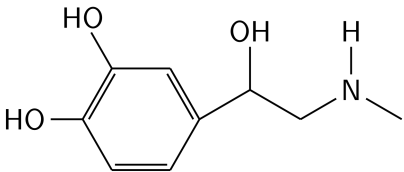
Organic chemistry for bioinformatics
Part 2: Naming molecules

After today you can

- recognise the functional groups and their priority
- give a molecule its name, step by step, and draw a molecule from its name
- decide whether a stereocentre is R or S
- write and read a SMILES string and look a molecule up in PubChem

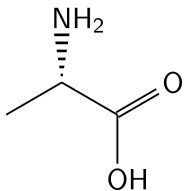


by end of lecture 2 you can name it
find its stereocentre
write its SMILES and look it up in Pubchem

Slides developed with the help of Claude (Anthropic), an AI assistant; content checked by S. J. Prohaska (2026).

One molecule, three descriptions

- lecture 1 ended with a teaser: "How do you give a computer a molecule including its handedness?"
- today: the name (for people) and the SMILES string (for computers)
- both must describe exactly one structure



structure (lecture 1)

(*S*)-2-aminopropanoic acid

name (part C, R/S in part D)

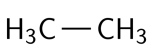
`C[C@@H](C(=O)O)N`

SMILES (part E)

trivial name: **L-alanine**

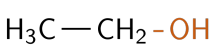
Functional groups: small part, big difference

- a functional group is a group of atoms with a typical behaviour
- it decides how a molecule reacts (lecture 3) and what it is called (today)
- C=C, C≡C and aromatic rings count as functional groups too, but in names they show up as -ene, -yne and benzene
- groups with heteroatoms (O, N, S) are ranked by priority



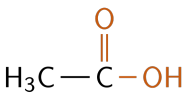
ethane

a gas, burns



ethanol

in wine and beer



ethanoic acid

in vinegar

The functional groups we use, by priority

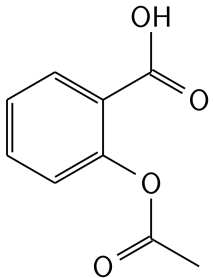
	class	group	suffix	prefix
1	carboxylic acid	$R-\overset{\overset{O}{\|\|}}{C}-OH$	-oic acid	carboxy-
2	ester	$R-\overset{\overset{O}{\|\|}}{C}-O-R'$	-oate	alkoxycarbonyl-
3	amide	$R-\overset{\overset{O}{\|\|}}{C}-NH_2$	-amide	carbamoyl-
4	nitrile	$R-C\equiv N$	-nitrile	cyano-
5	aldehyde	$R-\overset{\overset{O}{\|\|}}{C}-H$	-al	oxo-, formyl-
6	ketone	$R-\overset{\overset{O}{\|\|}}{C}-R'$	-one	oxo-

	class	group	suffix	prefix
7	alcohol	$R-OH$	-ol	hydroxy-
8	thiol	$R-SH$	-thiol	sulfanyl-
9	amine	$R-NH_2$	-amine	amino-
10	ether	$R-O-R'$	–	alkoxy- (methoxy-)
11	thioether	$R-S-R'$	–	alkylsulfanyl-

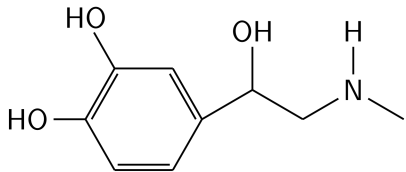
- highest priority in the molecule = principal group → suffix; all others → prefixes
- rule of thumb: acid > its derivatives > C=O > OH, SH > NH_2 > ethers

Action – Spot the groups

Name the class of every functional group you can find! (use the handout).



