23.6.17

Development of the fragment molecular orbital method combined with DFT and DFTB and applications to proteins

Dmitri G. Fedorov, Kazuo KitauraCD-FMat, AISTFIFC, Kyoto U. & AICS, RIKEN



National Institute of Advanced Industrial Science and Technology (AIST) 独立行政法人 産業技術総合研究所

Fragmentation

chemical, atom-wise definition



Fragmentation methods

Gordon, M. S., et al., Chem. Rev. 2012, 112, 632–672.



One-step: total energy from fragment energies

Two-step: compute density from fragment densities; then energy from the density. **Conglomerate**: overlapping fragments, otherwise (almost) non-overlapping.

FMO Methodology



1. Calculate fragments in the total Coulomb field.

2. Compute pairs of fragments.

3. Calculate total properties.

$$E = \sum_{I=1}^{N} E_{I} + \sum_{I>J}^{N} (E_{IJ} - E_{I} - E_{J})$$

Scaling of FMO

- Number of SCF monomers: linear, N
- Number of **dimers**: **SCF**: linear, $aN(a \approx 4)$ **ES**: quadratic [use multipole summations to add up quickly]
- Embedding (use point charges for far contributions).
 - Total: nearly linear in reality, quadratic in principle.

Quest for understanding

Mulliken:^a "the more accurate the calculations become, the more the concepts tend to vanish into thin air".

Can fragmentation methods condense concepts from thin air?

^a Mulliken, R. S. Molecular Scientists and Molecular Science: Some Reminiscences. *J. Chem. Phys.* **1965**, *43*, S2.

The many-body fragment expansion

D. G. Fedorov, et al. Acc. Chem. Res. 47 (2014) 2846, D. G. Fedorov, WIREs: CMS, in press.

Begin with the original equation:

$$E = \sum_{I=1}^{N} E_{I} + \sum_{I>J}^{N} (E_{IJ} - E_{I} - E_{J})$$

Include higher order terms,

$$E = \sum_{I=1}^{N} E_I + \sum_{I>J}^{N} \Delta E_{IJ} + \dots$$

Add and subtract reference (0) energies E_I^0

$$E = \sum_{I=1}^{N} E_{I}^{0} - \sum_{I=1}^{N} E_{I}^{0} + \sum_{I=1}^{N} E_{I} + \sum_{I>J}^{N} \Delta E_{IJ} + \dots$$
$$= \sum_{I=1}^{N} E_{I}^{0} + \sum_{I}^{N} \Delta E_{I} + \sum_{I>J}^{N} \Delta E_{IJ} + \dots$$



Generalised many-body expansion

Begin with the energy

$$E = \sum_{I=1}^{N} E_I^0 + \sum_{I=1}^{N} \Delta E_I + \sum_{I>J}^{N} \Delta E_{IJ} + \dots$$

Generalise to any size-extensive property A (energy, its gradient, Hessian, electron density etc):

$$A^{M} = A^{0} + \sum_{m=1}^{M} \Delta^{m} A$$

M is the many-body expansion level (typically, $M \le 3$, for FMO so far the largest *M* in literature is 4).

 $\begin{array}{l} History \ tour: \ Hankins, \ D.; \ Moskowitz, \ J. \ W.; \ Stillinger, \ F. \ H. \ J. \ Chem. \ Phys. \ 1970, \ 53, \ 4544-4554. \\ E(\mathbf{x}_{1}\cdots\mathbf{x}_{N}) = \sum_{i=1}^{N} E^{(1)}(\mathbf{x}_{i}) + \sum_{i< j=1}^{N} V^{(2)}(\mathbf{x}_{i}, \mathbf{x}_{j}) \\ + \sum_{i< j< k=1}^{N} V^{(3)}(\mathbf{x}_{i}, \mathbf{x}_{j}, \mathbf{x}_{k}) + \cdots + V^{(N)}(\mathbf{x}_{1}\cdots\mathbf{x}_{N}) \\ \end{array}$

Application to density

Reference density: isolated fragments (molecules).



Triple CT: three-body coupling between pair charge transfers.

The many-body expansion has a clear physical picture.

More on MBE Many-body expansion (MBE)

$$A^{M} = A^{0} + \sum_{m=1}^{M} \Delta^{m} A$$
, *N* is the number of fragments.

- MBE has an exact limit: A is exact for M=N.
- It is not known how to estimate the error, $\varepsilon^{M} = |A^{M} A|$.
- Green's function based reformulation of FMO provides interesting hints about estimating high-order terms.
- Provided that SCF for 1-body calculation converges, MBE also almost always converges.
- *M*-body FMO expansion describes full polarization at the *N*-body level and QM effects (charge transfer etc) at the *M*-body level.

Pair interaction energy decomposition analysis D. G. Fedorov et al., J. Phys. Chem. A 116 (2012) 704



 E_I'' (ΔE_I^{solv}) is the internal (solvation) energy of *I*



Contributions to the pair interaction energy, $\Delta E_{IJ} = \Delta E_{IJ}^{\text{ES}} + \Delta E_{IJ}^{\text{EX}} + \Delta E_{IJ}^{\text{CT+mix}} + \Delta E_{IJ}^{\text{DI}} + \Delta E_{IJ}^{\text{SOLV}}$

electrostatic (ES), **exchange-repulsion** (EX), **charge transfer** (CT+mix), **dispersion** (DI) and **solvent screening** (SOLV).

