better solvable. LCMS polar is a suitable technique when you would like to apply a fingerprinting procedure specifically on polar compounds. Lipid LCMS techniques are suitable as metabolic profiling techniques when you are specifically interested in lipid metabolism.

NMR-based metabolomics

What kind of samples can be analysed by NMR?

•All type of biological liquids (urine, plasma, cerebrospinal fluid, amniotic fluid, sperm, synovial fluid, saliva) or cellular or organ extracts

•All kind of biological samples such as biopsies of organs and cell cultures

Why is NMR competitive?

• NMR offers a direct biochemical window into a living system in a holistic way (no a priori selection)

• NMR is fully quantitative

• There is no need for special sample preparation (fractionation, derivatization, …)

• NMR is non-destructive and allows to completely recover the samples

• NMR has emerged into a high throughput analysis system with minimal sample preparation (cost effective)

• Nearly all metabolic intermediates have unique NMR signatures

Factors underlying the enormous challenges of metabolomics

Radius of a typical eukaryotic cell (meter)	5×10^{-6}
Volume of one cell (liter)	5×10^{-13}
Maximum number of cells in 10 g tissue	2×10^{10}
Maximum quantity (mole) of a metabolite with one copy/cell recoverable from 10 g cells/tissue	3×10^{-15}
Detection limit for MS (mole)	1×10^{-18}
Dynamic range limit for MS (factor)	1×10^{6}
Detection limit for ^{1}H NMR (mole)	1×10^{-9}
Dynamic range limit for NMR (factor)	1×10^{6}

NMR vs MS

NMR and MS dominate metabolomics research. "There is no one magic tool that can capture the diversity of composition and concentration which is present in a single sample," says Aram Adourian, senior director of advanced technologies at Beyond Genomics. "It really depends on the question you are asking. You need to have an array of tools available."

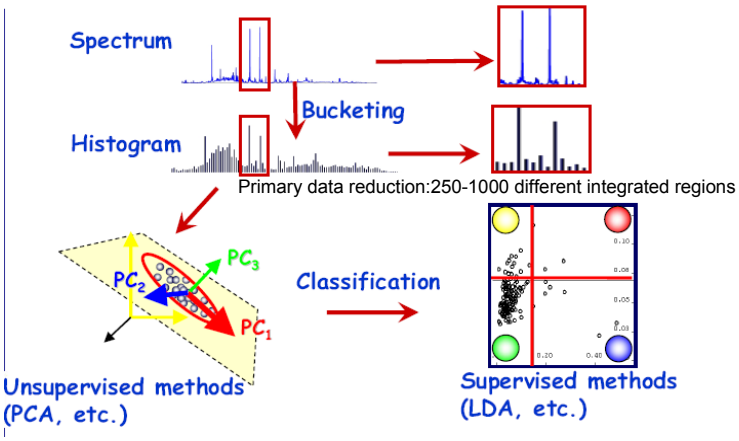
www.the-scientist.com
Amy Adams, "Metabolomics",
Volume 17 | Issue 8 | 38 | Apr. 21, 2003

Applications and examples of NMR-based metabonomics

APPLICATION	EXAMPLES
Classification of toxicity	Nephrotoxicity Hepatotoxicity Phospholipidosis Testicular toxicity Mitochondrial
Classification of disease	Inborn errors of metabolism Cancer (prostatic, brain, renal etc) Renal disease Diabetes Muscular Dystrophy
Investigation of physiological status	Diurnal variation Hormonal variation
Monitoring efficacy of therapeutic intervention	Renal transplantation (cyclosporin)
Functional genomics	Assessment of strain differences in animal models Evaluation of transgenic models
Characterisation of natural products	Batch to batch variation in commercial Feverfew

