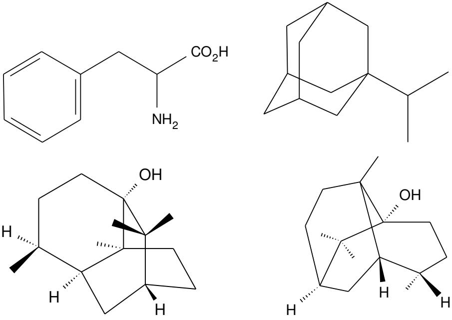
Ring Perception

Rings have profound influence on molecular properties:

1. small rings introduce strain into a molecule.
2. aromatic rings change physico-chemical properties.
3. rings present particular problemes in synthesis.



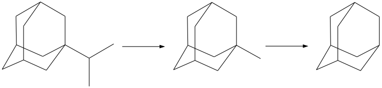
Downs GM et al., (1989), *J Chem Inf Comput Sci*, 29(3):172-187. DOI:10.1021/ci00063a007

Methods for Ring Perception

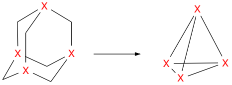
1. graph-theory based methods.
   - depth-first search: to find all cycles.
   - breadth-first search: fast for smallest cycles.
2. linear algebra based methods.
   - manipulation of incidence or adjacency matrix.
   - fundamental cycle basis.

Pre-processing of molecular graph:

1. Iteratively remove all nodes with degree 1 (resulting in the ring skeleton).



2. Merge ring nodes of degree 2 with corresponding neighbores (resulting in basic graph with fewer nodes).

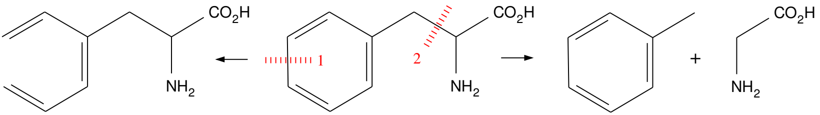


Facts on Cycles

The minimal number of cycles is given by the nullity μ

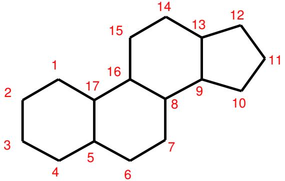
$$\mu = b - a + c \qquad \text{with } b = \text{bonds, } a = \text{atoms, c} = \text{compounds}$$

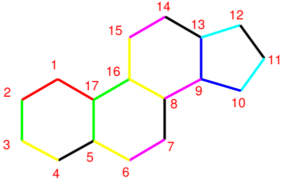
the number of edges which need to be broken to turn a cyclic into an acyclic graph. (Remember `SMILES` string generation!)

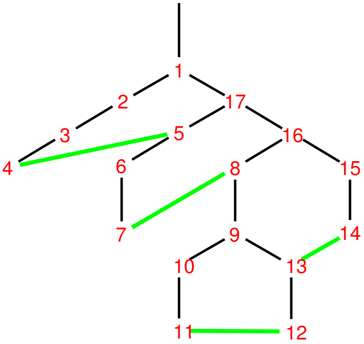


μ is also the cardinality (size) of the fundamental cycle basis.

Walk based Algorithms







1. Grow a spanning tree and remember *ring closure bonds*.
2. Walk from ring closure bonds towards root to common atom.
3. Add ring bonds to tree and do 2 to find ring systems.

Shelley CA, (1983) **Heuristic Approach for displaying chemical structures**, *J Chem Inf Comput Sci* **23** (2):61-65 |
DOI: 10.1021/ci00038a002

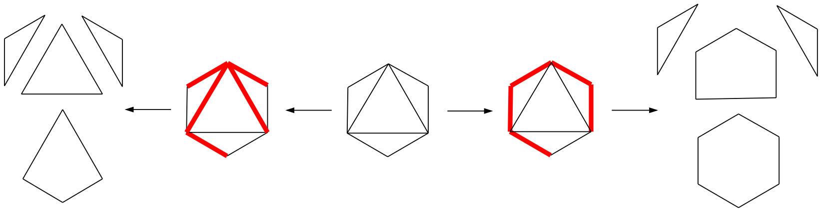
Spanning Trees and Cycles

Each spanning tree T has $V - 1$ edges.

