Supplementary Information

Anharmonic Vibrational Analysis of Biomolecules and Solvated Molecules Using Hybrid QM/MM Computations

Kiyoshi Yagi,[†] *Kenta Yamada,*[†] *Chigusa Kobayashi,*[§] *and Yuji Sugita**[†,§,//]

† Theoretical Molecular Science Laboratory, RIKEN Cluster for Pioneering Research, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan.

§ Computational Biophysics Research Team, RIKEN Center for Computational Science, 7-1-26 Minatojima-Minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan.

// Laboratory for Biomolecular Function Simulation, RIKEN Center for Biosystems Dynamics Research, 1-6-5 Minatojima-Minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan.

S1. Implementation of QM/MM in GENESIS

QM/MM in GENESIS is invoked by specifying four keywords in a newly set [QMMM] section as shown in Fig. S1 (a). The QM program is specified by a keyword, "qmtyp". The current version supports Gaussian09/16, Q-Chem 4.x, TeraChem 1.9, and DFTB+ 18.1. "qmcnt" and "qmexe" specify a template to generate an input file and a shell script to execute a QM program, respectively. The template file includes options for QM calculations, but without atomic coordinates and MM charges. As illustrated in Fig. S1 (b), the coordinates and charges are provided by GENESIS at runtime, and supplemented to the template file to generate an input. Then, the shell script is executed by calling a system function in Fortran. Using a shell script enables a flexible control of a QM program; for example, setting a local scratch folder, the number of cores, an archive of molecular orbitals for restarting the self-consistent field (SCF) calculations, etc. Examples are found in our GitHub repository

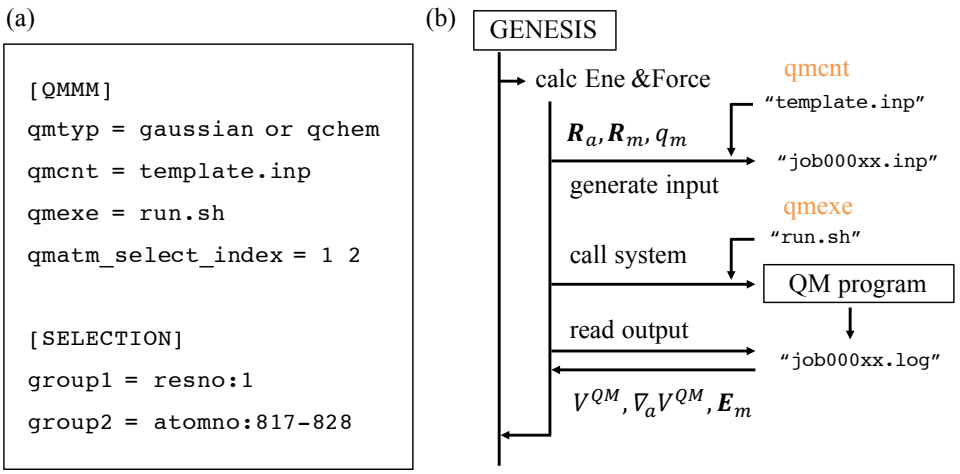

Figure S1. (a) Keywords for QM/MM calculations in an input file and (b) the scheme to interface GENESIS and QM programs.

(https://github.com/yagikiyoshi/QMMMscripts). When the quantum chemistry calculation is done and the control returns to GENESIS, the necessary data are retrieved from log files. The last keyword, "qmatm_select_index", specifies QM atoms using a selector function in GENESIS, which facilitates selection of atoms in terms of atom and residue numbers, segment ID, interatomic distance, etc. Note that link hydrogen atoms are automatically added by the program using the information of bond connectivity provided in a PSF file, thereby requiring no input about the QM and MM boundary.

The QM/MM calculation is often done in a two-step procedure, where the conventional MD simulation is fist performed using a force field to sample relevant structures, and then static QM/MM calculations are performed for those structures to obtain accurate molecular properties. For the convenience of such scheme, we have developed a tool to create QM/MM systems from a MD trajectory (qmmm_generator). Given a QM region and snapshot ID in the input, the tool wraps all molecules in a PBC box around the center of mass of QM atoms (or any center specified by user), and optionally truncates the MM region to generate coordinate files and PSF for each snapshot.

S2. QFF coefficients derived from numerical differentiations of gradient

The third- and fourth-order coefficients of QFF is computed by numerical differentiations of the gradient by the following formula.