Antti et al.
http://www.acc.umu.se/~tnkjtg/Chemometrics/Editorial

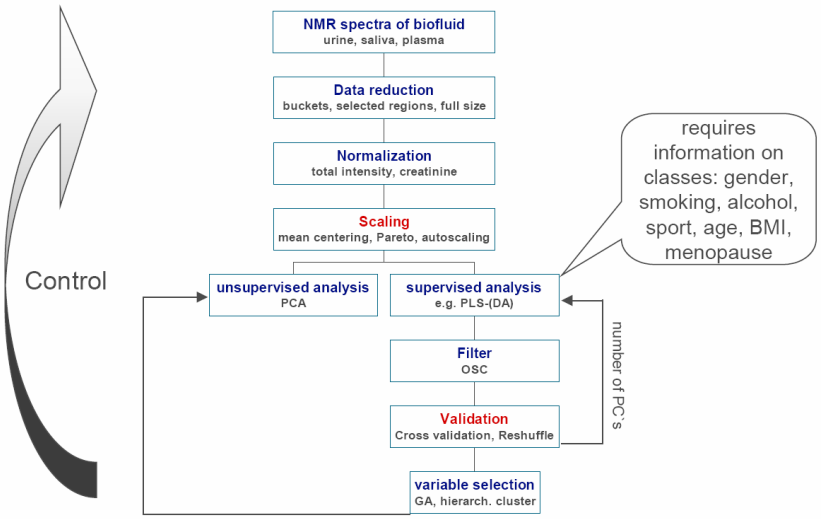
NMR-based metabolomics: the concept



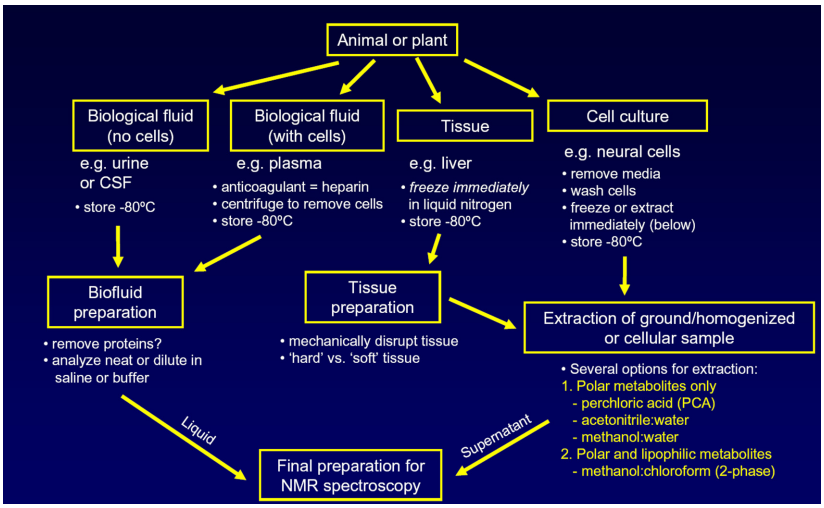
No *a priori* knowledge of the class of samples

Model for the prediction of independent data
Use class information to maximise separation among classes

NMR-based metabolomics (…)

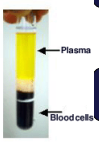


Sample preparation for NMR-based metabolomics



Plasma sample preparation

Blood collection

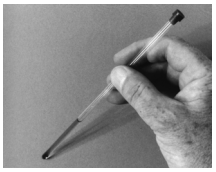


Plasma separation

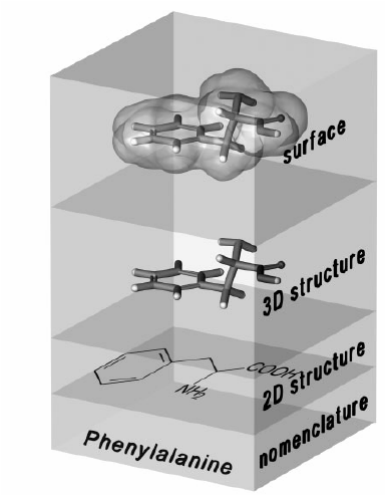
Filtration

Sample volume and pH adjustment

Ready for NMR spectroscopy



Chemical structure representations



Hierarchical scheme for representations of a molecule with different contents of structural information

Structure linear notations

The basic SMILES rules are:

1. Atoms are represented by their atomic symbols.
2. Hydrogen atoms automatically saturate free valences and are omitted (simple hydrogen connection).
3. Neighboring atoms stand next to each other.
4. Double and triple bonds are characterized by " = " and " # ", respectively.
5. Branches are represented by parentheses.
6. Rings are described by allocating digits to the two "connecting" ring atoms.

