QMC Summer School 2025

Molecular Workflows: Binding & Formation Energies

17 July 2025

Paul Kent, kentpr@ornl.gov

https://github.com/QMCPACK/qmc_summer_school_2025

Funding: U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division, as part of the Computational Materials Sciences Program and Center for Predictive Simulation of Functional Materials.

CPSFM | Center for Predictive Simulation of Functional Materials

Sandia National Laboratories

Argonne NATIONAL LABORATORY

NC STATE UNIVERSITY

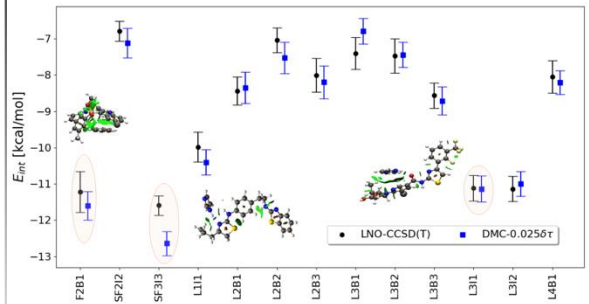
BROWN

OAK RIDGE National Laboratory

Goal: Illustrate a research level workflow for molecules

Recent Molecular Applications of QMC

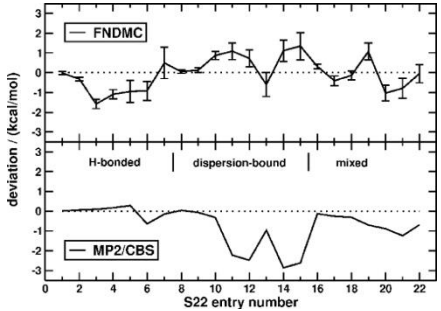
Benchmark accuracy Ligand-pocket interactions



0.3kcal/mol agreement for QMC and LNO-CCSD(T); LDA orbitals, ccECPs, 0.015 a.u. smallest timestep.

[QMeCha] M. Puleva et al. 10.26434/chemrxiv-2025-f6615

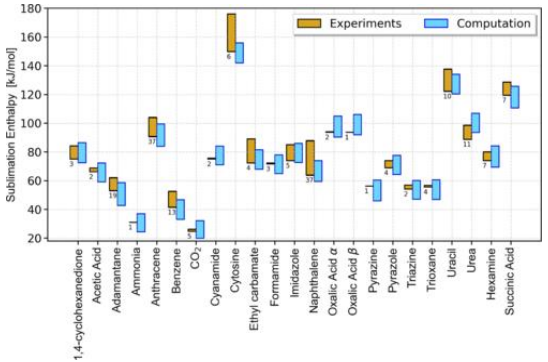
S22 test set for non-covalent interactions



0.68kcal/mol mean absolute deviation (MAD) for S22 test set using HF orbitals

[QMC@HOME] M. Korth at al. J. Phys. Chem. A **112** 2104 (2008)

Lattice Energies of Molecular Crystals



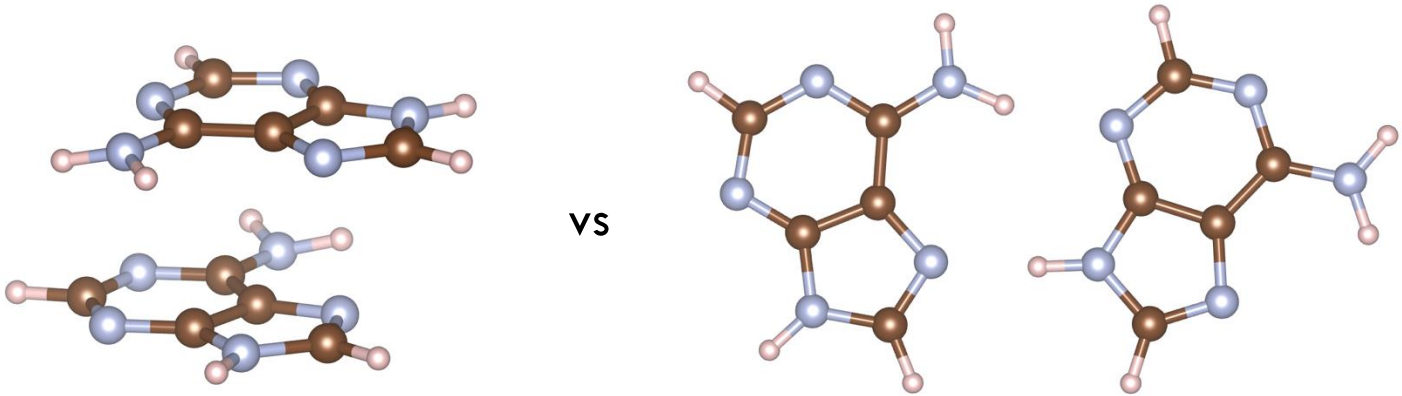
DMC for X23 dataset finds "same or better" reliability than expt (!).

