



MEDNARODNA
PODIPLOMSKA ŠOLA
JOŽEFA STEFANA

JOŽEF STEFAN
INTERNATIONAL
POSTGRADUATE SCHOOL

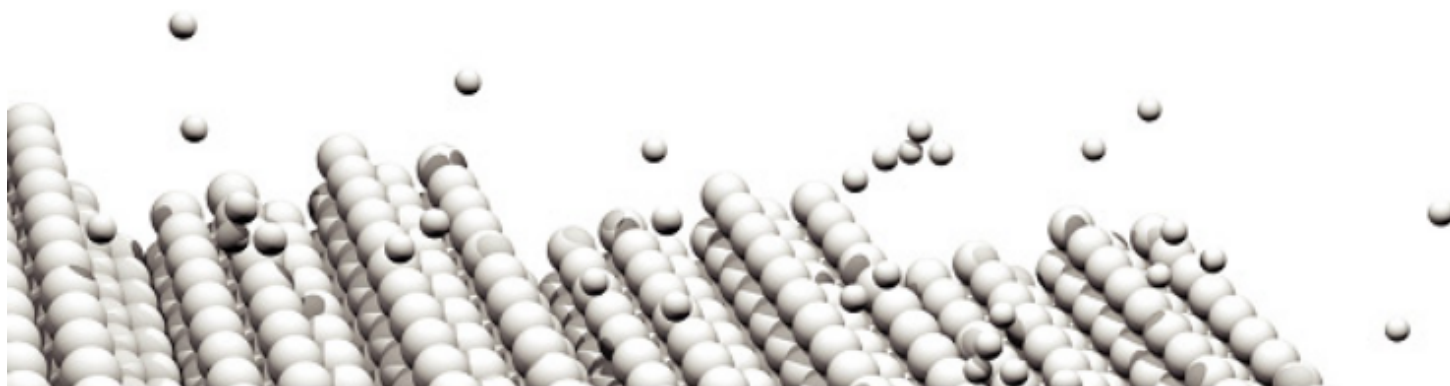


Delavnica: Molekulsko modeliranje in zapis SMILES za molekule

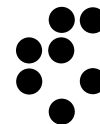
Anton Kokalj^{1,2} in Matic Poberžnik¹

¹Odsek za fizikalno in organsko kemijo, Institut "Jožef Stefan"

²Mednarodna podiplomska šola Jožefa Stefana



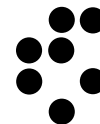
www.mps.si



Seznam vaj:

1. SMILES: format za zapis molekulskih struktur (**vaja-1**)
(SMILES = **S**implified **M**olecular-**I**nterface **L**ine-**E**nter **S**ystem)
2. Stabilnost *cis* in *trans* izomerov 1,2-difluoroetena (**vaja-2**)
3. Stabilnost *cis* in *trans* izomerov 2-butena (**vaja-3**)
4. Molekulske orbitale benzena (**vaja-4**)
5. Modeliranje po lastni izbiri (**vaja-5**)
6. Domoča naloga: singletni in tripletni kisik

Molekulsko modeliranje in zapis SMILES za molekule



Za potrebe delavnice smo pripravili mapo **MPS-2022**, ki se nahaja na namizju virtualnega operacijskega sistema. V tej mapi boste našli podmape **vaja-1** do **vaja-5**.

Za začetek odprite mapo **MPS-2022** (dvojni klik) in s pritiskom na desni gumb izberite opcijo **Open in Terminal**, kar bo odprlo ukazno lupino (terminal) na pravilni lokaciji. Znotraj terminala vnesite naslednja ukaza:

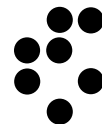
- **pwd**

```
/home/user/MPS-2022
```

- **ls**

```
ash_functions  delavnica.pdf  pseudo        vaja-1  vaja-4
calcmol        install.sh     README.md     vaja-2  vaja-5
calcmol.pwtk   lib           slides        vaja-3  xsf2_manipulator
```

Če ste dobili takšen izpis kot je na sliki, potem ste na pravem mestu in lahko začnemo z vajami. Če želite podrobneje spoznati ukazno lupino predlagamo ta **tutorial**.