1MR-QFF terms:

$$t_{iii} = \frac{\partial^2 g_i}{\partial Q_i^2} \simeq \frac{g_i(+\delta_i) + g_i(-\delta_i) - 2g_i}{\delta_i^2}, \tag{S1}$$

$$u_{iiii} = \frac{\partial^3 g_i}{\partial Q_i^3} \simeq \frac{g_i(+3\delta_i) - 3g_i(+\delta_i) + 3g_i(-\delta_i) - g_i(-3\delta_i)}{8\delta_i^3}. \tag{S2}$$

2MR-QFF terms:

$$t_{iij} = \frac{\partial^2 g_j}{\partial Q_i^2} \simeq \frac{g_j(+\delta_i) + g_j(-\delta_i) - 2g_j}{\delta_i^2}, \tag{S3}$$

$$u_{iiij} = \frac{\partial^3 g_j}{\partial Q_i^3} \simeq \frac{g_j(+3\delta_i) - 3g_j(+\delta_i) + 3g_j(-\delta_i) - g_j(-3\delta_i)}{8\delta_i^3}, \tag{S4}$$

$$u_{iijj} = \frac{\partial^3 g_j}{\partial Q_i^2 \partial Q_j}$$

$$\simeq \frac{1}{2\delta_i^2\delta_j}\left[g_j(+\delta_i, +\delta_j) - g_j(-\delta_i, +\delta_j) + g_j(+\delta_i, -\delta_j) - g_j(-\delta_i, -\delta_j)\right.$$

$$\left. - 2g_j(+\delta_j) + 2g_j(-\delta_j)\right]. \tag{S5}$$

3MR-QFF terms:

$$t_{ijk} = \frac{\partial^2 g_i}{\partial Q_j \partial Q_k} \simeq \frac{g_i(+\delta_j, +\delta_k) - g_i(-\delta_j, +\delta_k) - g_i(+\delta_j, -\delta_k) + g_i(-\delta_j, -\delta_k)}{4\delta_j\delta_k}, \tag{S6}$$

$$u_{iijk} = \frac{\partial^3 g_k}{\partial Q_i^2 \partial Q_j}$$

$$\simeq \frac{1}{2\delta_i^2\delta_j}\left[g_k(+\delta_i, +\delta_j) - g_k(-\delta_i, +\delta_j) + g_k(+\delta_i, -\delta_j) - g_k(-\delta_i, -\delta_j)\right.$$

$$\left. - 2g_k(+\delta_j) + 2g_k(-\delta_j)\right]. \tag{S7}$$

4MR-QFF terms:

$$u_{ijkl} = \frac{\partial^3 g_i}{\partial Q_j \partial Q_k \partial Q_l}$$

$$\simeq \frac{1}{8\delta_j\delta_k\delta_l}\left[g_i(+\delta_j,+\delta_k,+\delta_l) - g_i(-\delta_j,+\delta_k,+\delta_l) - g_i(+\delta_j,-\delta_k,+\delta_l)\right.$$

$$+ g_i(-\delta_j,-\delta_k,+\delta_l) - g_i(+\delta_j,+\delta_k,-\delta_l) + g_i(-\delta_j,+\delta_k,-\delta_l)$$

$$+ g_i(+\delta_j,-\delta_k,-\delta_l)$$

$$\left. - g_i(-\delta_j,-\delta_k,-\delta_l)\right]. \qquad \text{(S8)}$$

Here, g_i and $g_i(+\delta_i)$ are the i-th component of the gradient at the equilibrium geometry and at the grid point deviated by δ_i to the i-th direction, respectively. The number of grid points for calculating the 3MR and 4MR-QFF is,

$$N^{3MR-QFF} = 1 + 4f + 4\binom{f}{2} = 1 + 2f + 2f^2. \qquad \text{(S9)}$$

$$N^{4MR-QFF} = N^{3MR-QFF} + 8\binom{f}{3} = 1 + \frac{14f - 6f^2 + 4f^3}{3}. \qquad \text{(S10)}$$

