Organic chemistry for bioinformatics

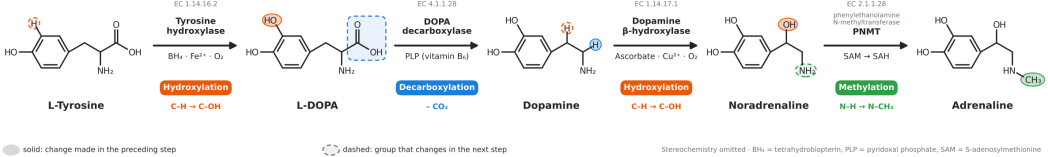
Part 1: From atoms to molecules

After today you can

- say how many bonds H, C, N, O, S (and P) can form
- write a molecule in four notations and read them
- explain why carbon is tetrahedral and why that matters for drawing
- recognise isomers and find a chiral carbon



Biosynthesis of adrenaline from tyrosine



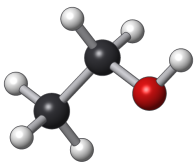
solid: change made in the preceding step

dashed: group that changes in the next step

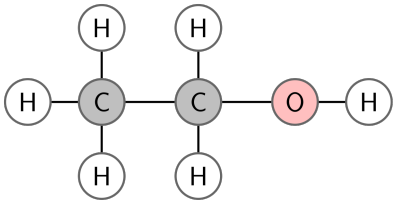
Stereochemistry omitted · BH_4 = tetrahydrobiopterin, PLP = pyridoxal phosphate, SAM = S-adenosylmethionine

Molecules are graphs

- atom = node, labelled with the element
- bond = edge
- every element allows only a fixed number of edges per node: its valency
- valid molecule = all node degrees correct
- chemistry = which graphs are allowed + what they look like in 3D



ball-and-stick (3D)



labelled graph

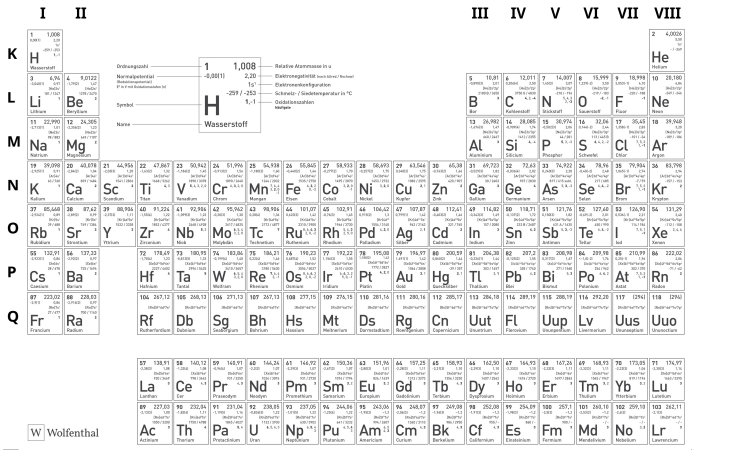
CCO

a text format for graphs
(SMILES, lecture 2)

Periodic table of the elements

- ordered by atom number
- row (“period”) = shell;

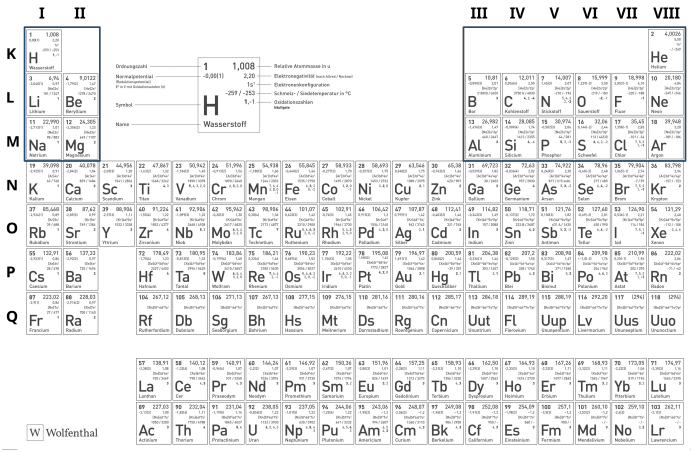
  max. number of e^- in the outer shell
- column (“(main) group”) = number of e^- in the outer shell
- same group → approx. similar properties



Periodic table of the elements

- ordered by atom number
- row (“period”) = shell;

