

C2184 Úvod do programování v Pythonu

Lekce 4 - samostatné úkoly

V hodině je nutné zvládnout 2 samostatné úkoly a zbyvajících je možné dodělat jako domácí úkol. Ke každému úloze je k dispozici šablona ve studijních materiálech.

A) Caesarova šifra

- napište funkci `decode()`, která podle zadaného klíče (celého čísla) provede posun znaků abecedy

BONUS:

- modifikujte skript, aby byl schopný dekodovat i řetězec `encoded_text_2`

B) Vlastní komparer pro funkci `sort()`

- v dokumentaci k Pythonu pohlédněte očekávané návratové funkce vlastního kompareru
- dopište kód kompareru, aby porovnával absolutní hodnoty čísel

BONUS:

- otestujte i pro absolutní čísla a v případě nutnosti modifikujte, aby se porovnávala pouze reálná část čísla

C) Zpracování CSV (Comma-separated values)

- zjistěte jaký je použit oddělovač sloupečků v přiložené proměnné
- spočítejte průměr, medián a rozptyl pro sloupečky `pKa` a `qH`

BONUS:

- spočítejte korelaci mezi sloupečky `pKa` a `qH`

D) Molekula ve formátu MOL

- napište funkci, která načte informace o molekule - pro každý atom vytvoří tuple s informacemi např. `('O', 0.4700, 2.5688, 0.0006)` a ty uloží do listu `[('O', 0.4700, 2.5688, 0.0006), ('O', -3.1271, -0.4436, -0.0003), ...]` a vrátí jako návratovou hodnotu
- formát molekuly je MDL Molfile¹
- spočítejte vzdálenost mezi atomy 3 a 8 (číslováno od 1), vazebný úhel mezi atomy 8, 7 a 9 a torzní úhel mezi 5, 9, 7 a 8

BONUS:

- spočítejte celkovou vazebnou délku (součet všech vazeb zapsaných v souboru)

Poznámka: Zkontrolované úkoly NE nahraďte do odevzdávací, do odevzdávací nahrajte pouze úkoly, které nestihnete během hodiny udělat.

¹podrobněji <http://c4.cabrillo.edu/404/ctfile.pdf>

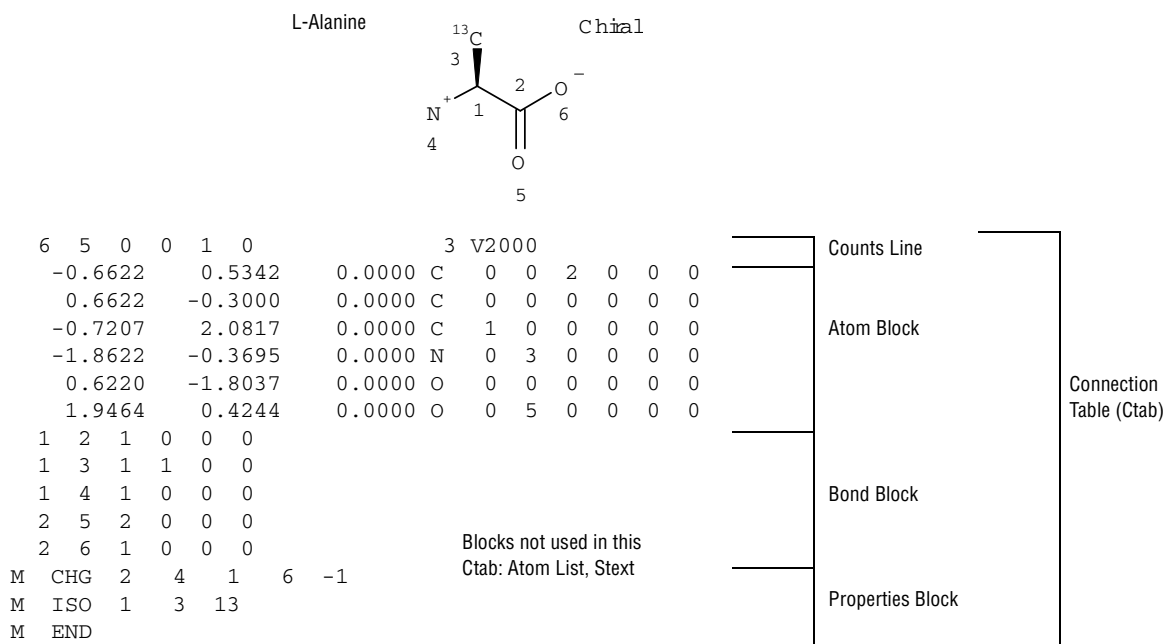
Chapter 2: The Connection Table [CTAB] (V2000)