S3. VSCF and VQDPT2 methods

In this study, the vibrational Schrödinger equation is solved based on *n*MR-PES derived from QM/MM calculations. The VSCF equation is first solved for the vibrational ground state,

$$\left[-\frac{1}{2}\frac{\partial^2}{\partial Q_i^2}+\left\langle\prod_{j\neq i}\phi_0^{(j)}\middle|V\middle|\prod_{j\neq i}\phi_0^{(j)}\right\rangle\right]\phi_0^{(i)}=\epsilon_0^{(i)}\phi_0^{(i)}, \qquad \text{(S11)}$$

which yields a set of one-mode function and energy, $\left\{\phi_{n_i}^{(i)}\right\}$ and $\left\{\epsilon_{n_i}^{(i)}\right\}$, respectively. In VQDPT2, the VSCF solution is improved by the second-order perturbation theory. Let us define VSCF configuration functions and the zero-th order energy as,

$$|\boldsymbol{n}\rangle=\prod_{i=1}^{f}\phi_{n_i}^{(i)}, \qquad \text{(S12)}$$

$$E_{\boldsymbol{n}}^{(0)}=\sum_{i=1}^{f}\epsilon_{n_i}^{(i)}. \qquad \text{(S13)}$$

We divide the VSCF configuration functions to those that are nearly degenerate to the target state, $\{|\boldsymbol{p}\rangle\}$, and others that are non-resonant, $\{|\boldsymbol{q}\rangle\}$. Then, the effective Hamiltonian matrix reads,

$$H_{\boldsymbol{pp'}}^{\text{eff}}=\langle\boldsymbol{p}|\widehat{H}_{vib}|\boldsymbol{p'}\rangle+\sum_{\boldsymbol{q}}\frac{\langle\boldsymbol{p}|\widehat{H}_{vib}|\boldsymbol{q}\rangle\langle\boldsymbol{q}|\widehat{H}_{vib}|\boldsymbol{p'}\rangle}{2}\left\{\frac{1}{E_{\boldsymbol{p}}^{(0)}-E_{\boldsymbol{q}}^{(0)}}+\frac{1}{E_{\boldsymbol{p'}}^{(0)}-E_{\boldsymbol{q}}^{(0)}}\right\}. \qquad \text{(S14)}$$

The diagonalization of this matrix yields the VQDPT2 energy and wavefunction.

The degenerate P space, $\{|\boldsymbol{p}\rangle\}$, and the complementary Q space, $\{|\boldsymbol{q}\rangle\}$, are specified by parameters, k and N_{gen}. For each target state (e.g., fundamentals), VSCF configurations, in which

the difference in quantum number is equal to or less than k, are selected for Q space. If a configuration of the Q space has the zero-th order energy closely degenerate to that of the target configuration, it is selected as a P space configuration. The same procedure is repeated setting the newly selected P space configurations to target state for N_{gen} times. See Ref.[1] for further details.

S4. On the width of Lorentz functions to construct the IR spectrum of a phosphate ion

The IR spectrum of a phosphate ion in solution shown in Fig. 4 was constructed using a Lorentzian line shape function,

$$\Gamma(\nu - \nu_i) = \frac{\frac{\gamma}{2}}{\pi\left((\nu - \nu_i)^2 + \frac{\gamma^2}{4}\right)},$$

where ν_i and γ are the vibrational excitation energy of a state i and a broadening factor, respectively. The weight averaged IR spectra obtained with different γ is shown in Fig. S2.