Adding an edge $E \setminus T$ to the spanning tree T will create a cycle.

Such a cycle is called a fundamental cycle (wrt. to T).

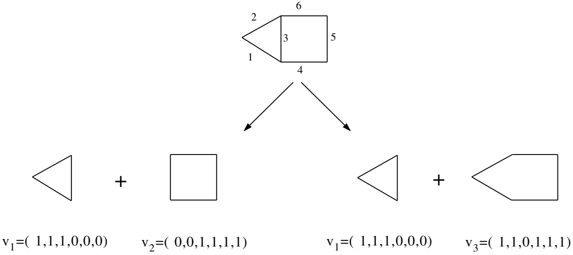
A graph with E edges has $E - V + 1$ fundamental cycles.



The set of fundamental cycles of any spanning tree T forms a basis for the cycle space.

Fundamental Cycle Basis

Fundamental cycle bases are not generally unique but they always contain μ cycles.



Cycles can be represented by edge-incidence vectors in $\{0,1\}^{|E|}$.

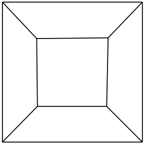
All cycles can be derived from the fundamental cycle basis by "`XOR`-ing" 1 to μ edge-incidence vectors from the cycle basis.

Ring Sets

Ring set	Contents
All cycles	all (simple) cycles
Beta-ring	3- and 4-edged simple cycles + linear independent from 3 ore more smaller beta-rings
ESER	heuristic selection of smallest simple cycles
Essential cycles	Intersection of all SSSR
ESSR	all simple faces and primary/secondary cut faces
K-rings/relevant cycles	Union of all SSSR
SER	K-rings + simple cycles that are fusions of pairs of them
Faces	simple faces faces
SSSR	μ smallest simple cycles (minimal cycle basis)

Which Ring set to choose?

1. Should in some way be "optimal" for the particular application.
2. Should be unique for a given structure.
3. Shoud be invariant (e.g. the processing order).
4. Should include a minimal and sufficent number of rings to describee the ring system.

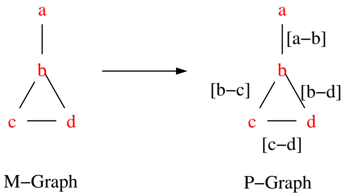


- 28 cycles ($6 \times$ 4-ring, $16 \times$ 6-ring, $6 \times$ 8-ring).
- 14 simple cycles.
- 6 simple faces.
- $\mu = 5$.

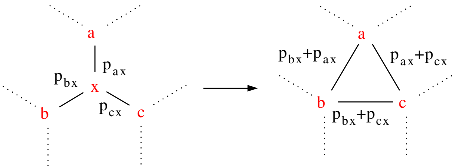
Just 4 of the simple faces cover already all edges and vertices.

Hanser Algorithm: exhaustive ring perception

1. Convert molecular graph to path graph

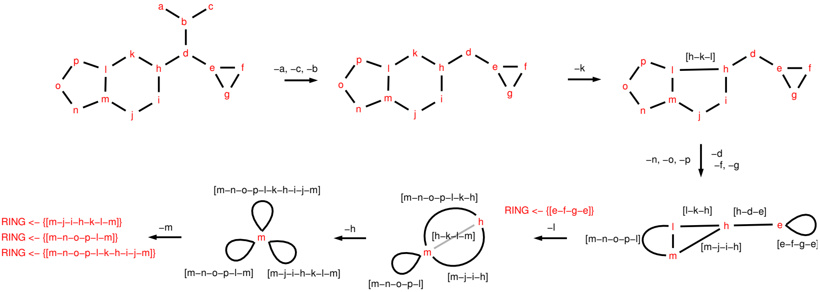


2. Perform graph reduction



Hanser T, Jauffret P, Kaufmann G, (1996), *J Chem Inf Comput Sci*, 36(6):1146-1152. DOI:10.1021/ci960322f