aspirin

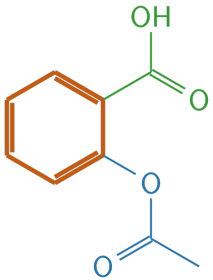


adrenaline

Hint: use the handout or slide 4

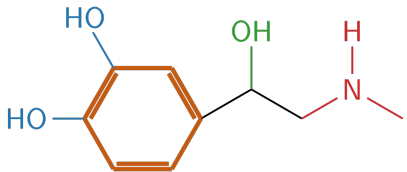
Action – Spot the groups

Name the class of every functional group you can find! (use the handout).



aspirin

carboxylic acid
ester
aromatic ring



adrenaline

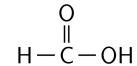
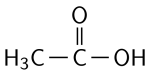
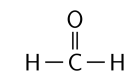
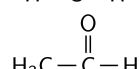
alcohol (2 × OH on the ring (→ phenol))
amine
aromatic ring

Why systematic names?

- a name must describe exactly one structure
- **test**: name $\to$ structure $\to$ name$'$ $\to$ structure$'$ $\to \infty$

  IUPAC: name $=$ name$'$ and

  structure $=$ structure$'$
- IUPAC (International Union of Pure and Applied Chemistry) gives rules to build such names
- many molecules also have a trivial name, which is often the one you find in biology papers

	trivial name	systematic name
	formic acid	methanoic acid
	acetic acid	ethanoic acid
	formaldehyde	methanal
	acetaldehyde	ethanal

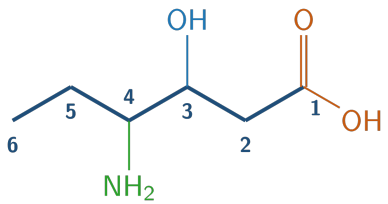
Our rules: IUPAC, slimmed down

- we follow the IUPAC rules (2013), but only a small part of them
- our test is weaker: name $\rightarrow$ structure $\rightarrow$ name$'$ $\rightarrow$ structure$'$
  only structure $=$ structure$'$ must be fulfilled
  but two different names may both be correct
- no stereo information in the name: R/S is a separate label (part D)

Support material

- handout *Nomenclature_practice_and_learn*: rules with worked examples
- handout *functional_groups*: the table on slide 4
- optional: book by Vollhardt, *Organische Chemie*

Anatomy of a name



4-amino-3-hydroxyhexanoic acid

prefixes with locants	**parent chain**	**suffix**
4-amino and 3-hydroxy	number of C (hex = 6)	the principal group
in alphabetical order	+ bonds (-an- = only single bonds)	(highest priority)

The naming algorithm

1. **Principal group:** find the group with the highest priority → suffix
2. **Parent chain:** the longest chain of carbons also containing the principal group in case of a tie → the chain with more multiple bonds
3. **Numbering:** lowest locant for the principal group, then for the multiple bonds, then for the substituents
4. **Substituents:** name them (prefixes) and sort them alphabetically; di-, tri-, tetra- do not count for the order
5. **Assemble:** prefixes with locants – parent (-ane, -ene, -yne) – suffix

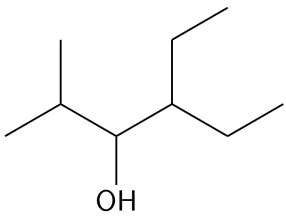
A fixed sequence of decisions: an algorithm, with tie-breaking rules.

Building blocks: chain length, bonds, substituents

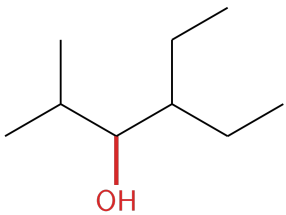
C	stem	C	stem
1	meth-	6	hex-
2	eth-	7	hept-
3	prop-	8	oct-
4	but-	9	non-
5	pent-	10	dec-

- only single bonds: -ane (e.g. propane)
- a C=C: -ene, a C≡C: -yne, with locant (e.g. but-2-ene, propyne)
- as a substituent: -yl (e.g. methyl, ethyl, propyl)
- several of the same: di-, tri-, tetra- (e.g. 2,3-diethyl. . . , 1,1,2-trimethyl. . .)
- rings: cyclo- (e.g. cyclohexane); benzene as a substituent: e.g. phenyl. . .
- English keeps the "e" before a consonant: butane-1,4-diamine, but butan-2-ol