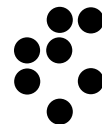


Specifikacija 3D molekulske strukture

Pri molekulskem modeliranju na podlagi prvih principov moramo kot vhodni podatek podati geometrijo molekule, tj. za vsak atom posebej moramo navesti tip atoma in njegove kartezične (x, y, z) koordinate. Kot primer je prikazana geometrija metana (CH_4) v kartezičnem zapisu (v enotah Å):

C	1.03130	-0.01731	-0.09339
H	0.66723	0.60288	-0.91540
H	0.66723	-1.03929	-0.21948
H	0.66723	0.38449	0.85473
H	2.12350	-0.01731	-0.09339

Takšna kartezična specifikacija molekulske geometrije ni prijazna do uporabnika. Na srečo obstajajo načini, kako generirati kartezične koordinate na podlagi simbolnega kemijskega zapisa. Tu se bomo osredotočili na zapis **SMILES** (*Simplified Molecular-Input Line-Entry System*). Gre za simbolični enovrstični zapis molekul, ki ga lahko s pomočjo programa kot je **Open Babel** pretvorimo v 3D kartezično obliko.

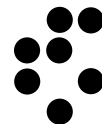


1. SMILES: format za zapis molekulskih struktur

Pravila za zapis **SMILES**:

1. atome zapišemo v oglatih oklepajih, npr. **[Al]**, **[C]**, **[S]**, ...
2. oglate oklepaje lahko izpustimo pri atomih tipa B, C, N, O, P, S, F, Cl, Br in I; ta zapis pomeni “nasičene” atome, nanje so vezani vodikovi atomi v skladu z valenco. Primer: **C** → CH₄, **N** → NH₃, **S** → H₂S
3. atomi B, C, N, O, P, S napisani z malimi črkami (**b**, **c**, **n**, **o**, **p**, **s**) pomenijo nenasičene atome. Primer: **c** → CH₃, **n** → NH₂, **s** → HS
4. kemijske vezi: **-** enojna vez, **=** dvojna vez, **#** trojna vez. Primer:
 - **NN** ali **N-N** → N₂H₄ oz. H₂N-NH₂
 - **nn** ali **N=N** → N₂H₂ oz. HN=NH
 - **N#N** → N₂ oz. N≡N
5. za *cis* in *trans* izomere uporabljamo **** in **/** poševnice. Primer:
 - **C/C=C/C** → trans-2-buten (poševnica **/** pomeni “spodaj”; tj. *zgoraj/spodaj/še-bolj-spodaj*)
 - **C/C=C\C** → cis-2-buten (poševnica **** pomeni “zgoraj”; tj. *zgoraj/spodaj\zgoraj*)

Več informacij o zapisu **SMILES** lahko najdete na [Wikipediji](#) in na tej [povezavi](#).



1. SMILES: format za zapis molekulskih struktur

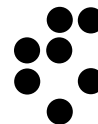
Najprej bomo zapisali preprosto molekulo etanola ($\text{C}_2\text{H}_5\text{OH}$). Prestavite se v mapo **vaja-1** in tam poženite program **obgui**, torej:

- **cd vaja-1**
- **obgui**

Odprlo se bo okno (glej sliko na naslednji strani), kjer bodite pozorni predvsem na to, da je pod **INPUT FORMAT** izbran “**smi -- SMILES format**” in obkljukana opcija **Input below**.

Pod **OUTPUT FORMAT** izberite “**xyz -- XYZ cartesian coordinates format**” in obkljukajte opciji “**Generate 3D coordinates**” in “**Display in firefox**”.

Etanol v formatu SMILES zapišemo kot **CCO**. Na strani **INPUT FORMAT** torej vpišite **CCO** in pritisnite gumb **CONVERT**. Odprlo se bo okno v brskalniku **firefox** s shematsko skeletno strukturo molekule etanola.



1. SMILES: format za zapis molekulskih struktur

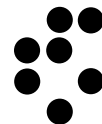
The screenshot shows the OpenBabelGUI application window. The 'INPUT FORMAT' dropdown is set to 'smi -- SMILES format'. The 'OUTPUT FORMAT' dropdown is set to 'xyz -- XYZ cartesian coordinates format'. The 'Generate 3D coordinates' checkbox is checked. The 'Output file' section has 'Output below only (no output file)' and 'Display in firefox' checked. The input text is 'CCO'. The output shows the 3D coordinates for 9 atoms (9 molecules converted).

Atom	X	Y	Z
C	1.08021	-0.05156	0.04668
C	2.59478	-0.05870	0.04513
O	3.07498	-0.90532	-0.98908
H	0.69421	0.60900	0.82804
H	0.68853	-1.06046	0.21303
H	0.69555	0.28352	-0.92218
H	2.98663	-0.41635	1.00213
H	2.98219	0.94892	-0.13139
H	2.73382	-1.79969	-0.81791

On the right, a Firefox browser window displays the 3D structure of ethanol (skeletal structure) with the label 'skeletna struktura etanola'.

Annotations in red text with arrows point to the 'smi -- SMILES format' dropdown (labeled 'zapis SMILES'), the 'Generate 3D coordinates' checkbox (labeled 'generirane 3D kartezične koordinate'), and the 3D structure of ethanol (labeled 'skeletna struktura etanola').