[CASINO] Pia et al. PRL **133** 046401 (2024)

We aim to find a reliable modeling approach for the adenine dimer

To focus on technical aspects, we skip the significant history in quantum chemical & experimental literature. The dimer is part of a larger investigation and a much larger set of molecules and geometries.

Individual runs (30 atoms, 100 electrons) use 64 cores for <24h.



Strategy

Compare DMC results with those of less costly DFT approximations and quantum chemical methods, and for different geometries.

We obtain the relaxed geometries with different DFT approximations, and use these to check the energy ordering and lowest energy structure with other DFTs and DMC. Many other routes possible, e.g., molecular dynamics snapshots.

Exploit QMC variational properties to check nodal surface with different DFT approximations.

Implementation

We only need to generalize the molecular tutorial ☺

Recommend building up step-by-step from simplest workflow

Key points in Nexus workflow script

See example files in session6_molecules

Machine settings for a slurm-based cluster; maximum number of submitted jobs set.

```
# machine and run settings
calc_cores = 64
qmc_bin="qmcpack"
settings(
    pseudo_dir = './pseudopotentials',  # Pseudopotential directory
    generate_only = 0,                  # only write input files, T/F
    results    = '',                    # Don't store results separately
    sleep      = 120,                   # Workflow polling frequency (sec)
#    machine    = 'ws'+str(calc_cores),   # Executing on simple workstation
    machine = 'baseline',               # a cluster with queuing system
    account = 'MAT269',                 # required account name in submitted jobs
    )

ourmachine=get_machine('baseline')
ourmachine.queue_size=24                # limit number of submitted jobs
```

Scan over structures and functionals

Key difference from molecular tutorial: loops over files & functionals. Simple standard python.

```
functionals_for_dft_runs = [ 'LDA', 'PBE0-D4', 'PBE0', 'PBE', 'R2SCAN', 'B3LYP', 'WB97M-V' ]
functionals_for_qmc_runs = [ 'LDA', 'PBE0', 'PBE', 'R2SCAN', 'B3LYP',  'WB97M-V' ]

directory = 'structures'
for struct in os.listdir(directory):
#...
    # Generate the physical system using the atoms found
    system = generate_physical_system(structure=filepath, **atoms_in_file)

    for dftfunc in functionals_for_dft_runs:
        # perform DFT
 #       scf = generate_pyscf(
 # ...
 #           )


        if dftfunc in functionals_for_qmc_runs:
            # convert orbitals to QMCPACK format
            c4q = generate_convert4qmc(
```

Workflow runs a series of interdependent QMC calculations

Workflow builds complexity cautiously:

Determinant only run

J1+J2 optimization then VMC

Further J1+J2 optimization then VMC

J1+J2+J3 optimization then VMC

DMC for 3 timesteps.