Example: Hanser at work



Function `Rings`(M-Graph):
$Rings \leftarrow \emptyset$
`Convert`(M-Graph, V, E)
while $V \neq \emptyset$ **do**
  choose $x \in V$
  `Remove`(x, V, E, $Rings$)
end while
Function `Convert`(M-Graph, V, E):
$V \leftarrow \emptyset, E \leftarrow \emptyset$
for all $x \in V$ **do**
  $v \leftarrow v \cup \{x\}$
end for
for all edges $(x, y) \in E$ **do**
  $p_{xy} \leftarrow (x, y)$
  $E \leftarrow E \cup \{p_{xy}\}$
end for

Function `Remove`(x, V, E, $Rings$):
for all paths $p_{xy}, p_{xz} \in E \times E$ **do**
  if $p_{xy} \otimes p_{xz} = \{x\}$ **then**
    $p_{yz} \leftarrow p_{xy} \oplus p_{xz}$
    $E \leftarrow E \cup \{p_{yz}\}$
  end if
end for
for all $p_{xy} \in E$ **do**
  if $x = y$ **then**
    $Rings \leftarrow Rings \cup \{p_{xy}\}$
    $E \leftarrow E - \{p_{xy}\}$
  end if
end for
$V \leftarrow V - \{x\}$

Ring Sets

A comparative tabulation of ring sets for the DBR database (based on a tabulation by Nickelsen[32]) is given in Table III. Block 1 gives the DBR structure with its number and nullity. Block 2 gives the cycle sizes and types present. Block 3 gives the comparative perception of type for each ring set. Block 4 gives the number of rings found of each type. Block 5 gives comments about particular types, where necessary. Other symbols are μ = nullity, G = ring size, e = simple cycle, β = β-ring, $\mathcal{S}$ = SSSR, $\mathcal{K}$ = $\mathcal{K}$-ring, $\mathcal{E}$ = ESSR, ✓ = included in ring set, and × = not included in ring set. The comment codes given in the note column are F = simple face, A = SSSR has to choose arbitrarily between symmetrical equivalents, D = Doppelpunkte exclude cycle as a simple cycle, I = sym-

Downs GM et al., (1989), *J Chem Inf Comput Sci*, 29(3):172-187. DOI:10.1021/ci00063a007

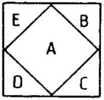
Ring Sets

1	2		3					4				5
Structure diagram	Cycle type	G	e	β	$\mathcal{S}$	$\mathcal{K}$	$\mathcal{E}$	β	$\mathcal{S}$	$\mathcal{K}$	$\mathcal{E}$	Note
DBR-1 $\mu = 2$ 	A,B	5	√	√	√	√	√	1	1	1	1	(F)
	A+B	6	√	√	×	×	√	1	0	0	1	(M)
DBR-2 $\mu = 2$ 	A,B	6	√	√	√	√	√	1	1	1	1	(F)
	A+B	6	√	√	×	×	√	1	0	1	1	(M)
DBR-3 $\mu = 4$ 	A,B,C,D	3	√	√	√	√	√	4	4	4	4	(F)
	A+B etc.	4	×	×	×	×	×	0	0	0	0	(N)
	A+B+C+D	4	√	√	×	×	√	1	0	0	1	(M)
	A+C etc.	6	×	×	×	×	×	0	0	0	0	(D)