Če ste vse naredili pravilno, ste dobili rezultat kot je prikazan na zgornji sliki.



1. SMILES: format za zapis molekulskih struktur

Za olajšanje dela s **SMILES** zapisom smo pripravili bližnjice, ki takšen zapis v ukazni vrstici avtomatsko pretvorijo v zapis s kartezičnimi koordinatami, ter odprejo s primernim programom. Ukazi, ki jih boste uporabljali so `smi`, `xsmi` in `asmi`.

Ukaz `smi` izpiše generirane kartezične koordinate v okno terminala, torej če vpišemo:

- `smi "CCO"`

bomo kot rezultat dobili x, y, z koordinate za vsak atom v molekuli etanola.

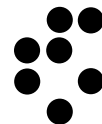
POZOR: pri uporabi orodij v terminalu moramo **SMILES** oznako zapisati z narekovaji. Torej: čeprav je SMILES zapis za etanol `CCO` v terminalu napišemo `"CCO"` !

Ukaza `xsmi` in `asmi` delujeta podobno, le da avtomatsko odpreta ustrezen program za ogled 3D strukture molekule; `xsmi` uporablja program `xcrysden`, medtem ko `asmi` uporablja `avogadro`.

- `xsmi "CCO"`
- `asmi "CCO"`

Za izris skeletne strukture molekule smo pripravili ukaza `smipng` in `smisvg`, npr.

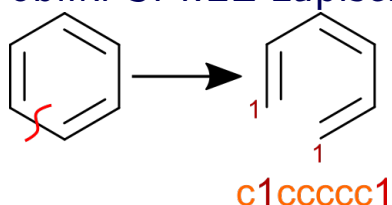
- `smipng "CCO"`



1. SMILES: format za zapis molekulskih struktur

Poizkusili bomo generirati kompleksnejšo strukturo z zapisom SMILES, tj. acetilsalicilno kislino (bolj znano kot aspirin). Postopek je sledeč:

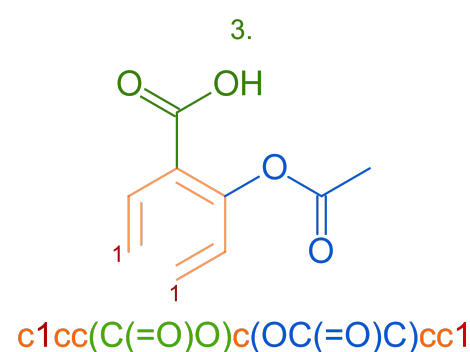
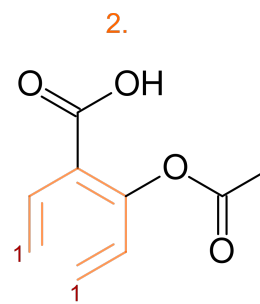
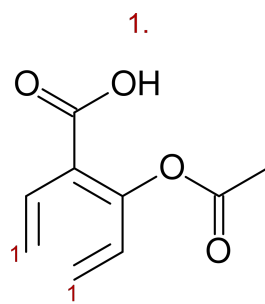
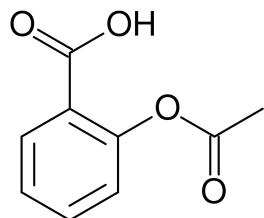
1. **Recept za ciklične spojine:** označite vsaj en par atomov v vsakem obroču s številom in prekinite obroč na tem mestu, nato sledite obroču od prvega do zadnjega označenega atoma. Npr. benzen v obliki SMILE zapišemo kot c1ccccc1.

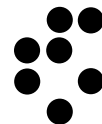


2. Izberite glavno verigo v strukturi
3. Stranske verige ločite od glavne z oklepaji na ustreznih mestih. Npr. etanojsko kislino bi zapisali kot CC(=O)O, kjer karbonylni kisik predstavlja stransko verigo.

Postopek je za acetilsalicilno kislino prikazan na spodnji sliki:

2D skeletna struktura



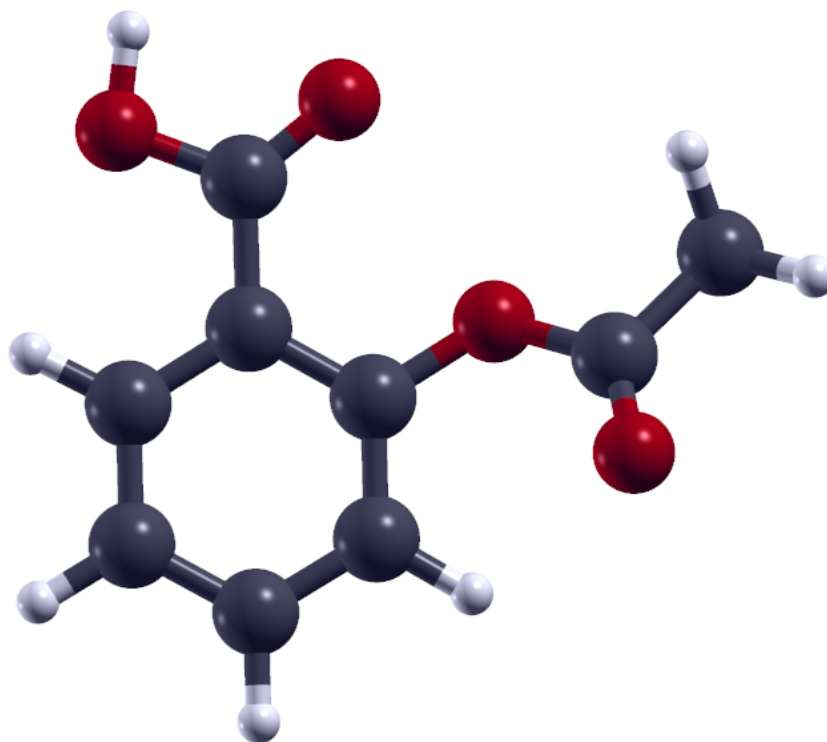


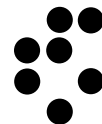
1. SMILES: format za zapis molekulskih struktur

3D strukturo acetilsalicilne kisline generirajte z ukazom:

- `xsmi "c1cc(C(=O)O)c(OC(=O)C)cc1"`

Bodite pozorni na to, da so nenasičeni ogljikovi atomi označeni z malo črko **c**, medtem ko so nasičeni ogljikovi atomi (oz. ogljikovi atomi, ki imajo eksplicitno navedene vse vezi) označeni z veliko črko **C**.





2. Stabilnost *cis* in *trans* izomerov 1,2-difluoroetena

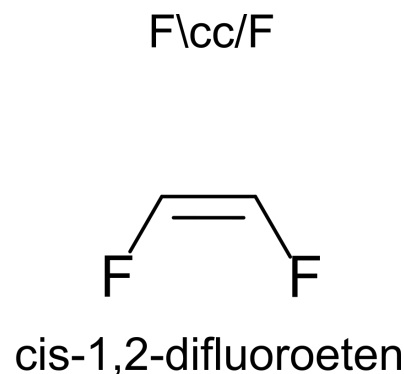
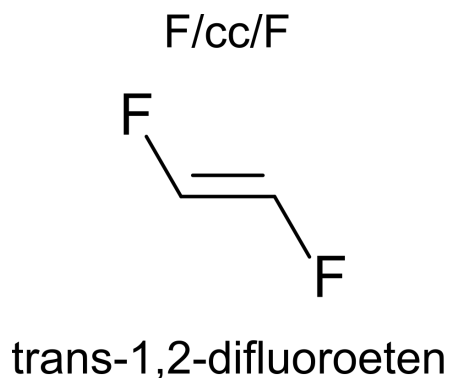
V drugi vaji bomo preverili kateri geometrični izomer 1,2-difluoroetena je stabilnejši.

Najprej izvedemo ukaz v terminalu:

- `cd ../vaja-2`

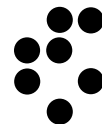
Sedaj se nahajamo v mapi `vaja-2`. Tako bodo vse datoteke, ki jih bomo generirali v tej mapi.

Skeletni strukturi izomerov in pripadajoča zapisa SMILE so izrisani na sliki.



Preverimo, če sta zapisa SMILE pravilna:

- `asmi "F/cc/F"`
- `asmi "F/cc\F"`



2. Stabilnost *cis* in *trans* izomerov 1,2-difluoroetena

Za izračune bomo uporabljali teorijo gostotnega funkcionala (angl. *Density Functional Theory (DFT)*), kot je implementirana v prosto dostopnem programskem paketu **Quantum ESPRESSO**. Za to delavnico smo pripravili ukaz **calcmol**, ki kot vhodni podatek zahteva le **SMILE** specifikacijo molekule in ime molekule. Za izpis opcij, ki so na voljo, pošnite ukaz brez dodatnih argumentov:

- **calcmol**

Ta ukaz izpiše:

```
Usage: calcmol [options] molecule SMILE
```

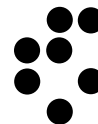
where:

```
molecule :    name of molecule
SMILE      :    SMILE specification of molecular structure
              (SMILE stands for "Simplified Molecular-Input Line-Entry")
```

Options are:

...