Dependencies allow for simultaneous execution of some steps.

```
# run VMC determinant only with QMCPACK. Sanity check / benchmark timings
qmc_hf = generate_qmcpack(
    identifier   = 'vmc_det_only',
    path         = Mysystem+'/'+dftfunc+'/vmc_det_only',
    job          = vmc_job,
    system       = system,
    pseudos      = 'ccecp',
    jastrows     = [],
    qmc          = 'vmc',
    warmupsteps  = 1000,
    blocks       = 100,
    steps        = max(int(48000/calc_cores),1),
    substeps     = 3,
    timestep     = 0.5,
    dependencies = orbdeps,
)


# add+optimize 2-body Jastrow
optJ2_1 = generate_qmcpack(
    identifier        = 'opt_J2_1',
    path              = Mysystem+'/'+dftfunc+'/opt_J2_1',
    job               = opt_job,
    system            = system,
    pseudos           = 'ccecp',
    J2                = True,          # 2-body B-spline Jastrow
    J1_rcut           = 4.0,           # 4 Bohr cutoff for J1
    J2_rcut           = 7.0,           # 7 Bohr cutoff for J2
    qmc               = 'opt',         # quartic variance optimization
    cycles            = 6,             # loop max of 6
    alloweddifference = 1e-3,          # increase allowed energy difference
    samples           = 96000,
    substeps          = 3,
    timestep          = 0.5,
    blocks            = 10,
    warmupsteps       = 1000,
    dependencies      = orbdeps,
)
```

Building and debugging the workflow step-by-step

I ran the workflow script a few sections at a time, commenting out unwanted parts:

For one input geometry file

- PySCF DFT calculations only
- Then determinant only VMC check
- Then initial optimization cycles
- …
- DMC at multiple timesteps (0.005, 0.015, 0.05)

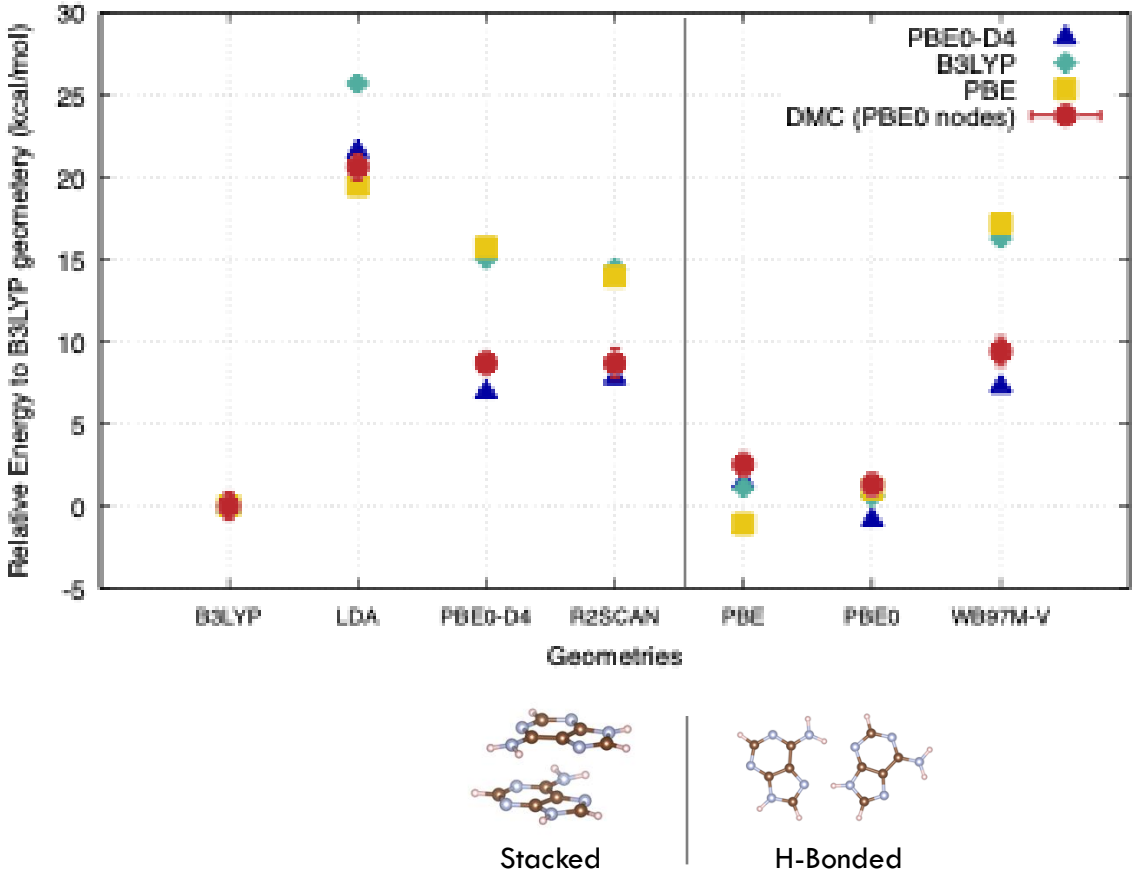
When rerunning the workflow, Nexus will only run the new calculations. Remove directories with modified runs to rerun. Once one geometry is working, add the full set.