Worked example 1

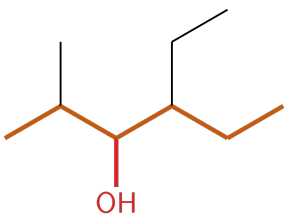


Worked example 1



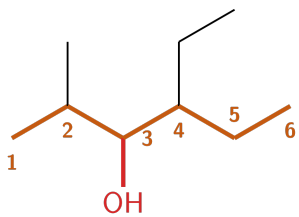
1. principal group: OH → **-ol**

Worked example 1



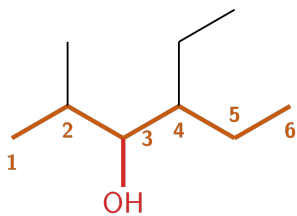
1. principal group: OH → **-ol**
2. longest chain through the C–OH:
   6 C → **hexan**
   (two choices of equal length give the same name)

Worked example 1



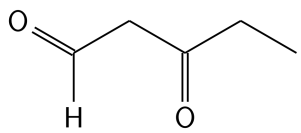
1. principal group: OH → **-ol**
2. longest chain through the C–OH:
   6 C → **hexan**
   (two choices of equal length give the same name)
3. numbering: OH gets 3 (from the other end: 4)
4. substituents: 4-ethyl, 2-methyl (alphabetically '**e**' before '**m**')

Worked example 1

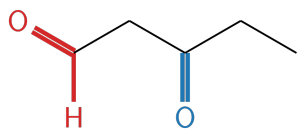


1. principal group: OH → **-ol**
2. longest chain through the C–OH:
   6 C → **hexan**
   (two choices of equal length give the same name)
3. numbering: OH gets 3 (from the other end: 4)
4. substituents: 4-ethyl, 2-methyl (alphabetically '**e**' before '**m**')
5. **4-ethyl-2-methylhexan-3-ol**

Worked example 2: two groups, one suffix

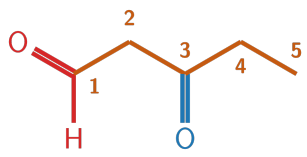


Worked example 2: two groups, one suffix



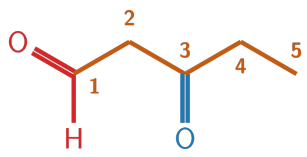
1. two groups: aldehyde (5) beats ketone (6)
   $\rightarrow$ suffix **-al**, the ketone becomes the prefix **oxo-**

Worked example 2: two groups, one suffix



1. two groups: aldehyde (5) beats ketone (6)
   $\rightarrow$ suffix **-al**, the ketone becomes the prefix **oxo-**
2. chain: 5 C, the CHO carbon is C1
   $\rightarrow$ **pentan**
3. ketone at C3 $\rightarrow$ **3-oxo**

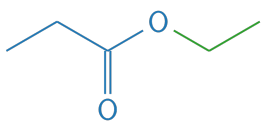
Worked example 2: two groups, one suffix



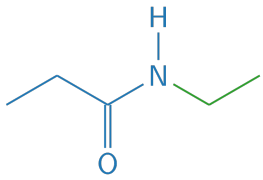
1. two groups: aldehyde (5) beats ketone (6)
   $\rightarrow$ suffix **-al**, the ketone becomes the prefix **oxo-**
2. chain: 5 C, the CHO carbon is C1
   $\rightarrow$ **pentan**
3. ketone at C3 $\rightarrow$ **3-oxo**
4. **3-oxopentanal**

Esters, amides and ethers: two parts, one name

- **ester**: alcohol part first, as a separate word; acid part with -oate
- **amide**: groups on the N get the locant N-
- **ether**: no suffix; the smaller part becomes a prefix (alkoxy-)



ethyl propanoate



N-ethylpropanamide

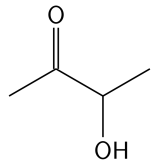


methoxyethane

Action – Silent Molecule Mail

1. Rows of 4. Student 1 gets a card and writes only the name on a slip.
2. Student 2 sees only the name and draws the structure.
3. Student 3 names the drawing, student 4 draws it again.
4. Compare the last drawing with the card: same structure = the row wins, even if the two names differ.
5. Second round: new card, start at the other end of the row.