To pomeni, da sta dva argumenta obvezna in sicer **molecule** in **SMILE**, medtem ko so ostali parametri (**[options]**) opcijski in kadar niso podani, zavzamejo privzete vrednosti. Pri tej vaji bomo uporabljali le privzete vrednosti.



2. Stabilnost *cis* in *trans* izomerov 1,2-difluoroetena

Najprej bomo izračunali optimizirano strukturo *trans*-1,2-difluoroetena. SMILE zapis te molekule je F/cc/F. Strukturo preverite z ukazom:

- `xsmi "F/cc/F"`

Ko se prepričate, da je struktura pravilna, pošnite račun z ukazom:

- `calcmol trans-1,2-difluoroeten "F/cc/F"`

Opazili boste, da se je v terminalu izpisala vrsta podatkov. Najprej se izpišejo vhodni podatki:

```
CALCMOL> =====  
CALCMOL> *** Building and calculating a "trans-1,2-difluoroeten" molecular structure  
CALCMOL> =====
```

```
Specified command-line : trans-1,2-difluoroeten F/cc/F
```

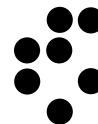
```
Structure name : trans-1,2-difluoroeten
```

```
SMILE : F/cc/F
```

```
Calculation type : RELAX
```

```
Summary of generated molecular structure from SMILE
```

```
Number of atoms : 6
```



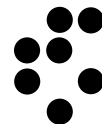
2. Stabilnost *cis* in *trans* izomerov 1,2-difluoroetena

Med pomembnejšimi podatki, ki so se izpisali v terminalu, so:

```
RELAX> =====  
RELAX> *** making a RELAX calculation of "trans-1,2-difluoroeten" via pw.x  
RELAX> =====  
  
Executable: /usr/bin/pw.x  
Running: /usr/bin/mpirun -np 2 pw.x -in \  
        pw.trans-1,2-difluoroeten.relax.in > pw.trans-1,2-difluoroeten.relax.out
```

Te vrstice pomenijo:

- za račun je bil uporabljen program **pw.x**
- ime vhodne datoteke je **pw.trans-1,2-difluoroeten.relax.in**: ta vsebuje avto-generirane podatke, ki so potrebni za izračun molekule
- ime izhodne datoteke je **pw.trans-1,2-difluoroeten.relax.out**: tu so zbrani podatki povezani z izračunom, kot npr. optimizirana geometrija in energija molekule. V kolikor gre z računom kaj narobe, moramo pogledati v izhodno datoteko, da odkrijemo težavo.



2. Stabilnost *cis* in *trans* izomerov 1,2-difluoroetena

Vrsta izračuna je bila **relax**; s tem programu **pw.x** naročimo, da optimizira geometrijo molekule (tj. poišče lokalni minimum energije) za dano vhodno geometrijo molekule.

Vsebino vhodne in izhodne datoteke lahko v terminalu pregledujemo z ukazom **less**:

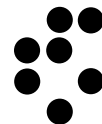
- **less pw.trans-1,2-difluoroeten.relax.out**

Naprej in nazaj se pomikamo s ↓ in ↑ puščicama. Za **izhod iz programa**, pritisnite tipko **q**.

Kako je potekala “relaksacija” lahko preverite s programom **xcrysden**, tj.:

- **xcrysden --pwo pw.trans-1,2-difluoroeten.relax.out**

Program je geometrijo optimiziral v nekaj korakih. Če vključite še opcijo za prikaz sil (**Display-->Forces**), lahko opazite kako se te manjšajo v vsakem naslednjem koraku “relaksacije”.



2. Stabilnost *cis* in *trans* izomerov 1,2-difluoroetena

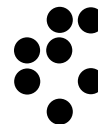
Energijo vsake geometrije med relaksacijo lahko izpišemo z ukazom `pwene`. Opazili boste, da se energija v vsakem koraku zmanjša, pri čemer je sprememba načeloma manjša iz koraka v korak. Manjša energija označuje stabilnejšo strukturo.

- `pwene pw.trans-1,2-difluoroeten.relax.out`

```
!   total energy           =   -122.54166088 ry
!   total energy           =   -122.54334123 ry
!   total energy           =   -122.54431915 ry
!   total energy           =   -122.54433606 ry
!   total energy           =   -122.54462686 ry
!   total energy           =   -122.54469206 ry
!   total energy           =   -122.54474000 ry
!   total energy           =   -122.54477263 ry
```

Energije so izpisane v “atomske” enote **Ry** (Rydberg); $1 \text{ Ry} = 1313 \text{ kJ/mol}$.

Stabilno (tj. optimizirano) strukturo dosežemo, ko je razlika med energijama posameznih korakov dovolj majhna in je vsota sil dovolj blizu nič.