SMILES

- *Simplified Molecular Input Line Entry Specification*
- A string of letters, numbers and other characters that specify the atoms, their connectivity, bond orders, & chirality
- http://www.daylight.com/smiles/f_smiles.html

Depiction	SMILES	Name
	O	water
	C	methane
	CC(=O)O	acetic acid
	C1CCCCC1	cyclohexane
	c1ccccc1	benzene
	c1ccccc1[N+](=O)[O-]	nitrobenzene

SMILES code	*Chemical structure*	*Compound name*
Atoms: Atoms are represented by their atomic symbols. Ambiguous two-letter symbols (e.g., Nb is not NB) have to be written in square brackets. Otherwise, no further letters are used. Free valences are saturated with hydrogen atoms.		
C	CH_4	methane
[Fe+2] or [Fe++]	Fe^{2+}	iron (II) cation
Bonds: Single, double, triple, and aromatic (or conjugated) bonds are indicated by the symbols " - ", " = ", " # " and " : ", respectively; single and aromatic bonds should be omitted.		
C=C	$H_2C=CH_2$	ethene
O=CO	HCOOH	formic acid
Disconnected structures in the molecule: Individual parts of the compound are separated by a period. The period indicates that there is no connection between atoms or parts of a molecule. The arrangement of the parts is arbitrary.		
[Na+].[OH-]	NaOH	sodium hydroxide

Structure linear notations

Branches: Branches are indicated within parentheses.

CC(=O)O —  — acetic acid

CC(C)C(=O)O —  — isobutyric acid

Cyclic structures: Rings are described by breaking the ring between two atoms and then labeling the two atoms with the same number.

C1CCCCC1 —  — cyclohexane

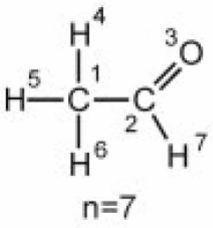
Aromaticity: Aromatic structures are indicated by writing all the atoms involved in lower-case letters.

o1cccc1 —  — furan

c12c(cccc1)cccc2 same as c1cc2ccccc2cc1 —  — naphthalene

Matrix representations

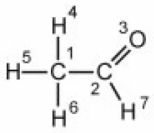
The matrix of a structure with n atoms consists of an array of $n \times n$ entries. A molecule with its different atoms and bond types can be represented in matrix form in different ways depending on what kind of entries are chosen for the atoms and bonds. Thus, a variety of matrices has been proposed: adjacency, distance, incidence, bond, and bond–electron matrices.



	1	2	3	4	5	6	7
1	0	1	0	1	1	1	0
2	1	0	1	0	0	0	1
3	0	1	0	0	0	0	0
4	1	0	0	0	0	0	0
5	1	0	0	0	0	0	0
6	1	0	0	0	0	0	0
7	0	1	0	0	0	0	0

Adiacency matrix of ethanal

Matrix representations



a)

	C1	C2	O3	H4	H5	H6	H7
C1	0	1.400	2.190	1.022	1.023	1.022	2.106
C2	1.400	0	1.123	1.999	1.982	1.999	1.022
O3	2.190	1.123	0	2.349	2.708	2.995	1.859
H4	1.022	1.999	2.349	0	1.668	1.661	2.895
H5	1.023	1.982	2.708	1.668	0	1.668	2.562
H6	1.022	1.999	2.955	1.661	1.668	0	2.336
H7	2.106	1.022	1.859	2.895	2.566	2.336	0

b)

	C1	C2	O3	H4	H5	H6	H7
C1	0	1	2	1	1	1	2
C2	1	0	1	2	2	2	1
O3	2	1	0	3	3	3	2
H4	1	2	3	0	2	2	3
H5	1	2	3	2	0	2	3
H6	1	2	3	2	2	0	3
H7	2	1	2	3	3	3	0

Distance matrices of ethanal with a) geometric distances in Å and b) topological distances. The matrix elements of b) result from counting the number of bonds along the shortest walk between the chosen atoms.

Matrix representations

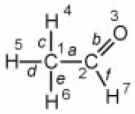
The incidence matrix is an $n \times m$ matrix where the nodes (atoms) define the columns (n) and the edges (bonds) correspond to the rows (m). An entry obtains the value of 1 if the corresponding edge ends in this particular node

	C1	C2	O3	H4	H5	H6	H7
a	1	1	0	0	0	0	0
b	0	1	1	0	0	0	0
c	1	0	0	1	0	0	0
d	1	0	0	0	1	0	0
e	1	0	0	0	0	1	0
f	0	1	0	0	0	0	1

a)

	C1	C2	O3	H4	H5	H6	H7
a	1	1					
b		1	1				
c	1			1			
d	1				1		
e	1					1	
f		1					1

b)



n=7; m=6

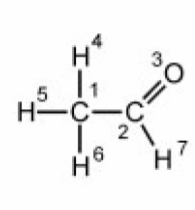
	C1	C2	O3
a	1	1	
b		1	1

c)

a) The redundant incidence matrix of ethanal can be compressed by b) omitting the zero values and c) omitting the hydrogen atoms. In the non-square matrix, the atoms are listed in columns and the bonds in rows.

