





QM/MM models for solvated and embedded systems (ground state energies, properties and reactivity)

Jeremy Harvey, KU Leuven, Belgium Lecture 3: Challenges of Sampling

Workshop on Theoretical Chemistry: Modelling of Environment Effects in Quantum Chemistry Mariapfarr, 20-23 February 2018 Scientific Organizer: Benedetta Mennucci

Free Energies in QM/MM

- Free energies are directly comparable to experiment, not potential energies
- In QM studies,V → G corrections can be obtained within the rigid-rotor, harmonic oscillator, particle in a box framework of statistical mechanics

$$S = Nk_B + Nk_B \ln\left(\frac{q(V,T)}{N}\right) + Nk_B T\left(\frac{\partial \ln q}{\partial T}\right)_V$$
$$q = q_t \times q_r \times q_v \qquad q_t = \left(\frac{2\pi mk_B T}{h^2}\right)^{\frac{3}{2}} V \quad q_v = \prod_i \left(1 - e^{\frac{-h\nu_i}{k_B T}}\right)^{-1}$$

 This implies that there is well-defined single minimum on the potential energy surface – extensions to small numbers ofadditional conformers are however straightforward.

Free Energies in Extended Systems

This is not applicable to large systems, with very large numbers of thermally populated conformers. A different approach to statistical mechanics must be used, based on *simulation* and approximate construction of the classical partition function:

$$q = \frac{1}{h^{3n}} \iiint \left(\frac{-\left\{ V(X) + \sum_{i} m_{i} v_{i}^{2}/2m_{i} \right\}}{k_{B}T} \right) \mathrm{d}X \mathrm{d}v$$

This partition function cannot be evaluated exactly (far too many degrees of freedom!)

The kinetic energy part can be separated and evaluated exactly (albeit neglecting quauntum mechanical effects)

The potential energy part is challenging

Free Energies in Extended Systems: 2



KU LEUVEN

Biased Sampling Methods

- Molecular dynamics does not visit regions of high μ very frequently. May require unreasonably long simulation. Instead, various biased simulation techniques are used, in which V(X) is modified to visit regions of interest more often
- Many such methods! Example: umbrella sampling, metadynamics.
- Umbrella sampling: multiple simulations are carried out using a set of modified potentials, typically of the form:

$$V_i(X) = V_0 + k(q - q_i)^2$$

• q is a coordinate chosen to match the reaction coordinate

Umbrella Sampling and WHAM

Each simulation yields n(q) hence $\mu(q)$. These are all different:

$$n_i(q) = \exp\left(\frac{-\mu_i(q)}{k_B T}\right) = \exp(-\beta\mu_i(q))$$

Need to *unbias* the simulations to obtain underlying potential of mean force. This is done using the weighted histogram analysis method (WHAM)

See: Kumar, Bouzida, Swendsen, Kollman, Rosenberg, J. Comp. Chem. 1992, 13, 1011; J. Kästner, WIREs Comp. Mol. Sci. 2011, 1, 932



WHAM
$$V_i(X) = V_0 + f_i^b(q(X))$$
 $V_i(X) = V_0 + k(q - q_i)^2$

Modified V are dependent on the biasing potentials $f_i^{b}(q)$ and the set $\{\lambda_i\}$. V_0 has $\lambda_i = 0 \forall i$. From observed distribution $P_{\{\lambda\}}$ with a given bias (i.e. a given set of λ values, usually $\lambda_i = I$, $\lambda_j = O(i \neq j)$), we want to estimate the unbiased probability distribution P_0 :

$$P_0(q') = \frac{\int \exp(-\beta V_0(X))\delta[q-q']dX}{\int \exp(-\beta V_0(X))dX}$$
$$P_i(q') = \frac{\int \exp(-\beta [V_0(X) + f_i^b(q)])\delta[q-q']dX}{\int \exp(-\beta [V_0(X) + f_i^b(q)])dX}$$

Because $f^{b}(q)$ only depends on q, not other coordinates, and because of the $\delta[q-q']$ term, can factorize:

$$P_i(q') = \exp\left(-\beta f_i^b(q')\right) \times \frac{\int \exp(-\beta V_0(X))\delta[q - q(X)]dX}{\int \exp\left(-\beta [V_0(X) + f_i^b(q)]\right)dX}$$
JNH Lect.3, 7

WHAM - 2 Matching:
$$P_0(q') = \frac{\int \exp(-\beta V_0(X))\delta[q-q']dX}{\int \exp(-\beta V_0(X))dX}$$

And: $P_i(q') = \exp(-\beta f_i^b(q')) \times \frac{\int \exp(-\beta V_0(X))\delta[q-q(X)]dX}{\int \exp(-\beta [V_0(X) + f_i^b(q)])dX}$
We get:
 $P_i(q') = P_i(q') \exp(-\beta f_i^b(q')) \times \int \exp(-\beta [V_0(X) + f_i^b(q)])dX$

$$P_0(q') = P_i(q') \exp\left(\beta f_i^b(q')\right) \times \frac{\int \exp\left(-\beta \left[V_0(X) + f_i(q)\right]\right) dX}{\int \exp(-\beta V_0(X)) dX}$$

Which can be rewritten as:

$$P_0(q') = P_i(q') \exp\left(\beta f_i^b(q')\right) \times \left\langle \exp\left(-\beta f_i^b(q)\right) \right\rangle$$

Because the last term in this expression (a mean value over the simulation) can be shown not to depend on q', one has:

$$\mu(q') = -RT \ln P_i(q') - f_i^b(q') + F_i$$

JNH Lect.3, 8 With F_i an unknown constant.