2. Stabilnost *cis* in *trans* izomerov 1,2-difluoroetena

Sedaj bomo postopek ponovili še za cis-1,2-difluoroeten (strukturo preverite z `xsmi`).

- `calcmol cis-1,2-difluoroeten "F/cc\F"`

Ko se račun konča, preverite končni energiji obeh izračunov z ukazom `pwene_list`:

- `pwene_list *.out`

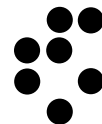
<code>pw.cis-1,2-difluoroeten.relax.out</code>	<code>-122.54736641 Ry</code>
<code>pw.trans-1,2-difluoroeten.relax.out</code>	<code>-122.54467922 Ry</code>

Kaj lahko sklepate iz dobljenega rezultata? Ste takšen rezultat pričakovali? Kolikšna je razlika v kJ/mol?

Namig: struktura z nižjo energijo je stabilnejša. Razliko v energiji lahko pretvorite z:

- `unitconv ry -122.54736641 - -122.54467922`

Opazka: da je cis-1,2-difluoroeten bolj stabilen kot trans-1,2-difluoroeten, je eksperimentalno dejstvo!



3. Stabilnost *cis* in *trans* izomerov 2-butena

V tej vaji bomo poizkusili vrniti kredibilnost kemijski intuiciji in izračunali razlike v energijah med *cis* in *trans* oblikama 2-butena.

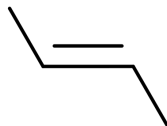
Najprej izvedemo ukaz v terminalu:

- `cd ../vaja-3`

Sedaj se nahajamo v mapi `vaja-3`. Tako bodo vse datoteke, ki jih bomo generirali v tej mapi.

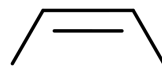
Skeletni strukturi izomerov in pripadajoča zapisa SMILE so izrisani na sliki.

C/c=C/C

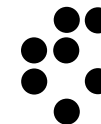


trans-2-buten

C\c=C/C



cis-2-buten



3. Stabilnost *cis* in *trans* izomerov 2-butena

Najprej preverimo, če sta zapisa SMILE pravilna:

- `asmi "C/cc/C"`
- `asmi "C/cc\C"`

Računa lahko poženete z ukazoma:

- `calcmol trans-2-buten "C/cc/C"`

in

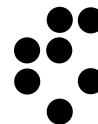
- `calcmol cis-2-buten "C/cc\C"`

Končne izračunane energije bomo ponovno izpisali z ukazom `pwene_list`:

- `pwene_list *.out`

<code>pw.trans-2-buten.relax.out</code>	-54.98825010 Ry
<code>pw.cis-2-buten.relax.out</code>	-54.98544648 Ry

Tokrat ni presenečenje in stabilnejši je *trans* izomer. Kolikšna je ta razlika v kJ/mol?



4. Molekulske orbitale benzena

V tej vaji bomo izrisali valenčne molekulske orbitale benzena.

Najprej izvedemo ukaz v terminalu:

- `cd ../vaja-4`

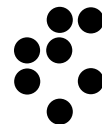
Sedaj se nahajamo v mapi `vaja-4`. Tako bodo vse datoteke, ki jih bomo generirali v tej mapi.

Benzen se v SMILE formatu zapiše kot `c1ccccc1`. Pravilnost zapisa preverimo z:

- `xsmi "c1ccccc1"`

Tokrat ne bomo uporabili privzetih vrednosti `calcmol` ukaza, ampak bomo vhodne parametre nekoliko spremenili, in sicer:

- Izbrali bomo način izračuna `-scf`, ki programu pove naj izračuna samo energijo vhodne geometrije molekule (tj. program ne bo izvedel geometrijske optimizacije). Tako bo račun hitrejši.
- Izbrali bomo tudi opcijo `-mo`, ki pove, da želimo izračunati valenčne molekulske orbitale izbrane molekule, in sicer do najnižje nezasedene orbitale (angl. *Lowest Unoccupied Molecular Orbital* – LUMO).



4. Molekulske orbitale benzena

Molekulske orbitale benzena torej izračunamo z ukazom:

- `calcmol -scf -mo benzen "c1ccccc1"`

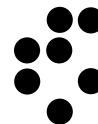
Opazili boste, da je namesto `relax` tokrat tip izračuna `scf`, poleg tega se po končanem izračunu s programom `pw.x` začne nov račun s programom `pp.x`, ki je zadolžen za “post procesiranje”, tj. za izračun molekulskih orbital.

Posamezne izračunane valenčne molekulske orbitale so zapisane v datotekah `psi2.benzen_K001_B001.xsf` do `psi2.benzen_K001_B016.xsf`. Najvišja zasedena valenčna molekulska orbitala (angl. *Highest Occupied Molecular Orbital* – HOMO) benzena ima indeks 15, LUMO orbitala pa indeks 16. Zakaj?