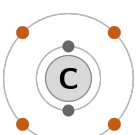
  max. number of e^- in the outer shell
- column (“(main) group”) = number of e^- in the outer shell
- same group $\rightarrow$ approx. similar properties



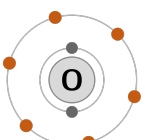
The six atoms of biochemistry: H, C, N, O, P, S

- these six make up roughly 99 % of the atoms in a human body
- valence electrons = electrons in the outermost shell; they form the bonds
- the group number tells you the valence electrons:
  - groups 1–2 directly,
  - groups 13–18 the last digit
- octet rule: atoms "want" 8 outer electrons (hydrogen: 2)

group	1	2	13	14	15	16	17	18
valence e⁻	1	2	3	4	5	6	7	8
	H							He
	Li	Be	B	**C**	**N**	**O**	F	Ne
	Na	Mg	Al	Si	**P**	**S**	Cl	Ar



C: 2 | 4



O: 2 | 6

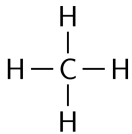
How many bonds can an atom form? The valency rule

Atom	Group	Valence e^-	**Bonds**	Lone pairs
H	1	1	**1**	0
C	14	4	**4**	0
N	15	5	**3**	1
O	16	6	**2**	2
S	16	6	**2**	2
P	15	5	**3 or 5**	1 or 0

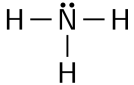
Rule: bonds $= 8 -$ valence electrons (H: $2 - 1$)

Exception: S and P (3rd period) can form more bonds.

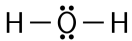
- S mostly 2 bonds;
- P nearly always occurs as phosphate with 5 bonds (one P=O, three P–O).



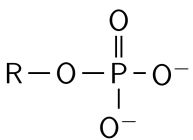
CH_4



NH_3



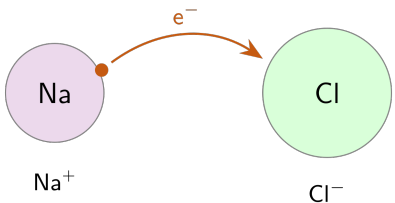
H_2O



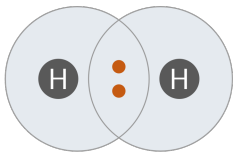
phosphate group (as in DNA, ATP)

Sharing or transferring electrons: covalent and ionic bonds

- ionic: one atom gives electrons to another; the ions attract (Na^+ Cl^-)
- covalent: atoms share an electron pair (H–H); all bonds inside organic molecules
- biochemistry has both: covalent skeleton plus charged groups in water (COO^-, NH_3^+, phosphate)



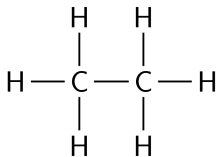
ionic



covalent

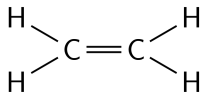
Single, double and triple bonds

- one bond = one shared electron pair; atoms can share 1, 2 or 3 pairs
- carbon always has 4 bonds in total: C–C, C=C, C≡C
- count the bonds of each carbon atom separately
- double bonds also occur with O and N (C=O, C=N): lecture 2



ethane

1+1+1+1 = 4



ethene

2+1+1 = 4

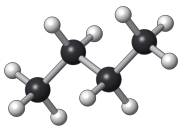


ethyne

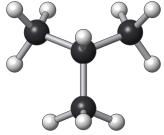
3+1 = 4

Hydrocarbons: only C and H

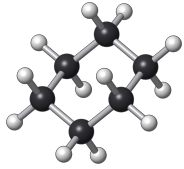
- saturated: only single bonds (alkanes, C_nH_{2n+2})
- unsaturated: at least one C=C or C≡C
- structures: linear, branched, ring
- every ring or double bond "costs" two H: C_6H_{14} (hexane) → C_6H_{12} (cyclohexane or hexene)
- names: meth-, eth-, prop-, but-, pent-, hex-



