XACS
Xiamen Atomistic Computing Suite
XACScloud.com

MLatom.com



AI-enhanced computational chemistry

Pavlo O. Dral
MLatom@XACS, Xiamen University
dral@xmu.edu.cn

6 May 2024

Pavlo O. Dral
dr-dral.com

Lecture 3

Why does DFT work? Improving upon it

Pavlo O. Dral

Task 3.1

Calculate with B3LYP/6-31G*:

1.Heats of formation of butane and isobutane.
2.Enthalpy of butane to isobutane isomerization.
3.Compare (give deviation) to experiment values from tasks 1 and 2 (e.g., use the NIST database).

Examples of input files:

https://webbook.nist.gov/chemistry/

Input file 1 for optimization

```
geomopt
method=B3LYP/6-31G*
qmprog=pyscf
xyzfile='...'
```

Should be run separately!

Input file 2 for freq

```
freq # Request frequency and thermochemistry calculations
method=B3LYP/6-31G*
qmprog=pyscf # Use PySCF instead of Gaussian
freqprog=pyscf
XYZfile=optgeoms.xyz # File with optimized geometry
```

Pavlo O. Dral

https://webbook.nist.gov/chemistry

Data compiled as indicated in comments:

ALS - Hussein Y. Afeefy, Joel F. Liebman, and Stephen E. Stein

GT - Glushko Thermocenter, Russian Academy of Sciences, Moscow

Quantity	Value	Units	Method	Reference	Comment
$\Delta_f H°_{gas}$	-30.03 ± 0.16	kcal/mol	Ccb	Pittam and Pilcher, 1972	*ALS*
$\Delta_f H°_{gas}$	-30.37 ± 0.16	kcal/mol	Cm	Prosen, Maron, et al., 1951	see Prosen and Rossini, 1945; *ALS*

Butane

Data compiled as indicated in comments:

ALS - Hussein Y. Afeefy, Joel F. Liebman, and Stephen E. Stein

GT - Glushko Thermocenter, Russian Academy of Sciences, Moscow

Quantity	Value	Units	Method	Reference	Comment
$\Delta_f H°_{gas}$	-32.07 ± 0.15	kcal/mol	Ccb	Pittam and Pilcher, 1972	*ALS*
$\Delta_f H°_{gas}$	-32.42 ± 0.13	kcal/mol	Cm	Prosen, Maron, et al., 1951	see Prosen and Rossini, 1945, Rossini, 1935; *ALS*

Isobutane

Pavlo O. Dral

After calculations finish, which will take some time, you will get results as follows:

Molecule/Reaction	B3LYP/6-31G*	Experiment	Error
butane $\Delta H_{\mathrm{f},298}$	-27.44	-30.03	2.59
isobutane $\Delta H_{\mathrm{f},298}$	-28.25	-32.07	3.82
isomerization $\Delta H_{\mathrm{r},298}$	-0.81	-2.04	1.23

- Errors for heats of formation are quite large - much more than coveted chemical accuracy (errors below 1 kcal/mol).
- However, isomerization enthalpy has error quite close to 1 kcal/mol.
- This is the general trend in quantum chemical calculations: the more similar structures, the less error is in the relative energies.
- **But aren't heats of formation relative energies?!**

Pavlo O. Dral

They are, but structures are very dissimilar, we need to atomize the molecule!

Example for calculating heat of formation of methane molecule:

$$\Delta_{\mathrm{at}}H_{298\,\mathrm{K}}(\mathrm{CH_4}) = \Delta_{\mathrm{f}}H^{\mathrm{exp}}_{298\,\mathrm{K}}(\mathrm{C}) + 4\Delta_{\mathrm{f}}H^{\mathrm{exp}}_{298\,\mathrm{K}}(\mathrm{H}) - \Delta_{\mathrm{f}}H^{\mathrm{exp}}_{298\,\mathrm{K}}(CH_4)$$

$$\Delta_{\mathrm{at}}H^{\mathrm{calc}}_{298\,\mathrm{K}}(\mathrm{CH_4}) = H^{\mathrm{calc}}_{298\,\mathrm{K}}(\mathrm{C}) + 4H^{\mathrm{calc}}_{298\,\mathrm{K}}(\mathrm{H}) - H^{\mathrm{calc}}_{298\,\mathrm{K}}(\mathrm{CH_4})$$

$$\Delta_{\mathrm{f}}H^{\mathrm{calc}}_{298\,\mathrm{K}}(CH_4) \approx \Delta_{\mathrm{f}}H^{\mathrm{exp}}_{298\,\mathrm{K}}(\mathrm{C}) + 4\Delta_{\mathrm{f}}H^{\mathrm{exp}}_{298\,\mathrm{K}}(\mathrm{H}) - \Delta_{\mathrm{at}}H^{\mathrm{calc}}_{298\,\mathrm{K}}(\mathrm{CH_4})$$