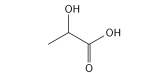
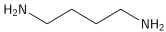
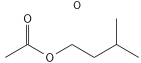
No talking, no phones – the handouts are allowed.



one of the cards (medium level)

15 min: 2 rounds · reveal

Reveal: who were the molecules?

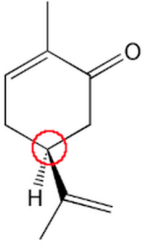
card	systematic name	trivial name
	2-hydroxypropanoic acid	**lactic acid**: sore muscles, yoghurt
	butane-1,4-diamine	**putrescine**: the smell of decay
	3-hydroxybutan-2-one	**acetoin**: butter aroma
	2-aminopropanoic acid	**alanine**: proteinogenic amino acid
	3-methylbutyl ethanoate	**isoamyl acetate**: banana flavour

Why R and S?

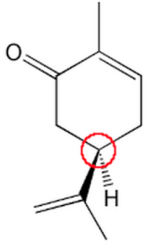
- enantiomers have the same name so far: the name says nothing about 3D
- but they can smell or act differently (carvone, lecture 1)
- the label (R) or (S) says which mirror image you mean
- rules by Cahn, Ingold and Prelog (CIP), 1956/66



caraway



(S)-carvone



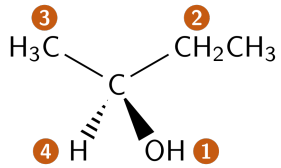
(R)-carvone



spearmint

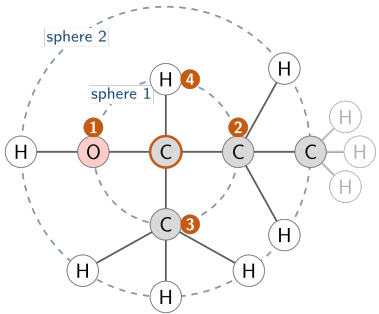
Step 1: rank the four groups (CIP priorities)

1. look at the atoms bonded directly to the stereocentre
   higher atomic number = higher priority
   $\rightarrow$ O > N > C > H
2. in case of a tie:
   go one sphere further out and compare the sets of atoms there
3. double bond: count the partner atom twice



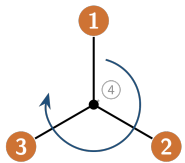
2-butanol:

- O of OH: O (8) $\rightarrow$ 1; H (1) $\rightarrow$ 4
- tie? C of ethyl vs. C of methyl: both C (6)
- next sphere: ethyl {C, H, H} > methyl {H, H, H}
- ethyl $\rightarrow$ 2; methyl $\rightarrow$ 3

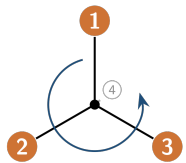


Step 2: look along the steering column

- turn the molecule so that group **4** points away from you (it is the steering column)
- the other three form the steering wheel
- follow $1 \to 2 \to 3$:
  clockwise = R, anticlockwise = S



R: clockwise

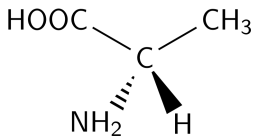


S: anticlockwise

Action – R or S? Ask the die

Your tetrahedral die, numbered 1–4,
is a model for the **(R)** configuration.
With 4 pointing away, 1 → 2 → 3 runs clockwise.