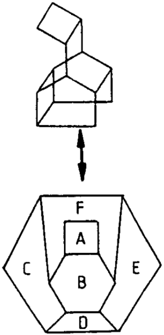
Ring Sets

1	2		3					4				5
Structure diagram	Cycle type	G	e	β	$\mathcal{S}$	$\mathcal{K}$	$\mathcal{E}$	β	$\mathcal{S}$	$\mathcal{K}$	$\mathcal{E}$	Note
DBR-10 $\mu = 7$												
	A, …,F	4	√	√	√	√	√	6	6	6	6	(F)
	G	6	√	√	√	√	√	1	1	1	1	(F)
	A+B+···+G	6	√	√	×	√	√	1	0	1	1	(I)
	A+B etc.	6	×	×	×	×	×	0	0	0	0	(N)
	A+B+G etc.	8	√	×	×	×	×	0	0	0	0	(S)
	A+B+C+G etc.	8	√	×	×	×	×	0	0	0	0	(S)

Ring Sets

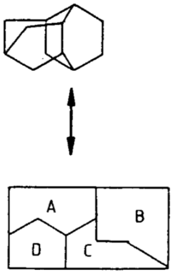
1	2		3					4				5
Structure diagram	Cycle type	G	e	β	$\mathcal{S}$	$\mathcal{K}$	$\mathcal{E}$	β	$\mathcal{S}$	$\mathcal{K}$	$\mathcal{E}$	Note
DBR-14 $\mu = 5$ 	B,C,D,E	3	√	√	√	√	√	4	4	4	4	(F)
	A	4	√	√	√	√	√	1	1	1	1	(F)
	A+B+C+D+E	8	×	×	×	×	×	0	0	0	0	(N)

Ring Sets

1	2		3					4				5
Structure diagram	Cycle type	G	e	β	$\mathcal{S}$	$\mathcal{K}$	$\mathcal{E}$	β	$\mathcal{S}$	$\mathcal{K}$	$\mathcal{E}$	Note
DBR-26 $\mu = 6$ 												
	A,D	4	√	√	√	√	√	2	2	2	2	(F)
	C,E	5	√	√	√	√	√	2	2	2	2	(F)
	B	6	√	√	√	√	√	1	1	1	1	(F)
	A+B+···+F	6	√	√	√	√	√	1	1	1	1	(O)
	B+C+D+E	6	√	√	×	√	√	1	0	1	1	(R)
	B+A+F	6	√	√	×	√	√	1	0	1	1	(P)
	A+B+C+F, A+B+E+F	7	×	×	×	×	×	0	0	0	0	(S)
	A+B+C+D+F, A+B+D+E+F	7	×	×	×	×	×	0	0	0	0	(S)
	F	8	×	×	×	×	×	0	0	0	0	(N)
	B+C+D, B+D+E	9	×	×	×	×	×	0	0	0	0	(S)

Ring Sets

DBR-28 $\mu = 4$



1	2		3					4				5
Structure diagram	Cycle type	G	e	β	$\mathcal{S}$	$\mathcal{K}$	$\mathcal{E}$	β	$\mathcal{S}$	$\mathcal{K}$	$\mathcal{E}$	Note
DBR-28 $\mu = 4$	D	5	√	√	√	√	√	1	1	1	1	(F)
	A,B,C	6	√	√	√	√	√	3	3	3	3	(F)
	B+C	6	√	√	×	√	√	1	0	1	1	(P)
	A+B+C+D	7	√	×	×	×	√	0	0	0	1	(M)
	A+D	7	√	√	×	×	×	1	0	0	0	(S)
	A+B+C	8	×	×	×	×	×	0	0	0	0	(S)
	A+C+D	9	×	×	×	×	×	0	0	0	0	(S)
	C+D, B+C+D	9	×	×	×	×	×	0	0	0	0	(N)
	A+B, A+C	10	×	×	×	×	×	0	0	0	0	(N)
	A+B+D	11	×	×	×	×	×	0	0	0	0	(N)

SSSR Algorithmic Approach 1: Horton

- compute a sufficiently large set of cycles
- sort them by weight
- initialize B to empty set
- go through the cycles C in order of increasing weight
- add C to B if is independent of B
- use Gaussian elimination to decide independence
- in order to make the approach efficient, one needs to identify a small set of cycles which is guaranteed to contain a minimum basis