Example Results

Geometries strongly depend on functional.

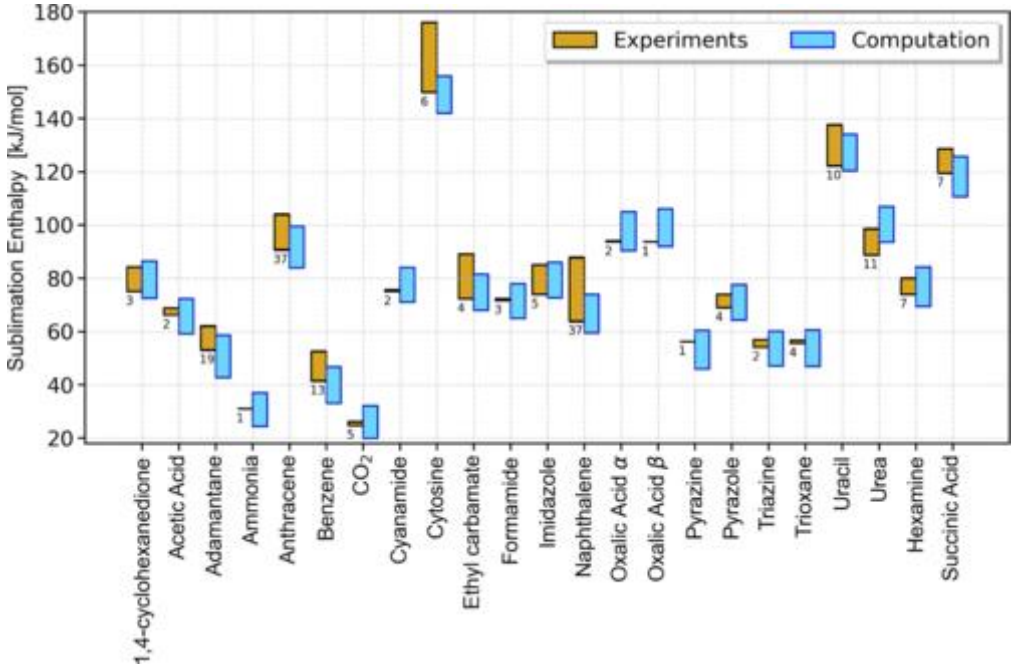
Functionals with dispersion corrections better track the DMC energies.

These workflows can produce a substantial number of jobs and output data.
Make a plan!



Molecular crystals could be run similarly

Limited changes to workflow needed to repeat X23 paper. E.g. Switch to QE/periodic boundary conditions. Authors used 0.01 a.u. timestep, LDA orbitals, J1+J2+J3, "simple" finite size corrections. "1000 CPUh for ~0.5kJ/mol with 504 electrons". QMC sections need little change.



Pia et al.
PRL **133** 046401 (2024)

Summary

The tutorial workflows for molecules are very close to those needed for research level calculations.

I recommend building & testing workflows step-by-step for both insight & ease of debugging.

As shown in recent papers, the “standard recipe” of single determinant trial wavefunctions and DMC gives a highly accurate starting point.