1. rank the groups of L-alanine (step 1)
2. turn the die: 4 where the H is, 1 where the NH_2 is
3. 2 and 3 match the drawing → (R);
   2 and 3 are swapped → (S)
   i.e. (mirror image of the die)

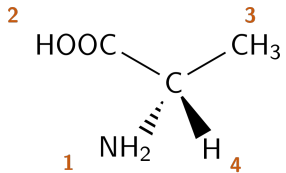


L-alanine

Action – R or S? Ask the die

Your tetrahedral die, numbered 1–4,
is a model for the **(R)** configuration.
With 4 pointing away, 1 → 2 → 3 runs clockwise.

1. rank the groups of L-alanine (step 1)
2. turn the die: 4 where the H is, 1 where the NH_2 is
3. 2 and 3 match the drawing → (R);
   2 and 3 are swapped → (S)
   i.e. (mirror image of the die)



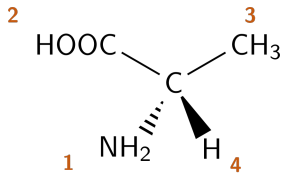
L-alanine

$NH_2 > COOH > CH_3 > H$

Action – R or S? Ask the die

Your tetrahedral die, numbered 1–4,
is a model for the **(R)** configuration.
With 4 pointing away, 1 → 2 → 3 runs clockwise.

1. rank the groups of L-alanine (step 1)
2. turn the die: 4 where the H is, 1 where the NH_2 is
3. 2 and 3 match the drawing → (R);
   2 and 3 are swapped → (S)
   i.e. (mirror image of the die)



L-alanine

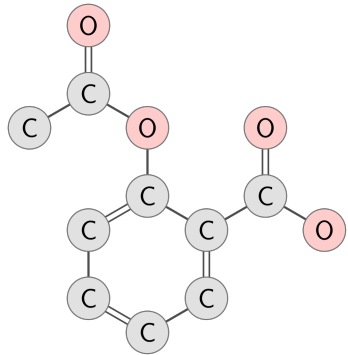
$NH_2 > COOH > CH_3 > H$

(S): 2 and 3 are swapped

SMILES: a molecule as a string

Simplified **M**olecular **I**nput **L**ine **E**ntry **S**ystem

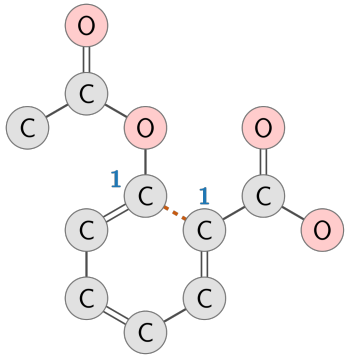
1. leave out the H atoms; atoms are nodes, bonds are edges



SMILES: a molecule as a string

Simplified **M**olecular **I**nput **L**ine **E**ntry **S**ystem

1. leave out the H atoms; atoms are nodes, bonds are edges
2. break each ring once to get a spanning tree; label both ends with the same digit (1)



Side note: number of cuts = bonds − atoms + 1 (cyclomatic number); aspirin: $13 - 13 + 1 = 1$

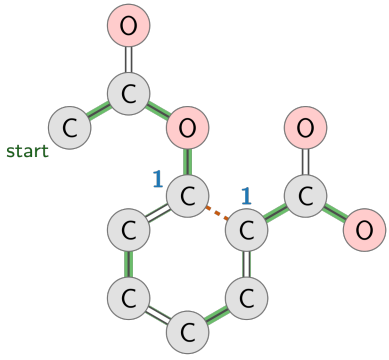
SMILES: a molecule as a string

Simplified **M**olecular **I**nput **L**ine **E**ntry **S**ystem

1. leave out the H atoms; atoms are nodes, bonds are edges
2. break each ring once to get a spanning tree; label both ends with the same digit (1)
3. walk along the longest path; write atoms and bonds (single: nothing, double =, triple #); side branches in ( )

`CC(=O)OC1=CC=CC=C1C(=O)O`

aspirin



Side note: number of cuts = bonds − atoms + 1 (cyclomatic number); aspirin: 13 − 13 + 1 = 1