KU LEUVEN

WHAM - 3

Each biased simulation yields an estimate of the overall free energy curve, with its own off-set F_i :

$$\mu(q') = -RT \ln P_i(q') - f_i^b(q') + F_i$$

Each estimate is accurate near the corresponding centre of the biasing potential. They can be combined using WHAM to yield an overall free energy curve

WHAM is rigorous but from the point of view of QM/MM is **very expensive** since it requires multiple molecular dynamics simulations



Simulation with QM/MM

QM/MM potential energy surfaces can be used instead of MM forcefields to carry out simulations of extended systems

For semiempirical QM/MM, this is more or less routine since the cost of the QM/MM energy and gradient is similar to that for MM

For DFT, this is used occa-

$$F = ma$$

$$\frac{\partial V(t)}{\partial x} = F = \frac{\Delta \dot{x}(t \to t + \Delta t)}{\Delta t}$$

$$\dot{x}(t) = v(t) = \frac{\Delta x(t \to t + \Delta t)}{\Delta t}$$

sionally, especially with AIMD-type DFT methods

Umbrella sampling typically requires tens of simulations, each at least 10 ps long. With a 1 fs timestep, this requires > $10^5 E$ + grad! JNH Lect.3, 10

Approximate QM/MM Free Energy Barriers

Against this background, various approaches are used to estimate free energy barriers in the framework of QM/MM

The simplest is to assume $\Delta G^{\ddagger} \cong \Delta E^{\ddagger}$ (reasonable for reactant complex to TS, as no change in molecularity).

Still in this limit, there is the difficulty that each initial structure will yield a different ΔE^{\ddagger} . Which one to take?

Many early QM/MM studies just take one value, sometimes not even checking that the corresponding reactant state and TS are connected.



Approximate QM/MM $\Delta E^{\ddagger} - 2$

- A (slightly) more reliable (but not so rigorous...) $\langle \Delta E^{\ddagger} \rangle = \frac{\sum_{i=1}^{n} \Delta E_{i}^{\ddagger}}{n}$
- Based on ΔE_i^{\ddagger} values computed from starting structures sampled from an MD simulation, and hence assumed to be equally likely.
- Reaction from starting structures leading to high ΔE_i^{\ddagger} are exponentially less likely, so a further refinement is $\langle \Delta E^{\ddagger} \rangle = -RT \ln \left\{ \frac{\sum_{i=1}^n \exp(-\beta \Delta E_i^{\ddagger})}{n} \right\}$ to 'Boltzmann' weight
- This approach yields very high contributions from the smallest ΔE_i^{\ddagger} , and is subject to artefacts ('rare' high-energy reactant states). See: How Many Conformations Need to Be Sampled to Obtain Converged QM/MM Energies? The Curse of Exponential Averaging, U. Ryde, J. Chem. Theory Comput. 2017, **13**, 5745.

JNH Lect.3, 12

KU LEUVEN

Approximate QM/MM $\Delta E^{\ddagger} - 3$

Full sampling as in Umbrella Sampling would imply something like:

$$\left\langle \Delta E^{\ddagger} \right\rangle = -RT \ln \left\{ \frac{\sum_{i} \exp\left(-\beta \Delta E_{i}^{\ddagger}\right) \exp\left(-\beta E_{i}^{\text{reac}}\right)}{\sum_{i} \exp\left(-\beta E_{i}^{\text{reac}}\right)} \right\}$$

As the reactant state energies E_i^{reac} vary a lot, in practice a huge number of energy barriers are needed to get a meaningful average. In practice, Boltzmann averaging with some 'expert' supervision (discarding unreasonable structures, pre-picking of reasonable ones) is close to state of the art nowadays (though clearly unsatisfactory) See: Ulf Ryde (previous slide), or Averaging Techniques for Reaction Barriers in QM/MM Simulations, A. M. Cooper & J. Kästner, ChemPhysChem 2014, **15**, 3264.

Dual-Level Approach for QM/MM Free Energies

Estimating a relative free energy on an accurate potential energy surface from the relative free energy on a less accurate (but less demanding) potential energy surface is attractive ΔE_{T-R}

See review: Eliot Boulanger and J. N. Harvey, QM/MM methods for free energies and photochemistry, Curr. Opinion. Struct. Bio. 2018, **49**, 72.



Dual-Level Approaches

Switching from the low-level reference state to the high-level target can be written in terms of free energy perturbation:

$$\Delta G^{R \to T} = \sum_{m} \delta G^{\lambda_m \to \lambda_{m+1}}$$

This can however be as expensive as full free energy calculation, unless approximations are used

See review: Duarte, Amrein, Blaha-Nelson & Kamerlin, Recent advances in QM/MM methods free energy calculations using reference potentials, Biochim. Biophys. Acta BBA- General, 2015, **1850**, 954.



Dual-Level Approaches – 2

One approximation: linear response approximation:

$$\Delta G^{R \to T} \approx \left\{ \left\langle E_T - E_R \right\rangle_R + \left\langle E_T - E_R \right\rangle_T \right\} / 2$$

KU LEUVEN

Many other approximations, involving e.g. averaging to improve numerical convergence, e.g. construct mean solvent field from MD, and perform QM/MM only in presence of mean solvent



Lecture III Conclusions

- Given a QM/MM Hamiltonian, obtaining free energies remains very challenging
- "Cheap" method typically used in QM studies is not really appropriate
- For the specific case of free energy barriers, some account of conformational complexity is usually mandatory
- Techniques such as umbrella sampling are available and accurate but often too expensive for QM/MM
- Approximate dual-level techniques are currently being developed