HOMO orbitalo lahko narišete s programom `xcrysden`:

- `xcrysden --xsf psi2.benzen_K001_B015.xsf`

Nato sledite navodilom na naslednji strani.



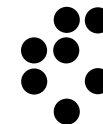
4. Molekulske orbitale benzena

Navodila kako izrisati orbitale s programom **xcrysden** po izvršitvi ukaza:

```
xcrysden --xsf psi2.benzen_K001_B015.xsf
```

- Uporabite meni **Tools-->Data Grid**. Ko se odpre okno, pritisnite gumb **[OK]**
- Odprlo se bo okno, ki kontrolira izris izopovršine; tukaj morate izbrati primerno izovrednost (**Isovalue**), dobra izbira v tem primeru je 0.005
- Obkljukajte tudi opcijo **Render +/- Isovalue** in pritisnite gumb **[Submit]**
- Izberite željen pogled na strukturo
- Shranite izbrano orientacijo preko menuja **File-->Save Current State**, npr. v datoteko z imenom **stanje.xcrysden**
- V terminalu poskusite ta ukaz z drugimi orbitalami (npr. LUMO):

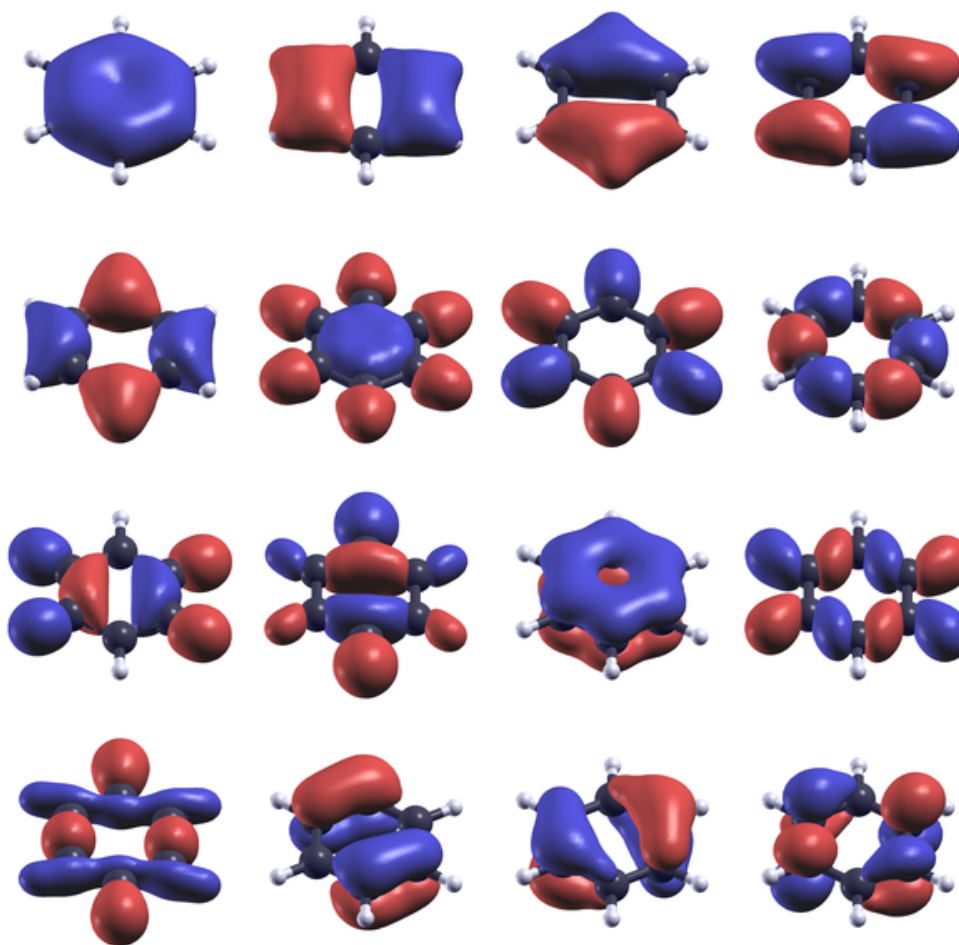
```
xcrysden --xsf psi2.benzen_K001_K016.xsf --script stanje.xcrysden
```

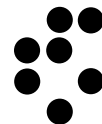


4. Molekulske orbitale benzena

Za referenco so tukaj izrisane vse valenčne molekulske orbitale benzena. Sami lahko to sliko generirate z ukazom `plot-psi2.sh` in sicer:

- `./plot-psi2.sh`





5. Modeliranje po lastni izbiri

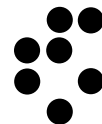
Zdaj ste na vrsti vi. Omejitvi sta edino domišljija in računska moč prenosnika (ne izbirajte prevelikih molekul).