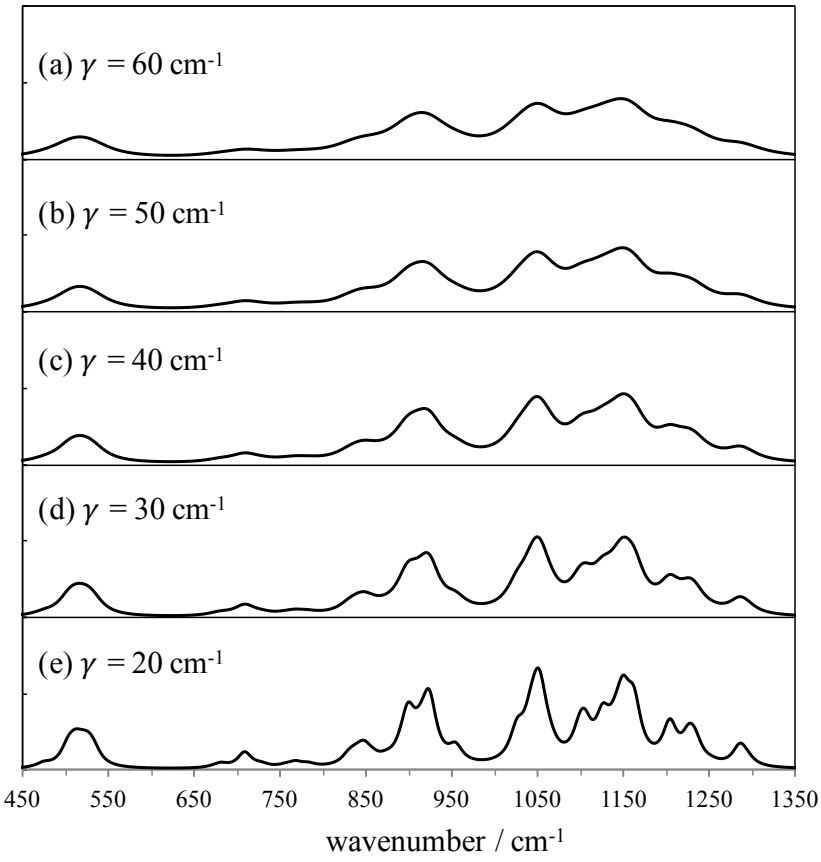

Figure S2. The weight averaged IR spectrum obtained with γ = 20, 30, 40, 50, and 60 cm^{-1}.

S5. On the electronic structure of $Fe^{III}(NO)$

From the result in Fig. 6, the electronic structure of $Fe^{III}(NO)$ can be understood in terms of two types of orbital interaction diagrams. Fig. S3 (a) illustrates a case where an axial ligand (L) weakly interacts with Fe such that the energy levels of *d* orbitals (d_{xz} and d_{yz}) are separated from π^* orbitals of NO. In this case, the character of the orbital is kept unchanged, and the radical electron of NO in the π^* orbital is transferred to one of the *d* orbitals of Fe, literally forming a complex, $Fe^{II}(NO^+)$. Consequently, the two π^* orbitals of NO are both unoccupied, thereby yielding a linear Fe-NO perpendicular to the heme plane. A typical such ligand is histidine, which we confirmed a formation of linear Fe-NO by a QM cluster calculation using imidazole in place of thiolate (data not shown). On the other hand, Fig. S3 (b) illustrates the other extreme, where the axial ligand (e.g. the thiolate of cysteine) strongly interacts with Fe, so that the energy levels of *d* and π^* orbitals are sufficiently close to form strongly mixed, hybrid orbitals. In this case, the unpaired electrons of Fe and NO occupy the hybrid orbital, resulting in a partial occupation of π^* orbitals. The partial occupation of π^* orbitals not only breaks the symmetry to make the NO molecule tilted, but also weakens the N-O bond resulting in a lowering of the NO stretching frequency.

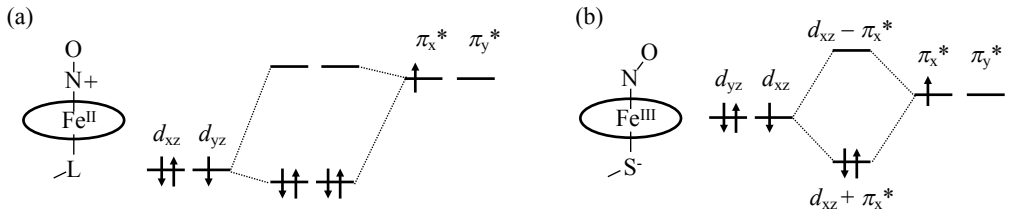

Figure S3. Schematic diagram of orbital interaction giving rise to (a) linear NO and (b) tilted NO.

An alternative configuration, an open-shell singlet (os-1et) state, has one of the unpaired electrons in π^* orbitals of NO, and the other electron with opposite spin in *d* orbitals of Fe^{III}. In other words, the radical electrons remain in unperturbed orbitals.

Table S1. The type of simulation, simulation time (in ns), time step (in fs), and force constants (in kcal mol^{-1} $Å^{-2}$) of harmonic positional restraints for heavy atoms employed in MD simulations for P450nor.

Step	Type[a]	Time	Time step	Restraints	
				Backbone	Other atoms[b]
1[c]	Minimize	5000 steps	-	fixed	fixed
2	NVT	3.0	2.0	fixed	fixed
3	NPT	3.0	2.0	fixed	fixed
4[c]	NVT	3.0	2.0	fixed	fixed
5	NVT	3.0	2.0	5.0	2.0
6	NVT	3.0	2.0	5.0	1.0
7	NVT	3.0	2.0	2.5	1.0
8	NVT	3.0	2.0	2.5	0.5
9	NVT	3.0	2.0	1.0	0.5
10	NVT	3.0	2.0	1.0	0.1
11	NVT	3.0	2.0	1.0	0.0
12	NVT	3.0	2.0	0.5	0.0
13	NVT	3.0	2.0	0.1	0.0
14	NVT	10.0	2.5	0.0	0.0
15[d]	NVT	70.0	2.5	0.0	0.0

a. The temperature and pressure were controlled at 300 K and 1 atm, respectively.

b. Sidechain, heme NO unit, and crystal water molecules.

c. The last snapshot of step 1and 4 is XtalV and XtalW, respectively.

d. Production run. NVT1 to 5 are taken from this trajectory.