Overview

A connection table (Ctab) contains information describing the structural relationships and properties of a collection of atoms. The atoms may be wholly or partially connected by bonds. Such collections may, for example, describe molecules, molecular fragments, substructures, substituent groups, polymers, alloys, formulations, mixtures, and unconnected atoms. The connection table is fundamental to all of Elsevier MDL's file formats.

The following figure shows the connection table of a simple molecule (alanine) with the various data blocks identified. The connection table corresponds to the following alanine molecule. The atom numbers on the structure correspond to atom numbers in the Ctab. An atom number is assigned according to the order of the atom in the Atom Block.

Figure 3 Connection table organization illustrated using alanine



The format for a Ctab block is:

Counts line:	Important specifications here relate to the number of atoms, bonds, and atom lists, the chiral flag setting, and the Ctab version.
Atom block:	Specifies the atomic symbol and any mass difference, charge, stereochemistry, and associated hydrogens for each atom.
Bond block:	Specifies the two atoms connected by the bond, the bond type, and any bond stereochemistry and topology (chain or ring properties) for each bond.
Atom list block:	Identifies the atom (number) of the list and the atoms in the list.
Stext (structural text descriptor) block:	Used by MDL ISIS/Desktop programs.
Properties block:	Provides for future expandability of Ctab features, while maintaining compatibility with earlier Ctab configurations.

The detailed format for each block outlined above follows:

Note: A blank *numerical* entry on any line should be read as “0” (zero). Spaces are significant and correspond to one or more of the following:

- Absence of an entry
- Empty character positions within an entry
- Spaces between entries; single unless specifically noted otherwise

The FORTRAN format for coordinate information in the V2000 CTfile format is typically F10.4.

The Counts Line

```
aaabbblllffcccsssxxrrrpppiiimmvvvvv
```

Where:

aaa	= number of atoms (current max 255)*	[Generic]
bbb	= number of bonds (current max 255)*	[Generic]
lll	= number of atom lists (max 30)*	[Query]
fff	= (obsolete)	
ccc	= chiral flag: 0=not chiral, 1=chiral	[Generic]
sss	= number of stext entries	[MDL ISIS/Desktop]
xxx	= (obsolete)	
rrr	= (obsolete)	
ppp	= (obsolete)	
iii	= (obsolete)	
mmm	= number of lines of additional properties, including the M END line. No longer supported, the default is set to 999.	[Generic]

* These limits apply to MACCS-II, REACCS, and the MDL ISIS/Host Reaction Gateway, but not to the MDL ISIS/Host Molecule Gateway or MDL ISIS/Desktop.

For example, the counts line in the Ctab shown in the previous figure shows six atoms, five bonds, the CHIRAL flag *on*, and three lines in the properties block:

```
6 5 0 0 1 0          3 V2000
```

The Atom Block

The Atom Block is made up of atom lines, one line per atom with the following format:

```
xxxxx.xxxxxyyyy.yyyyzzzz.zzzz aaaddcccssshhbbbvvhHrrriimmnnnee
```

where the values are described in the following table:

Figure 4 Meaning of values in the atom block

Field	Meaning	Values	Notes
x y z	atom coordinates		[Generic]
aaa	atom symbol	entry in periodic table or L for atom list, A, Q, * for unspecified atom, and LP for lone pair, or R# for Rgroup label	[Generic, Query, 3D, Rgroup]
dd	mass difference	-3, -2, -1, 0, 1, 2, 3, 4 (0 if value beyond these limits)	[Generic] Difference from mass in periodic table. Wider range of values allowed by M ISO line, below. Retained for compatibility with older Ctabs, M ISO takes precedence.
ccc	charge	0 = uncharged or value other than these, 1 = +3, 2 = +2, 3 = +1, 4 = doublet radical, 5 = -1, 6 = -2, 7 = -3	[Generic] Wider range of values in M CHG and M RAD lines below. Retained for compatibility with older Ctabs, M CHG and M RAD lines take precedence.
sss	atom stereo parity	0 = not stereo, 1 = odd, 2 = even, 3 = either or unmarked stereo center	[Generic] Ignored when read.
hhh	hydrogen count + 1	1 = H0, 2 = H1, 3 = H2, 4 = H3, 5 = H4	[Query] H0 means no H atoms allowed unless explicitly drawn. Hn means atom must have n or more H's in excess of explicit H's.
bbb	stereo care box	0 = ignore stereo configuration of this double bond atom, 1 = stereo configuration of double bond atom must match	[Query] Double bond stereochemistry is considered during SSS only if both ends of the bond are marked with stereo care boxes.
vvv	valence	0 = no marking (default) (1 to 14) = (1 to 14) 15 = zero valence	[Generic] Shows number of bonds to this atom, including bonds to implied H's.
HHH	H0 designator	0 = not specified, 1 = no H atoms allowed	[MDL ISIS/Desktop] Redundant with hydrogen count information. May be unsupported in future releases of Elsevier MDL software.
rrr	Not used		
iii	Not used		
mmm	atom-atom mapping number	1 - number of atoms	[Reaction]
nnn	inversion/retention flag	0 = property not applied 1 = configuration is inverted, 2 = configuration is retained,	[Reaction]
eee	exact change flag	0 = property not applied, 1 = change on atom must be exactly as shown	[Reaction, Query]

With CTab version V 2000, the `dd` and `ccc` fields have been superseded by the `M ISO`, `M CHG`, and `M RAD` lines in the properties block, described below. For compatibility, all releases since `M ACCS-II 2.0`, `REACCS 8.1`, and `MDL ISIS 1.0`:

- Write appropriate values in both places if the values are in the old range.
- Use the atom block fields if there are no `M ISO`, `M CHG`, or `M RAD` lines in the properties block.

Support for these atom block fields may be removed in future releases of Elsevier MDL software.

The Bond Block

The Bond Block is made up of bond lines, one line per bond, with the following format:

```
111222tttsssxxrrrccc
```

where the values are described in the following table:

Figure 5 Meaning of values in the bond block

Field	Meaning	Values	Notes
111	first atom number	1 - number of atoms	[Generic]
222	second atom number	1 - number of atoms	[Generic]
ttt	bond type	1 = Single, 2 = Double, 3 = Triple, 4 = Aromatic, 5 = Single or Double, 6 = Single or Aromatic, 7 = Double or Aromatic, 8 = Any	[Query] Values 4 through 8 are for SSS queries only.
sss	bond stereo	Single bonds: 0 = not stereo, 1 = Up, 4 = Either, 6 = Down, Double bonds: 0 = Use x-, y-, z-coords from atom block to determine cis or trans, 3 = Cis or trans (either) double bond	[Generic] The wedge (pointed) end of the stereo bond is at the first atom (Field 111 above)
xxx	not used		
rrr	bond topology	0 = Either, 1 = Ring, 2 = Chain	[Query] SSS queries only.
ccc	reacting center status	0 = unmarked, 1 = a center, -1 = not a center, Additional: 2 = no change, 4 = bond made/broken, 8 = bond order changes 12 = 4+8 (both made/broken and changes); 5 = (4 + 1), 9 = (8 + 1), and 13 = (12 + 1) are also possible	[Reaction, Query]

The Atom List Block [Query]

Note: Newer programs use the `M ALS` item in the properties block in place of the atom list block. The atom list block is retained for compatibility, but information in an `M ALS` item supersedes atom list block information.

Made up of atom list lines, one line per list, with the following format:

```
aaa kSSSSn 111 222 333 444 555
```