  Horton set: for any edge $e = (a, b)$ and vertex v take the cycle $C_{e,v}$ consisting of e and the shortest paths from v to a and b. $O(nm)$ cycles, Gaussian elimination, running time $O(nm^3)$

Horton, J.D.: *A polynomial-time algorithm to find the shortest cycle basis of a graph.*

SIAM J.C. 16(2), 358-366 (1987)

SSSR Algorithmic Approach 1: Horton

- compute a sufficiently large set of cycles
- sort them by weight
- initialize B to empty set
- go through the cycles C in order of increasing weight
- add C to B if is independent of B
- use Gaussian elimination to decide independence
- in order to make the approach efficient, one needs to identify a small set of cycles which is guaranteed to contain a minimum basis
  Horton set: for any edge $e = (a, b)$ and vertex v take the cycle $C_{e,v}$ consisting of e and the shortest paths from v to a and b.
  $O(nm)$ cycles, Gaussian elimination, running time $O(nm^3)$

Horton, J.D.: *A polynomial-time algorithm to find the shortest cycle basis of a graph*.

SIAM J.C. 16(2), 358-366 (1987)

SSSR Algorithmic Approach 2: de Pina

- construct basis iteratively, assume partial basis is $\{C_1, \ldots, C_i\}$
- compute a vector S orthogonal to $C_1, \ldots, C_i$.
- find a cheapest cycle C having a non-zero component in the direction S, i.e., $\langle C, S \rangle = 0$
- add C to the partial basis
- C is not the cheapest cycle independent of the partial basis
- it is the shortest vector with a component in direction S.
- correct

De Pina, J.C.: *Applications of shortest path methods*. Ph.D. thesis, University of Amsterdam (1995)

SSSR / MCB

Type	Authors	Approach	Running time
undirected	Horton, 87	Horton	$O(m^3 n)$
	de Pina, 95	de Pina	$O(m^3 + mn^2 \log n)$
	Golinsky/Horton, 02	Horton	$O(m^\omega n)$
	Berger/Gritzmann/de Vries, 04	de Pina	$O(m^3 + mn^2 \log n)$
	Kavitha/Mehlhorn/Michail/Paluch, 04	de Pina	$O(m^2 n + mn^2 \log n)$
	Mehlhorn/Michail, 07	Horton-Pina	$O(m^2 n / \log n + mn^2)$
directed	Kavitha/Mehlhorn, 04	de Pina	$O(m^4 n)$ det, $O(m^3 n)$ Monte Carlo
	Liebchen/Rizzi, 04	Horton	$O(m^{1+\omega} n)$
	Kavitha, 05	de Pina	$O(m^2 n \log n)$ Monte Carlo
	Hariharan/Kavitha/Mehlhorn, 05	de Pina	$O(m^3 n + m^2 n^2 \log n)$
	Hariharan/Kavitha/Mehlhorn, 06	de Pina	$O(m^2 n + mn^2 \log n)$ Monte Carlo
	Mehlhorn,Michail 07	Horton-Pina	$O(m^3 n)$ det, $O(m^2 n)$ Monte Carlo

open problem: faster algorithms

Summary of Properties

Ring set	Type	Unique	Contains basis	Contains MCB	Contains $\mathcal{R}(G)$	Size
SSSR	G	No	Yes	MCB	No	$\mathcal{O}(m)$
ESER/DESER	G	Yes	No	No	No	exp
Faces(G)	P	Yes	Yes	No	No	exp
ESSR	P	Yes	Yes	No	No	exp
SSCE	G	Yes	No	No	No	exp
β-rings	P	No	Yes	No	No	$\mathcal{O}(m + n^4)$
SER	G	No	Yes	Yes	No	exp
Elementary cycles	G	Yes	Yes	Yes	Yes	exp
$\mathcal{K}$-rings Relevant Cycles	G	Yes	Yes	Yes	Yes	exp