Najprej izvedemo ukaz v terminalu:

- `cd ../vaja-5`

Sedaj se nahajamo v mapi `vaja-5`. Tako so bodo vse datoteke, ki jih boste generirali v sklopu te vaje nahajale v tej mapi.

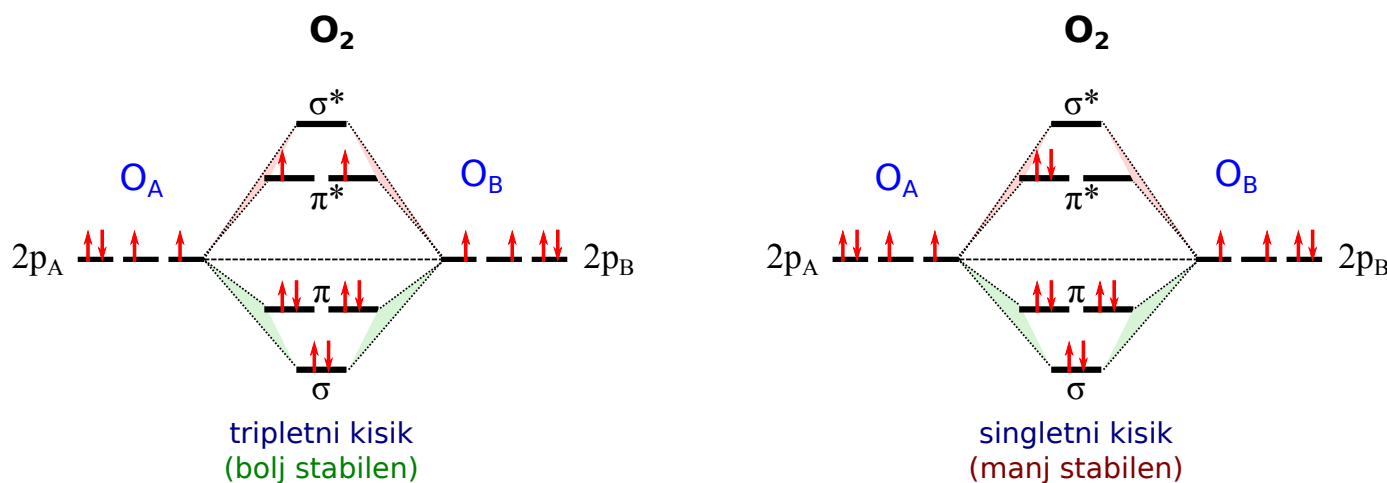




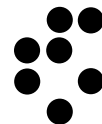
Domača naloga: singletni in tripletni kisik

Kisik je izjema med (preprostimi) molekulami, saj je paramagneten, medtem ko je velika večina molekul nemagnetnih (npr. glej [ta eksperiment](#)). Razlog je v tem, da je osnovno stanje kisika tripletno, kar pomeni, da ima molekula O_2 dva nesparjena elektrona – ti so vir paramagnetnosti. (Vsi permanentni magneti imajo v osnovnem stanju nesparjene elektrone).

Elektronska struktura tripletnega in singletnega kisika je prikazana tu.



Hundovo pravilo pravi, da elektroni zasedejo energijsko enakovredna stanja (orbitale) posamično. To je razlog, zakaj je tripletni kisik bolj stabilen od singletnega.



Domača naloga: singletni in tripletni kisik

Izračunajmo energiji singletnega in tripletnega kisika. Tripletni kisik izračunamo z ukazom:

- `calcmol -magn 2 02-triplet 0=0`

Opcija `-magn 2` zahteva dva nesparjena elektrona (tj. `-magn 2` zahteva “magnetizacijo” z dvema nesparjenima elektronoma). Singletni kisik izračunamo z ukazom:

- `calcmol -smear 0.001 02-singlet 0=0`

Opcijo `-smear 0.001` smo uporabili, da račun lažje konvergira. Singletni kisik ni najstabilnejše (tj. osnovno) stanje, zato račun brez te opcije ne bi konvergiral.

Optimizirani energiji obeh kisikov dobimo z ukazom:

- `pwene_list *02-*.out`

<code>pw.02-triplet.relax.out</code>	<code>-63.47792179 Ry</code>
<code>pw.02-singlet.relax.out</code>	<code>-63.39018847 Ry</code>

Kolikšna je razlika v energiji med singletnim in tripletnim kisikom v kJ/mol?