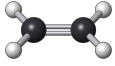
butane, C_4H_{10}
linear



isobutane, C_4H_{10}
branched



cyclohexane, C_6H_{12}
ring



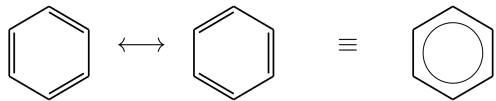
ethene, C_2H_4
double bond

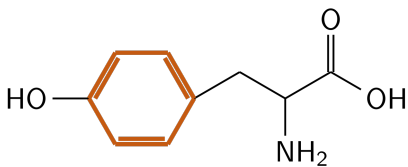


ethyne, C_2H_2
triple bond

Aromatic rings

- benzene, C_6H_6: six carbons in a ring with alternating double and single bonds, but all six bonds are equal (edges are 1.5 bonds)
- unusually very stable, does not react like a normal C=C
- very common in biochemistry: tyrosine, phenylalanine, tryptophan, DNA bases, adrenaline
- for now: recognise the hexagon





tyrosine, the aromatic ring highlighted

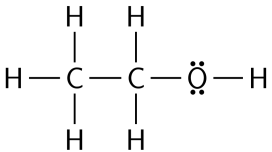
Notation 1: Molecular formula (Summenformel)

- atoms and their counts: C_2H_6O
- convention: C first, then H, then the others alphabetically
- tells you the composition only, nothing about connections
- question: how many different molecules have the formula C_2H_6O?

$$C_2H_6O\ ?$$

Notation 2: Structural formula (Lewis / Valenzstrichformel)

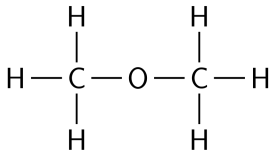
- every atom written, every bond a line (double bond = two parallel lines)
- lone pairs as lines or two dots (often omitted)
- check: does every atom have the right number of bonds?
- same formula, different structure = constitutional isomers



ethanol, C_2H_6O

lone pairs on O shown

liquid, boiling point 78 °C



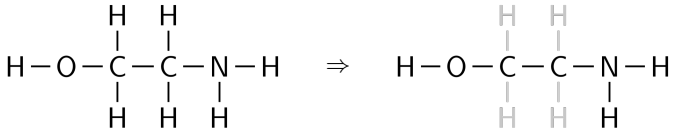
dimethyl ether, C_2H_6O

lone pairs on O omitted

gas, boiling point −24 °C

Notation 3: Skeletal formula (Skelettformel)

1. bonds are lines; a double bond is two parallel lines
2. every line end and every corner is a carbon atom (C not written)
3. hydrogens on carbon are not written
4. all other atoms (N, O, S, P . . .) are written, together with the hydrogens attached to them



(1) full Lewis structure (2) H on carbon greyed out

(3) carbon labels greyed out (4) **skeletal formula**

2-aminoethanol, C_2H_7NO

Reading skeletal formulas: a quick test

How many C, H, N, O atoms does this molecule have?

- task: write the molecular formula
- tip:

  number of C = corners + ends

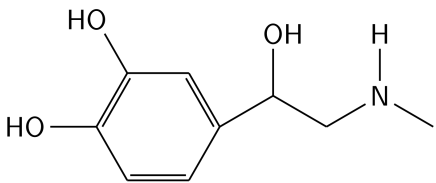
  number of H =

  = sum of

  (4 − number of bonds per carbon)

  over all carbons +

  + number of H (on N and O)



Reading skeletal formulas: a quick test

How many C, H, N, O atoms does this molecule have?

- task: write the molecular formula
- tip:

  number of C = corners + ends

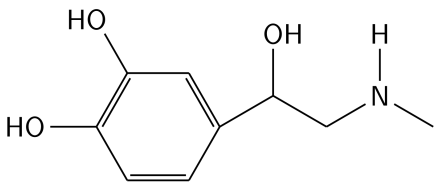
  number of H =

  = sum of

  (4 − number of bonds per carbon)

  over all carbons +

  + number of H (on N and O)



adrenaline: you build it in lecture 3

solution: $C_9H_{13}NO_3$

Exercise: build it, then write it

1. In pairs, build a molecule with the model kit using C, N, O, S and H.
   Constraints: all valencies correct (no free holes, no extra bonds); N, O and S must not be bonded to each other.
   Bonus: include one double bond or one ring.
2. Write it as: molecular formula, Lewis formula, skeletal formula.
3. Swap drawings with the next pair: can they rebuild your molecule from the drawing? Check the valencies together.

Wedge-dash, the 4th notation, follows in part D.

