Cuby: integrative framework for computational chemistry

Jan Řezáč

Institute of Organic Chemistry and Biochemistry, AS CR

October 13, 2015

Cuby framework

- Software framework for computational chemistry[1]

- Provides unified acces to multiple software packages and programs
- Implements and automates common computational protocols
- Modular structure allows construction of complex workflows

cuby

Chemistry in Ruby

[1] http://cuby4.molecular.cz

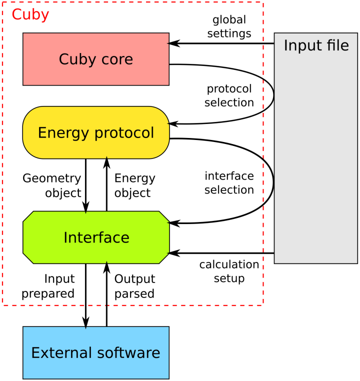
Cuby framework

- Written in Ruby language[2]
  - Modern and powerful
  - Accessible to beginners
  - Runs anywhere
  - Comes from Japan!

[2] http://www.ruby-lang.org

Structure of the framework

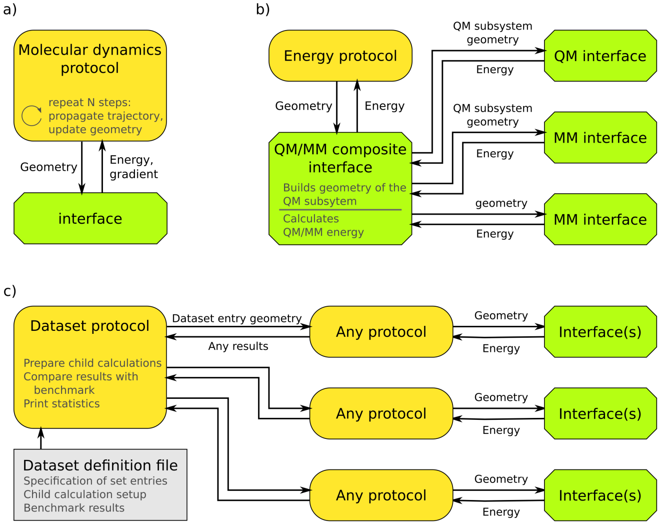
- Object representation of data
- Cuby core managing the calculation
- Two types of modules: *protocols* and *interfaces*



Modularity

- Modular structure also at the level of code
- Modules are fully interchangeable
- Can be arbitrarily combined by the user (from the input file)

Modularity



a)

Molecular dynamics protocol

repeat N steps: propagate trajectory, update geometry

Geometry

Energy, gradient

interface

b)

Energy protocol

Geometry

Energy

QM/MM composite interface

Builds geometry of the QM subsytem

Calculates QM/MM energy

QM subsystem geometry

Energy

QM interface

QM subsystem geometry

Energy

MM interface

geometry

Energy

MM interface

c)

Dataset protocol

Prepare child calculations
Compare results with benchmark
Print statistics

Dataset definition file

Specification of set entries
Child calculation setup
Benchmark results

Dataset entry geometry

Any results

Any protocol

Geometry

Energy

Interface(s)

Any protocol

Geometry

Energy

Interface(s)

Any protocol

Geometry

Energy

Interface(s)

Use cases

- Run calculations in various software without building specific inputs for each
- Combine method A with protocol B not available in code implementing A
- Enjoy automation of tasks built into the available protocols
- Construct complex workflows from the available modules
- Extend the code to do whatever you want

Interfaces

External software

AMBER (MM)
CFour (Coupled clusters)
CHARMM (MM)
deMON (DFT)
DFTB+ (SCC-DFTB)
Gaussian (QM)
Molpro (QM)
MOPAC (Semiempirical)
MRCC (Coupled clusters)
Orca (QM)
Psi4 (QM)
Turbomole (QM)

Composite interfaces

Counterpoise correction
CBS extrapolation
Fragmentation
Potential mixer
Numerical derivatives
QM/MM

Modifiers

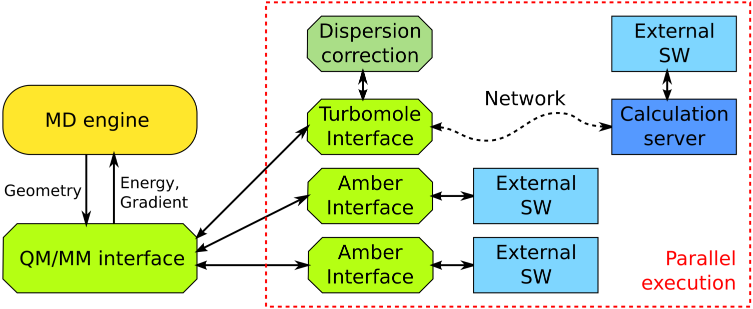
DFT-D dispersion
DFT-D3 dispersion
H4 correction for SQM
X correction for PM6
Restraints
Water droplet restraint