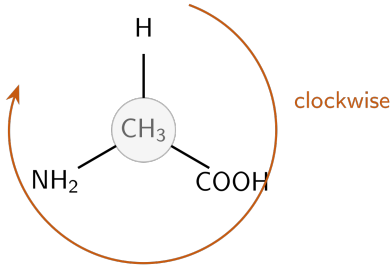
SMILES in one table

what	how	example
atoms	element symbols, H usually left out	`CCO` ethanol
double, triple bond	`=`, `#`	`C=C` ethene, `CC#C` propyne
branch	in brackets	`CC(=O)O` acetic acid
ring	a digit at both ends of the broken bond	`C1CCCCC1` cyclohexane
aromatic ring	alternating bonds, or lower-case atoms	`C1=CC=CC=C1` = `c1ccccc1`
stereocentre	`@` or `@@`	next slide

- one molecule has many valid SMILES (`CCO`, `OCC`, `C(C)O` ...)
- databases use one canonical SMILES: a graph canonicalisation problem

Handedness in SMILES: @ and @@

- look from the atom written before the stereocentre towards the stereocentre
- look at neighbours of the stereocentre
- use their order in the SMILES string to determine if they run `@@` clockwise or `@` anticlockwise
- note: @/@@ depends on the writing order, R/S does not
- L-alanine: `C[C@@H](C(=O)O)N`
  from CH_3: H → COOH → NH_2 clockwise



L-alanine, seen from the CH_3 group

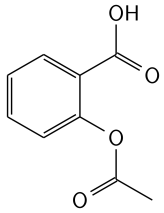
`C`	`[C@@H]`	`(C(=O)O)`	`N`
↑	↑	↑	↑
CH_3	H	COOH	NH_2
1st	2nd	3rd	4th

Action – Decode, then ask PubChem

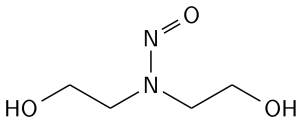
1. in pairs: draw the structure (4 min)
   a) `CC(=O)OC1=CC=CC=C1C(=O)O`
   b) `OCCN(N=O)CCO`
2. paste the strings into **pubchem.ncbi.nlm.nih.gov**: compare structure, name and the safety section (5 min)
3. write two SMILES for ethanol: why does PubChem find one compound? (2 min)
4. draw one Molecule Mail card in the PubChem Sketcher and let it name the molecule (4 min)

Action – Decode, then ask PubChem

1. in pairs: draw the structure (4 min)
   a) `CC(=O)OC1=CC=CC=C1C(=O)O`
   b) `OCCN(N=O)CCO`
2. paste the strings into **pubchem.ncbi.nlm.nih.gov**: compare structure, name and the safety section (5 min)
3. write two SMILES for ethanol: why does PubChem find one compound? (2 min)
4. draw one Molecule Mail card in the PubChem Sketcher and let it name the molecule (4 min)



a) aspirin (2-acetyloxybenzoic acid)



b) *N*-nitrosodiethanolamine: carcinogenic, a contaminant in cosmetics

PubChem and the PubChem Sketcher

PubChem (NIH, free)

- search by name, trivial name, SMILES, formula
- structure 2D and 3D, IUPAC name, SMILES, properties, safety
- 3D view: rotate the conformer with the mouse

`pubchem.ncbi.nlm.nih.gov`

PubChem Sketcher

- draw a molecule; the SMILES is written while you draw
- paste a SMILES to get the drawing
- search PubChem directly from the drawing

`pubchem.ncbi.nlm.nih.gov/edit3/index.html`

Summary and outlook

- functional groups decide reactivity and name; priority picks the suffix
- naming is an algorithm: principal group, chain, numbering, prefixes, assemble
- R/S: rank the groups, look along the lowest one, clockwise = R
- SMILES: depth-first walk over a spanning tree; @/@@ for handedness

Next time: what functional groups do – reactions, and the enzymes that catalyse them.

These slides were made with an AI assistant (Claude). Every name, structure and R/S label was checked by hand: do the same with AI answers.