Colour key of the kit



C: 4 holes



H: 1 hole



O: 2 holes



N: 3 holes

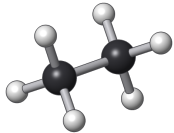


S: 2 holes

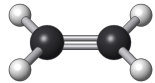
springs: double bonds

10 min: 4 build · 4 write · 2 swap

Geometry I: bond length, bond energy, bond angle



ethane: about 109.5°



ethene: about 120°, planar



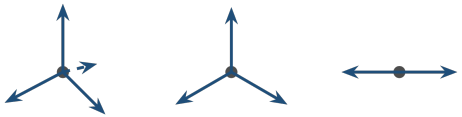
ethyne: 180°, linear

	Bond	Length	Energy (kJ/mol)	Angle
ethane	C–C	154 pm	347	about 109.5°
ethene	C=C	134 pm	614	about 120°
ethyne	C≡C	120 pm	839	180°

- more shared electron pairs: shorter and stronger bond
- the angle depends on the number of neighbours of the carbon: 4 / 3 / 2

Why these angles? Electron pairs repel

- bonds and lone pairs are clouds of negative charge: they get as far apart as possible (VSEPR)
- 4 neighbours: tetrahedral, 109.5° (CH_4)
- 3 neighbours: planar, 120° (ethene)
- 2 neighbours: linear, 180° (ethyne, CO_2)
- lone pairs count too: NH_3 107°, H_2O 104.5°
- you may hear sp^3 / sp^2 / sp: these are just names for 4 / 3 / 2 neighbours



4: tetrahedral 3: planar 2: linear



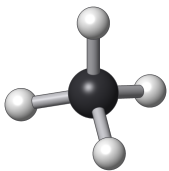
NH_3: 107°
1 lone pair

H_2O: 104.5°
2 lone pairs

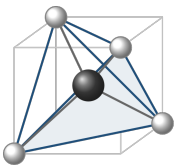
VSEPR ... Valence Shell Electron Pair Repulsion (VSEPR) theory

The tetrahedral carbon

- methane: carbon in the centre, hydrogens at the corners of a tetrahedron, all angles 109.5°
- every carbon with four single bonds has this shape
- a flat drawing hides this, so we need a notation for 3D



rotate it live: MolView / PubChem 3D viewer



H on alternate corners of a cube

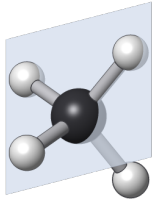


the kit model

Notation 4: Wedge-dash (Keilstrichformel)

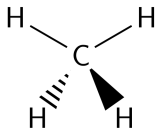
- plain line: bond in the plane of the paper
- solid wedge: bond pointing towards you
- dashed wedge: bond pointing away from you
- typical: two bonds in the plane, one wedge, one dash

Mini-task (2 min): take one carbon with four substituents from your molecule and redraw it with wedge-dash. (No such carbon? Use the CH_2 of ethanol.) The rest is homework.

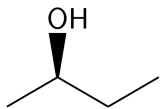


model with paper plane

$\Rightarrow$



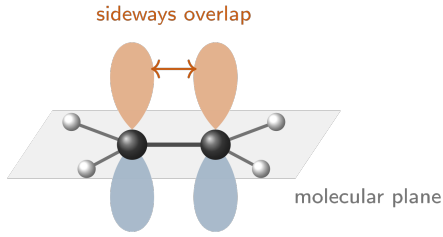
wedge-dash

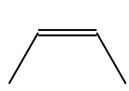


2-butanol in skeletal wedge-dash

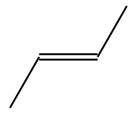
Double bonds are rigid

- single bond: the two ends can rotate like an axle
- double bond: no rotation; all six atoms of ethene lie in one plane
- reason: the second bond is formed by sideways overlap of two p orbitals; twisting would break it
- consequence: cis- and trans-but-2-ene are different molecules



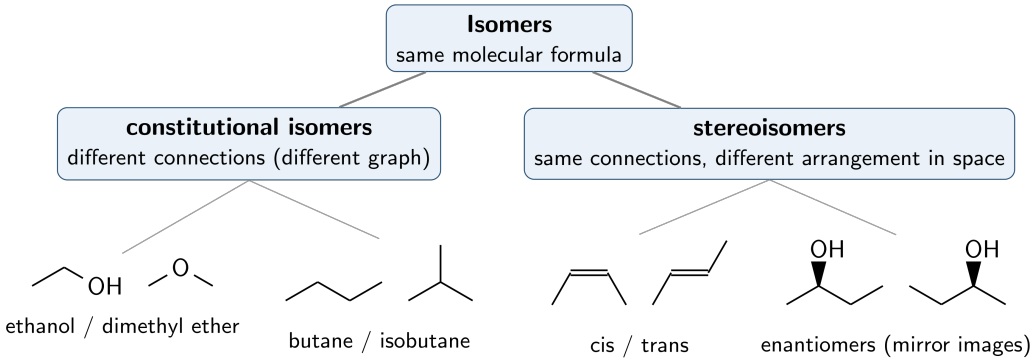


cis-but-2-ene



trans-but-2-ene

Isomers: same formula, different molecules

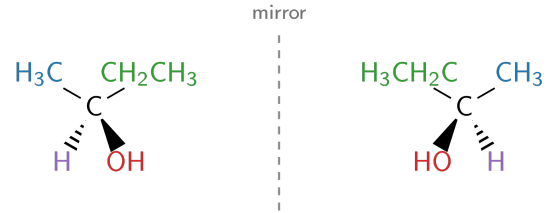


Rotation around single bonds does not make a new isomer.