Protocols

- Elementary calculations: energy, gradient, Hessian ...
- Composite quantities: reaction energies, interaction energies (up to 3rd order)
- Drivers: geometry optimization, molecular dynamics, PES scans, external (Gaussian)
- Data set processing
- Geometry manipulation / measurements
- Multi-step protocol (conditional execution)

Example: QM/MM MD

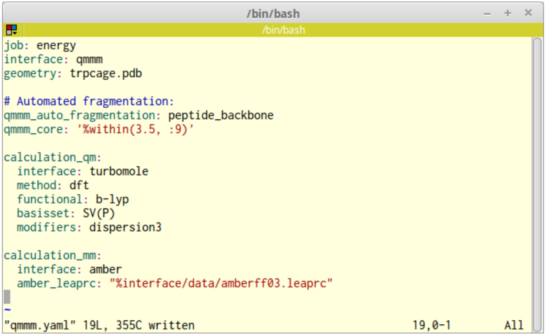
- Any method can be used for both
- Interface nesting possible – multi-layer schemes
- Automated QM region creation for proteins[3]



[3] J. Řezáč *et al.*, Molecules 2015, 20 (3), 4780–4812.

Example: QM/MM MD

- Any method can be used for both
- Interface nesting possible – multi-layer schemes
- Automated QM region creation for proteins[3]

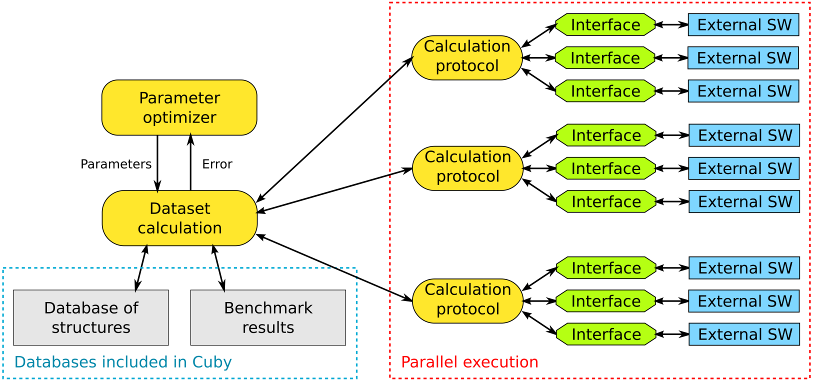


```
job: energy
interface: qmmm
geometry: trpcage.pdb

# Automated fragmentation:
qmmm_auto_fragmentation: peptide_backbone
qmmm_core: '%within(3.5, :9)'

calculation_qm:
  interface: turbomole
  method: dft
  functional: b-lyp
  basisset: SV(P)
  modifiers: dispersion3

calculation_mm:
  interface: amber
  amber_leaprc: "%interface/data/amberff03.leaprc"

~
"qmmm.yaml" 19L, 355C written                                 19,0-1        All
```

[3] J. Řezáč *et al.*, Molecules 2015, 20 (3), 4780–4812.

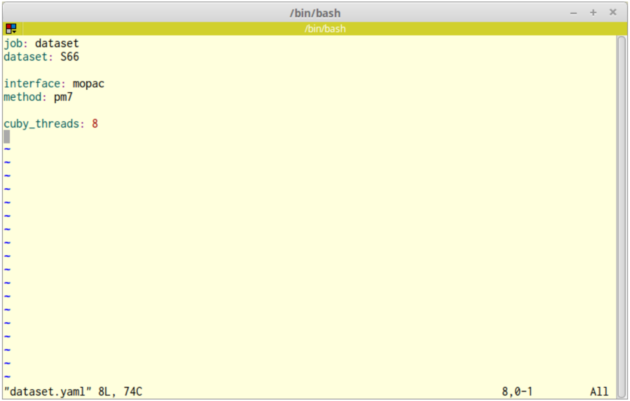
Example: data sets

- Data sets can be defined for repeated use
- Built-in sets of benchmark data
- Cuby prints nice table and plots