Pavlo O. Dral

After calculations finish, which will take some time, you will get results as follows:

Molecule/Reaction	B3LYP/6-31G*	Experiment	Error
butane $\Delta H_{\mathrm{f},298}$	-27.44	-30.03	2.59
isobutane $\Delta H_{\mathrm{f},298}$	-28.25	-32.07	3.82
isomerization $\Delta H_{\mathrm{r},298}$	-0.81	-2.04	1.23

- Errors for heats of formation are quite large - much more than coveted chemical accuracy (errors below 1 kcal/mol).
- However, isomerization enthalpy has error quite close to 1 kcal/mol.
- This is the general trend in quantum chemical calculations: the more similar structures, the less error is in the relative energies.
- In other words, the difference in energies of similar structures is usually small due to the ***error cancellation***.
- That's why DFT works. It banks on error cancellation and you can expect decent results if you calculate relative energies of similar structures.

Pavlo O. Dral

Any ideas, how can we improve upon DFT result?

Pavlo O. Dral

1. Exploite error cancellation

$$\Delta H^{\mathrm{DFT}}_{\mathrm{r},298} = H^{\mathrm{DFT}}_{298}(\mathrm{isobutane}) - H^{\mathrm{DFT}}_{298}(\mathrm{butane})$$
$$\Delta H_{\mathrm{f},298}(\mathrm{isobutane}) = \Delta H^{\mathrm{exp}}_{\mathrm{f},298}(\mathrm{butane}) + \Delta H^{\mathrm{DFT}}_{\mathrm{r},298}$$

Task 3.2

Calculate the heat of formation of isobutane using above equations at B3LYP/6-31G*.

Report:

1. What is the resulting heat of formation?
2. Error in this heat of formation?

Pavlo O. Dral

1. Exploite error cancellation

$$\Delta H^{\mathrm{DFT}}_{\mathrm{r},298} = H^{\mathrm{DFT}}_{298}(\mathrm{isobutane}) - H^{\mathrm{DFT}}_{298}(\mathrm{butane})$$
$$\Delta H_{\mathrm{f},298}(\mathrm{isobutane}) = \Delta H^{\mathrm{exp}}_{\mathrm{f},298}(\mathrm{butane}) + \Delta H^{\mathrm{DFT}}_{\mathrm{r},298}$$

Task 3.2

Calculate the heat of formation of isobutane using above equations at B3LYP/6-31G*.

Report:

1. What is the resulting heat of formation?
2. Error in this heat of formation?

Pavlo O. Dral

In fact, there are many elaborate schemes which were designed to get the reasonable estimates from DFT calculations (check out isodesmic, homodesmotic reactions, etc.).

Pavlo O. Dral

2. Single-point higher-level corrections

Task 3.3

Calculate the isomerization energy of butane to isobutane by performing single-point calculations at CCSD(T)/cc-pVDZ (this will take you a minute; cc-pVTZ will take you half an hour...). Use the geometries optimized at B3LYP/6-31G* in your previous Task 3.1.

Report:
1.What is the resulting isomerization energy?
2.How does it compare to the experimental isomerization enthalpy?
Example of input file:

```
method=CCSD(T)/cc-pVDZ
qmprog=pyscf
yestfile=ccsd-t_energy.dat
XYZfile=optgeoms.xyz # File with geometry optimized with DFT
```

Pavlo O. Dral

CCSD(T)/cc-pVDZ//B3LYP/6-31G* energies for:

- butane (-157.9898929453521 Hartree).
- isobutane (-157.9921731980625).
- the isomerization energy is -1.43 kcal/mol.
- this is pretty close to the experimental isomerization enthalpy (difference is only 0.61 kcal/mol, much better than with DFT!).

Pavlo O. Dral

In principle, you need to calculate isomerization enthalpy, not energy. For this, you can also re-use your old DFT calculations and add ZPVE and thermal correction at B3LYP to CCSD(T) energies. This can be your advanced exercise.

Pavlo O. Dral

3. Use machine learning!

1. All above solutions are still quite time-consuming and not very straightforward to use as you need to do many manual manipulations.
2. A better way to do is to use **AIQM1 or ANI1-ccx**! They are both faster and more accurate and you will get the correct heats of formation out-of-the-box and do not need to any more tricks as the relative energies will automatically be better You can try yourself.
3. Use DFT but its version improved with ML. Universal models following this route start to emerge, with examples being DM21 and CF22D.
4. Build your own ”Δ-learning” model (next lectures...)