Connection tables

A major disadvantage of a matrix representation for a molecular graph is that the number of entries increases with the square of the number of atoms in the molecule. What is needed is a representation of a molecular graph where the number of entries increases only as a linear function of the number of atoms in the molecule.



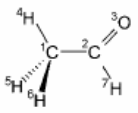
atom index	element	1st index of atom	bond order	2nd index of atom	bond order
1	C	2	1		
2	C			3	2
3	O				

File formats

Table 2-5. The most important file formats for exchange of chemical structure information.

File format	*Suffix*	*Comments*	*Support*	*Ref.*
MDL Molfile	*.mol	Molfile; the most widely used connection table format	*www.mdli.com*	50
SDfile	*.sdf	Structure-Data file; extension of the MDL Molfile containing one or more compounds	*www.mdli.com*	50
RDfile	*.rdf	Reaction-Data file; extension of the MDL Molfile containing one or more sets of reactions	*www.mdli.com*	50
SMILES	*.smi	SMILES; the most widely used linear code and file format	*www.daylight.com*	20, 21
PDB file	*.pdb	Protein Data Bank file; format for 3D structure information on proteins and polynucleotides	*www.rcsb.org*	53
CIF	*.cif	Crystallographic Information File format; for 3D structure information on organic molecules	*www.iucr.org/iucr-top/cif/*	55
JCAMP	*.jdx, *.dx, *.cs	Joint Committee on Atomic and Molecular Physical Data; structure and spectroscopic format	*www.jcamp.org/*	56
CML	*.cml	Chemical Markup Language; extension of XML with specialization in chemistry	*www.xml-cml.org*	57–59

Molfiles



1.	NSC7594 acetaldehyde	
2.	JITcc1serve09180215543D 0 0.00000 0.00000CNCI NS	Header block
3.		
4.	7 6 0 0 0 0 0 0 0 0 0999 V2000	Counts line
5.	0.0000 0.0000 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0	Atom block
6.	1.5000 0.0000 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0	
7.	2.1200 -1.0200 -0.0200 O 0 0 0 0 0 0 0 0 0 0 0 0	
8.	-0.3567 -0.4872 -0.8834 H 0 0 0 0 0 0 0 0 0 0 0 0	
9.	-0.3567 -0.5215 0.8636 H 0 0 0 0 0 0 0 0 0 0 0 0	
10.	-0.3567 1.0086 0.0198 H 0 0 0 0 0 0 0 0 0 0 0 0	
11.	2.0245 0.9324 0.0183 H 0 0 0 0 0 0 0 0 0 0 0 0	
12.	1 2 1 0 0 0 0	Bond block
13.	2 3 2 0 0 0 0	
14.	1 4 1 0 0 0 0	
15.	1 5 1 0 0 0 0	
16.	1 6 1 0 0 0 0	
17.	2 7 1 0 0 0 0	
18.	M END	Properties block

Connection table (Ctab)

Molfiles

Description	Number of atoms	Number of bonds	Number of atom lists	(obsolete)	Chiral flag	Other properties ignored for Molfiles	Number of additional properties	Current Ctab version
Column	123	456	789	1 012	345	2 3 6789012345678901234567890	123	456789
Data	7	6	0	0	0	0 0 0 0 0	999	V2000

Description	Cartesian coordinates (x, y, z)			(space)	Atom symbol	Mass difference	Charge	9 miscellaneous properties
Column	1 1234567890	2 1234567890	3 1234567890	1	234	56	789	4 012...
Data	0.0000	0.0000	0.0000		C	0	0	0...
	1.5000	0.0000	0.0000		C	0	0	0...
	2.1200	-1.0200	-0.0200		O	0	0	0...

Human Metabolome Database
Location: http://www.hmdb.ca/
Description: The Human Metabolome Database (HMDB) is a freely available electronic database containing detailed information about small molecule metabolites found in the human body. It is intended to be used for applications in metabolomics, clinical chemistry, biomarker discovery and general education. The database is designed to contain or link three kinds of data: 1) chemical data, 2) clinical data, and 3) molecular biology/biochemistry data. The database currently contains nearly 2500 metabolite entries including both water-soluble and lipid soluble metabolites as well as metabolites that would be regarded as either abundant (> 1 uM) or relatively rare (< 1 nM). Additionally, approximately 5500 protein (and DNA) sequences are linked to these metabolite entries. Each MetaboCard entry contains more than 90 data fields with half of the information being devoted to chemical/clinical data and the other half devoted to enzymatic or biochemical data. Many data fields are hyperlinked to other databases (KEGG, PubChem, MetaCyc, ChEBI, PDB, Swiss-Prot, and GenBank) and a variety of structure and pathway viewing applets. The HMDB database supports extensive text, sequence, chemical structure and relational query searches.