History tour: EDA (Kitaura, K.; Morokuma, K. *Int. J. Quant. Chem.* **1976**, *10*, 325–340). $\Delta E = \Delta E^{\text{ES}} + \Delta E^{\text{EX}} + \Delta E^{\text{CT}} + \Delta E^{\text{mix}}$

Subsystem analysis of binding

D. G. Fedorov et al., J. Phys. Chem. A 120 (2016) 2218.



Proteins in solution and gas phase Y. Nishimoto et al., PCCP 2016, 18, 22047–22061.



HOMO-LUMO gap of Chignolin

full DFT/PCM/6-31G* (not FMO)

<mark>zwitterionic</mark>

method	Gas phase	solution
HF	6.07	10.90
PBE	(0.09)	1.96
BLYP	(0.03)	2.02
BOP	<mark>(0.04)</mark>	2.03
B3LYP	0.21	4.13
CAM-B3LYP	0.84	7.16
LC-BLYP	2.51	8.77
LC-BOP	2.78	8.79
M11	3.54	9.36
DFTB	0.00	3.17

three functionals diverged



What happens without FMO?

In gas phase:

- Chignolin with PBE, BLYP, BOP in gas phase is metallic, charges flow freely; atomic charges are unchemical (the charge of Gly1 with BLYP, a cation, is -2.8).
- Adding long-range corrections increases the gap magically: atomic charges are as expected (Gly1 with LC-BLYP: 0.9)

In solution:

• Solvent embedding stabilizes HOMO; all methods work; Gly1 charges are BLYP: 1.0, LC-BLYP: 1.0.

What happens with FMO?



In gas phase:

Negative gap: HOMO of one fragment can be above LUMO of another (divergence of SCF and MBE).

In solution:

HOMO is stabilised, no problems.

Pair interactions: DFTB vs DFT

Trp-cage protein (PDB: 1L2Y), 20 residues



1IO5 protein (1961 atoms) is 4840 times faster with DFTB vs DFT.

Optimization of protein structure

FMO-DFTB3/PCM vs experiment (RMSD)



Examples of fragmentation



Electronic excitations in proteins

M. Chiba et al., J. Comput. Chem. 29 (2008) 2667

Photoactive yellow protein (PDB: 2PHY), Cys69

FMO-TDDFT(LC-BLYP)/6-31G*, singlet, eV

Lebarge transfer	3.22
experiment	2.78

CT contributions of residues

-0.05 eV (Ala-067) -0.04 (Phe-096)



Metal ion – water coordination

Y. Nishimoto et al., J. Comput. Chem. 38 (2017) 406

 $Na^++(H_2O)_n$ (*n*=64 for validation, 473 for production) FMO-DFTB, 100 ps NVT MD: (15 days on 1 PC node)



Coordination number 8.9 (FMO2) and 6.5 (FMO3) (others predict: 5-6).

Fullerite cluster optimisation

Y. Nishimoto et al., J. Chem. Theory Comput. 10 (2014) 4801



Fullerite cluster (74.4 \times 74.4 \times 1.4 nm³, 1 030 440 atoms) optimised. Cell length: 14.04 Å (experiment) and 13.91 Å (calculated).

Enzymatic reaction profile

H. Nakata et al., J. Chem. Theory Comput. 11 (2015) 3053

Proton transfer between Triosephosphate isomerase and phosphoglycolohydroxamic (PGH) (9227 atoms).



RHF/D/6-31G* level with PCM

FMO-DFTB drug discovery for GPCR

I. Morao et al., J. Comput. Chem., in press.



Correlation between experiment and DFTB

Correlation between MP2/6-31G* and DFTB



DFTB is about 1000 times faster than MP2 for these protein-ligand complexes (about 600 atoms).

Generalised distributed data interface (GDDI)

D. G. Fedorov et al., J. Comput. Chem., 25 (2004) 872

Divide all CPU cores into groups.





FMO3-MD/6-31G**

(H₂O)₆₄ on supercomputer *Mira* (S. R. Pruitt et al., JCTC, 2016)

Red: ideal Blue: actual

What next? FMO→GMO.

T. Shimazaki et al., J. Chem. Phys. 146 (2017) 084109.

The group molecular orbital (GMO) is based on Huzinaga equations: $(\hat{F}_I + \hat{P}_I)\varphi_i^I = \varepsilon_i^I \varphi_i^I$.

$$E = \sum_{I}^{N} E_{I}$$





The GMO Fock matrix includes embedding and a projection operator.

property	FMO	GMO
<i>n</i> -body level	(1),2,3,(4)	1
basis redundancy	1 atom per boundary	many atoms in tail groups
embedding	1e+2e Coulomb	1e+2e Coulomb+exchange
projection operator	boundary atom split	fragment WF orthogonality
accuracy control	n, fragment size	group size, tail size
program	GAMESS	in-house, Python

Software for FMO http://staff.aist.go.jp/d.g.fedorov

Computations:



(M. S. Gordon, ISU)

GUI:



FIJ



Acknowledgements



H. Nakata Y. Nishimoto A. Heifetz Y. Alexeev K. Kitaura

Takeshi Nagata, Stephan Irle, Mandy Green, Mahito Chiba, Pavel Avramov, Spencer Pruitt, Mark Gordon, Tomomi Shimazaki, Takahito Nakajima, etc etc

Conclusions

FMO is a fast way to do large scale calculations.

IR and Raman spectra, excited states, RDF can be computed.

MBE and interaction analyses are useful to gain insight.

Some systems and properties are problematic: cluster models of covalent crystals, metallic nanoparticles, diffuse functions (large basis sets).