Example: data sets

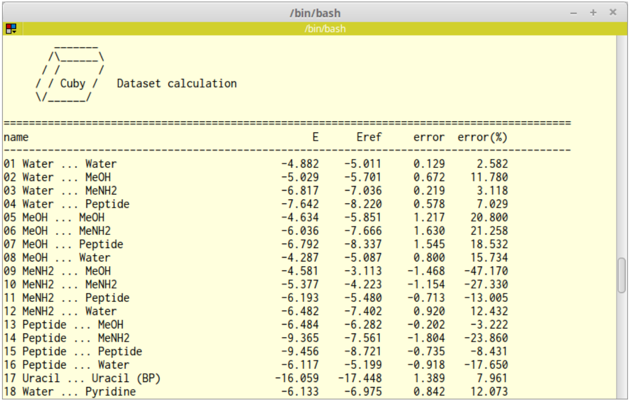
- Data sets can be defined for repeated use
- Built-in sets of benchmark data
- Cuby prints nice table and plots



```
job: dataset
dataset: S66

interface: mopac
method: pm7

cuby_threads: 8

~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
"dataset.yaml" 8L, 74C                                              8,0-1         All
```

Example: data sets

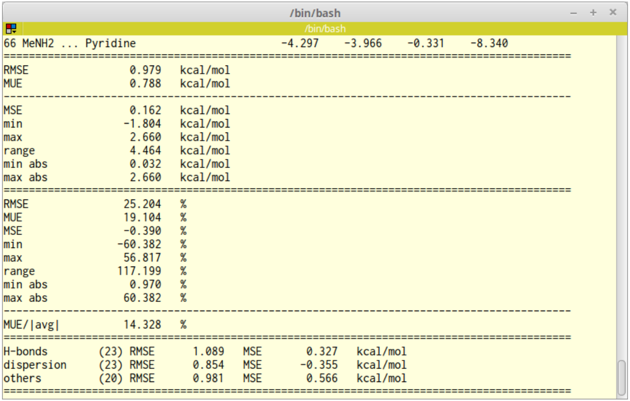
- Data sets can be defined for repeated use
- Built-in sets of benchmark data
- Cuby prints nice table and plots



```
        _______
       /\______\
      / /      /
     / / Cuby /   Dataset calculation
     \/______/

==========================================================================================
name                                             E      Eref     error  error(%)
------------------------------------------------------------------------------------------
01 Water ... Water                          -4.882    -5.011     0.129     2.582
02 Water ... MeOH                           -5.029    -5.701     0.672    11.780
03 Water ... MeNH2                          -6.817    -7.036     0.219     3.118
04 Water ... Peptide                        -7.642    -8.220     0.578     7.029
05 MeOH ... MeOH                            -4.634    -5.851     1.217    20.800
06 MeOH ... MeNH2                           -6.036    -7.666     1.630    21.258
07 MeOH ... Peptide                         -6.792    -8.337     1.545    18.532
08 MeOH ... Water                           -4.287    -5.087     0.800    15.734
09 MeNH2 ... MeOH                           -4.581    -3.113    -1.468   -47.170
10 MeNH2 ... MeNH2                          -5.377    -4.223    -1.154   -27.330
11 MeNH2 ... Peptide                        -6.193    -5.480    -0.713   -13.005
12 MeNH2 ... Water                          -6.482    -7.402     0.920    12.432
13 Peptide ... MeOH                         -6.484    -6.282    -0.202    -3.222
14 Peptide ... MeNH2                        -9.365    -7.561    -1.804   -23.860
15 Peptide ... Peptide                      -9.456    -8.721    -0.735    -8.431
16 Peptide ... Water                        -6.117    -5.199    -0.918   -17.650
17 Uracil ... Uracil (BP)                  -16.059   -17.448     1.389     7.961
18 Water ... Pyridine                       -6.133    -6.975     0.842    12.073
```