ChemSpider
Location: http://www.chemspider.com/
Decription: ChemSpider is a free access service providing a structure centric community for chemists. Providing access to millions of chemical structures and integration to a multitude of other online services ChemSpider is the richest single source of structure-based chemistry information.

KEGG: Kyoto Encyclopedia of Genes and Genomes
Location: http://www.genome.jp/kegg/
Description: The goal of this website is to build a bioinformatics resource as complete computer representation of the cell, the organism, and the biosphere, which will enable computational prediction of higher-level complexity of cellular processes and organism behaviors from genomic and molecular information. For metabolites especially http://www.genome.ad.jp/kegg/ligand.html, which includes possibilities to search for chemical formula, name, exact mass, and pathway.

METLIN Metabolite Database
Location: http://metlin.scripps.edu/
Description: METLIN is a metabolite database for metabolomics containing over 15,000 structures, it is also represents a data management system designed to assist in a broad array of metabolite research and metabolite identification by providing public access to its repository of current and comprehensive mass spectral metabolite data.

ChEBI
Location: http://www.ebi.ac.uk/chebi/
Description: Chemical Entities of Biological Interest is a freely available dictionary of molecular entities focused on ‘small’ chemical compounds. The term ‘molecular entity’ refers to any constitutionally or isotopically distinct atom, molecule, ion, ion pair, radical, radical ion, complex, conformer, etc., identifiable as a separately distinguishable entity. The molecular entities in question are either products of nature or synthetic products used to intervene in the processes of living organisms.

PubChem
Location: http://pubchem.ncbi.nlm.nih.gov/
Description: PubChem provides information on the biological activities of small molecules. Including:
PubChem Compound: Search unique chemical structures using names, synonyms or keywords. Links to available biological property information are provided for each compound.
PubChem Substance: Search deposited chemical substance records using names, synonyms or keywords. Links to biological property information and depositor web sites are provided.
PubChem BioAssay: Search bioassay records using terms from the bioassay description, for example "cancer cell line". Links to active compounds and bioassay results are provided.
Structure Search: Search PubChem's Compound database using a chemical structure as the query. Structures may be sketched or specified by SMILES, MOL files, or other formats.

Comparative Toxicogenomics Database (CTD)
Location: http://ctd.mdibl.org/
Description: The Comparative Toxicogenomics Database (CTD) elucidates molecular mechanisms by which environmental chemicals affect human disease.
Chemical–gene/protein interactions and chemical– and gene–disease relationships are curated from the published literature, and integrated with diverse data (chemicals, genes/proteins, human diseases, references, sequences, vertebrate and invertebrate organisms, and the Gene Ontology) to facilitate environmental health research.

Lipid Maps
Location: http://www.lipidmaps.org/
Description: This website is focussed on the lipid section of the metabolome by developing an integrated metabolomic system capable of characterizing the global changes in lipid metabolites ("lipidomics"). All enzymes and other proteins involved in lipid metabolism will be mapped and integrated into the networks of lipidomics.

Lipid Library
Location: http://www.lipidlibrary.co.uk/
Description: Includes: definitions, structures, composition, occurrence, biochemistry and functions of most types of fatty acids and lipids.
Comments: This website is regularly updated also with the most recent literature on the analysis of lipids and many practical tips can be found. The site is really a must for lipid scientists.

PRIMe
Location: http://prime.psc.riken.jp/
Description: RIKEN Metabolomics platform, is a Web-based service for metabolomics and transcriptomics as systems for understanding life. This project measures standard metabolites by means of multi-dimensional NMR spectroscopy, GC/MS, LC/MS, and CE/MS. We also provide unique tools for metabolomics, transcriptomics, and integrated analysis of a range of other "-omics" data.

MassBank
Location: www.massbank.jp
Desciption: This site presents the database of comprehensive, high-resolution mass spectra of metabolites. Supported by the JST-BIRD project, it offers various query methods for standard spectra from Keio Univ., RIKEN PSC, and others.