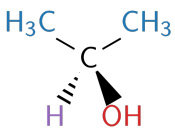
Chirality: carbons with four different substituents

- chiral = not superimposable on its mirror image (like your hands)
- a carbon with four different substituents is a chiral centre
- the two mirror images are enantiomers
- two identical substituents: not chiral





2-butanol: two enantiomers



propan-2-ol: two identical CH_3 $\rightarrow$ **not chiral**

Hands-on: build the mirror image

1. Build 2-butanol (central C with H, OH, CH_3, CH_2CH_3).
2. Build its mirror image next to it.
3. Try to superimpose them: can you rotate one so every atom matches?
4. Control: do the same with propan-2-ol.

Question: how many different arrangements of four different groups around a tetrahedral carbon exist?



Hands-on: build the mirror image

1. Build 2-butanol (central C with H, OH, CH_3, CH_2CH_3).
2. Build its mirror image next to it.
3. Try to superimpose them: can you rotate one so every atom matches?
4. Control: do the same with propan-2-ol.

Question: how many different arrangements of four different groups around a tetrahedral carbon exist?

Two (the two enantiomers)

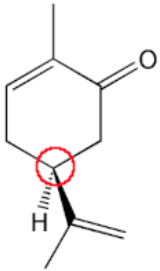


Same bonds, different smell: carvone

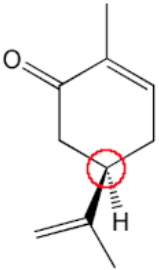
- same formula, same boiling point, same density; but the nose is made of chiral receptors and tells them apart



Kuemmel



(S)-Carvon



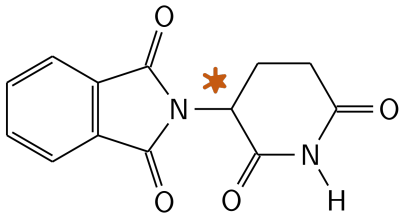
(R)-Carvon



Minze

Same bonds, different effect: Contergan (thalidomide)

- sedative and anti-nausea drug (Contergan, Grünenthal), on sale in Germany from 1957, withdrawn in 1961
- taken in pregnancy: severe malformations in roughly 10 000 children worldwide (estimates vary)
- thalidomide is chiral; the popular story says one enantiomer is sedative and the other harmful
- but the enantiomers interconvert in the body, so a pure enantiomer would not have prevented the damage
- lesson: chirality matters for drugs, and so does careful testing
- today thalidomide is used again under strict control (e.g. multiple myeloma)

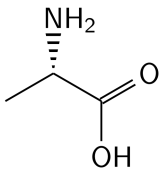


thalidomide; * = stereocentre

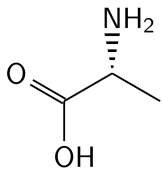


Life is homochiral

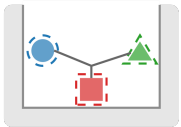
- proteins are built from L-amino acids only; the sugars in glucose and in DNA/RNA are D
- enzymes are chiral: they recognise one enantiomer
- lecture 3: every product of the adrenaline pathway is a single enantiomer, even though the pictures leave out the 3D detail
- D and L are older labels; R and S follow in lecture 2



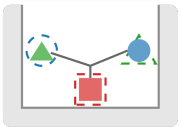
L-alanine



D-alanine



fits



mirror image: no fit

Summary and outlook

- molecules are graphs with valency constraints: H 1, C 4, N 3, O 2, S 2, P 3 or 5
- notations: formula (counts), Lewis (graph), skeletal (compact), wedge-dash (3D)
- double bonds are rigid
- carbon with four substituents is tetrahedral
- carbon with four different substituents is chiral
- constitutional isomers – same formula, different connections
- enantiomers – same connection, different arrangement in 3D

Teaser: how do you give a computer a molecule including its handedness?

As a string.