Example: data sets

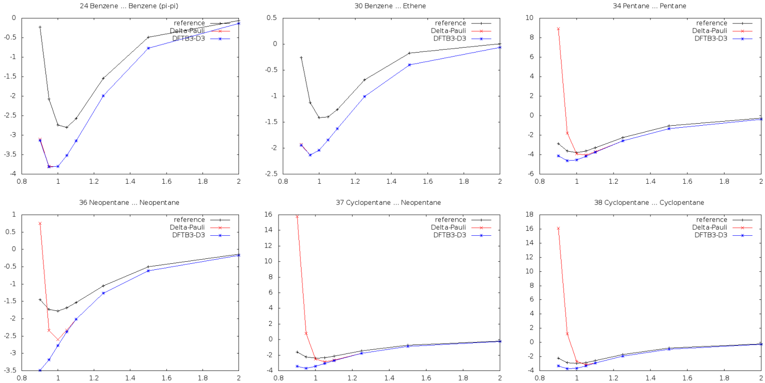
- Data sets can be defined for repeated use
- Built-in sets of benchmark data
- Cuby prints nice table and plots



```
66 MeNH2 ... Pyridine                           -4.297    -3.966    -0.331    -8.340
===========================================================================================
RMSE                 0.979   kcal/mol
MUE                  0.788   kcal/mol
-------------------------------------------------------------------------------------------
MSE                  0.162   kcal/mol
min                 -1.804   kcal/mol
max                  2.660   kcal/mol
range                4.464   kcal/mol
min abs              0.032   kcal/mol
max abs              2.660   kcal/mol
===========================================================================================
RMSE                25.204   %
MUE                 19.104   %
MSE                 -0.390   %
min                -60.382   %
max                 56.817   %
range              117.199   %
min abs              0.970   %
max abs             60.382   %
-------------------------------------------------------------------------------------------
MUE/|avg|           14.328   %
===========================================================================================
H-bonds        (23) RMSE      1.089   MSE       0.327   kcal/mol
dispersion     (23) RMSE      0.854   MSE      -0.355   kcal/mol
others         (20) RMSE      0.981   MSE       0.566   kcal/mol
===========================================================================================
```

Example: data sets

- Data sets can be defined for repeated use
- Built-in sets of benchmark data
- Cuby prints nice table and plots



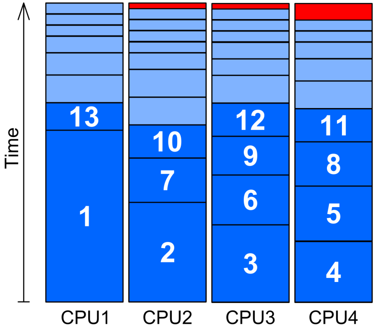
Input file

- Based on the YAML format[4]
- Simple to edit but powerful when needed
- Geometry can be read from external file in many common formats
- Structured input allows construction of complex workflows without programming

[4] http://yaml.org

Paralellization

- Applicable to all composite calculations
- Implemented in the core of the framework
- Effortless access for both user and developer
- Can be combined with parallel execution of external programs



Other features

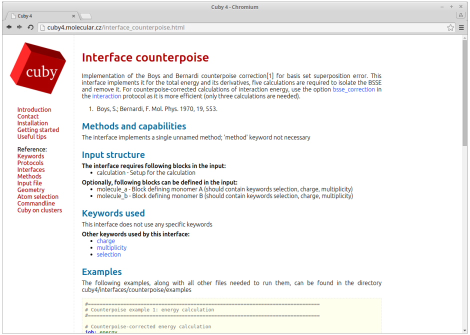
- Powerful atom selection language for manipulating geometries
- Structure generation from SMILES

  ```
  2 C=C -> C1CCC1
  ```

- Interactive shell
- Integration with job managment system on clusters

Documentation

- Online at http://cuby4.molecular.cz
- Accessible from commandline
- Generated from source code - always up to date
- More than 100 examples



Interface counterpoise

Implementation of the Boys and Bernardi counterpoise correction[1] for basis set superposition error. This interface implements it for the total energy and its derivatives, five calculations are required to isolate the BSSE and remove it. For counterpoise-corrected calculations of interaction energy, use the option bsse_correction in the interaction protocol as it is more efficient (only three calculations are needed).

1. Boys, S.; Bernardi, F. Mol. Phys. 1970, 19, 553.

Methods and capabilities

The interface implements a single unnamed method; 'method' keyword not necessary

Input structure

The interface requires following blocks in the input:

- calculation - Setup for the calculation

Optionally, following blocks can be defined in the input:

- molecule_a - Block defining monomer A (should contain keywords selection, charge, multiplicity)
- molecule_b - Block defining monomer B (should contain keywords selection, charge, multiplicity)

Keywords used

This interface does not use any specific keywords

Other keywords used by this interface:

- charge
- multiplicity
- selection

Examples

The following examples, along with all other files needed to run them, can be found in the directory cuby4/interfaces/counterpoise/examples

```
#===============================================================================
# Counterpoise example 1: energy calculation
#===============================================================================

# Counterpoise-corrected energy calculation
job: energy
```

For programmers

- Cuby libraries – object representation of computational chemistry data
- Interface and protocol modules completely separated
  - Easy to add/modify
  - No need to modify/understand the core
- Performance-critical pieces of code written in C

Our projects

- Benchmark calculations of non-covalent interactions
- Development of corrections for semiempirical methods
- QM/MM simulations
- In-silico drug design

Summary

Cuby4

Close to public release
Open source
http://cuby4.molecular.cz
rezac@uochb.cas.cz

Acknowledgements

Pavel Hobza group at IOCB, Prague

Cuby development supported by
Czech Science Foundation