Pavlo O. Dral

LETTER | DECEMBER 08 2003

Combined first-principles calculation and neural-network correction approach for heat of formation

LiHong Hu; XiuJun Wang; LaiHo Wong; GuanHua Chen

Check for updates

J. Chem. Phys. 119, 11501–11507 (2003)
https://doi.org/10.1063/1.1630951 **Article history**

Split-Screen PDF Share Tools

Correcting with NN taking number of atoms as input!

Despite their success, the results of first-principles quantum mechanical calculations contain inherent numerical errors caused by various intrinsic approximations. We propose here a neural-network-based algorithm to greatly reduce these inherent errors. As a demonstration, this combined quantum mechanical calculation and neural-network correction approach is applied to the evaluation of standard heat of formation $\Delta_f H^{\ominus}$ for 180 small- to medium-sized organic molecules at 298 K. A dramatic reduction of numerical errors is clearly shown with systematic deviation being eliminated. For example, the root-mean-square deviation of the calculated $\Delta_f H^{\ominus}$ for the 180 molecules is reduced from 21.4 to $3.1\,\mathrm{kcal\,mol^{-1}}$ for $\mathrm{B3LYP/6\text{-}311+G(d,p)}$ and from 12.0 to $3.3\,\mathrm{kcal\,mol^{-1}}$ for $\mathrm{B3LYP/6\text{-}311+G(3df,2p)}$ before and after the neural-network correction.

Pavlo O. Dral

3. Use machine learning!

Use DFT but its version improved with ML. Universal models following this route start to

RESEARCH

QUANTUM CHEMISTRY

Pushing the frontiers of density functionals by solving the fractional electron problem

James Kirkpatrick[1]*†, Brendan McMorrow[1]†, David H. P. Turban[1]†, Alexander L. Gaunt[1]†, James S. Spencer[1], Alexander G. D. G. Matthews[1], Annette Obika[1], Louis Thiry[2], Meire Fortunato[1], David Pfau[1], Lara Román Castellanos[1], Stig Petersen[1], Alexander W. R. Nelson[1], Pushmeet Kohli[1], Paula Mori-Sánchez[3], Demis Hassabis[1], Aron J. Cohen[1,4]*

Density functional theory describes matter at the quantum level, but all popular approximations suffer from systematic errors that arise from the violation of mathematical properties of the exact functional. We overcame this fundamental limitation by training a neural network on molecular data and on fictitious systems with fractional charge and spin. The resulting functional, DM21 (DeepMind 21), correctly describes typical examples of artificial charge delocalization and strong correlation and performs better than traditional functionals on thorough benchmarks for main-group atoms and molecules. DM21 accurately models complex systems such as hydrogen chains, charged DNA base pairs, and diradical transition states. More crucially for the field, because our methodology relies on data and constraints, which are continually improving, it represents a viable pathway toward the exact universal functional.

Kirkpatrick *et al.*, *Science* **374**, 1385–1389 (2021) 10 December 2021

Pavlo O. Dral

3. Use machine learning!

Use DFT but its version improved with ML. Universal models following this route start to emerge, with examples being DM21 and CF22D.

Our benchmarks show that AIQM1 is both faster and more accurate than DM21/6-31G* for standard tests, but DM21 has broader element coverage.

Pavlo O. Dral

Homework 3.1.

Calculate with ANI-1ccx:
1.Heats of formation of butane and isobutane.
2.Enthalpy of butane to isobutane isomerization.
3.Compare to experiment values from tasks 1 and 2 (e.g., use the NIST database).
4.Is it better than B3LYP/6-31G*?

Example of input file for geometry optimization:

```
geomopt
ANI-1ccx
xyzfile='...'
```

Example of input file for frequency calculations:

```
freq # Request frequency and thermochemistry calculations
ANI-1ccx
freqprog=pyscf
XYZfile=optgeoms.xyz # File with optimized geometry
```

Pavlo O. Dral

Homework 3.2

This is more advanced and optional homework for extra credit. Calculate the **enthalpy** at 298 K (not energy!) of the isomerization of butane to isobutane by performing single-point calculations at CCSD(T)/cc-pVDZ (this will take you a minute; cc-pVTZ will take you half an hour…). Use the geometries, zero-point energies, and thermal corrections obtained at B3LYP/6-31G*.

Report:

1. What is the resulting isomerization enthalpy?
2. How different is it from isomerization energy?
3. How does it compare to the experimental isomerization enthalpy?

Pavlo O. Dral