Table S2. Representative geometric parameters in the active site of P450nor [bond lengths (r in Å), bond angles (θ in degree), and dihedral angles (ϕ in degree)], and the N-O and Fe-N harmonic stretching vibrations [ω in cm^{-1}] obtained from QM/MM calculations with different sizes of QM region for the NVT4 snapshot, together with the numbers of QM atoms (N_{QM}) and basis functions (N_{bf}).

	QM region A[a]	QM region B[b]	QM region C[c]	QM region D[d]
$r_{Fe–NO}$	1.653	1.650	1.653	1.653
$r_{N–O}$	1.137	1.136	1.136	1.137
$r_{Fe–S}$	2.310	2.314	2.314	2.306
$\theta_{Fe–N–O}$	165.0	168.3	168.7	165.7
$\phi_{NA–Fe–N–O}$	130.8	105.9	103.3	126.3
$\phi_{Cβ–S–N–O}$	125.3	103.8	103.3	123.6
$\omega_{N–O}$	2017.4	2025.2	2026.8	2016.3
$\omega_{Fe–NO}$	628.4	628.7	628.1	629.9
N_{QM}	79	88	102	99
N_{bf}	892	964	1107	1129

[a] A ferric heme-NO unit and a side chain of Cys352 (-$C_\beta H_2 S$).

[b] Three water molecules within 3.5 Å of NO ligand in addition to the QM region A.

[c] Two water molecules forming hydrogen bonds with O atom of Cys352, and backbone of Cys352 and Ile353 (C_α, C, O, and H atoms of Cys352 and N, C_α, H atoms of Ile353) in addition to the QM region B.

[d] C_α, C, O, and H atoms of Cys352, backbone atoms of Ile353 and Ala354, and N, C_α, H atoms of Glu355 in addition to the QM region A.

REFERENCES

1. Yagi, K.; Otaki, H., Vibrational Quasi-Degenerate Perturbation Theory with Optimized Coordinates: Applications to Ethylene and Trans-1,3-Butadiene. *J. Chem. Phys.* **2014,** *140*, 084113.
2. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A., et al. *Gaussian 09 Rev. D.01*, Gaussian Inc., Wallingford, CT, 2009.
3. Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988,** *37*, 785-789.
4. Becke, A. D., Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993,** *98*, 5648-5652.
5. Dunning, T. H., Jr., Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989,** *90*, 1007-1023.
6. Singh, U. C.; Kollman, P. A., An approach to computing electrostatic charges for molecules. *J. Comp. Chem.* **1984,** *5*, 129-145.
7. Besler, B. H.; Merz Jr., K. M.; Kollman, P. A., Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **1990,** *11*, 431-439.
8. Tosha, T.; Nomura, T.; Nishida, T.; Saeki, N.; Okubayashi, K.; Yamagiwa, R.; Sugahara, M.; Nakane, T.; Yamashita, K.; Hirata, K., et al., Capturing an initial intermediate during the P450nor enzymatic reaction using time-resolved XFEL crystallography and caged-substrate. *Nat. Comm.* **2017,** *8*, 1585.
9. Dennington, R.; Keith, T.; Millam, J. *GaussView, Version 5*, Semichem Inc., Shawnee Mission, KS, 2009.
10. Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H., Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **2011,** *7*, 2284-2295.
11. Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H., PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011,** *7*, 525-537.

12. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004,** *25*, 1605-1612.
13. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; Mackerell, A. D., Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *J. Chem. Theory Comput.* **2012,** *8*, 3257-3273.
14. Meuwly, M.; Becker, O. M.; Stote, R.; Karplus, M., NO rebinding to myoglobin: a reactive molecular dynamics study. *Biophys. Chem.* **2002,** *98*, 183-207.
15. Mishra, S.; Meuwly, M., Nitric Oxide Dynamics in Truncated Hemoglobin: Docking Sites, Migration Pathways, and Vibrational Spectroscopy from Molecular Dynamics Simulations. *Biophys. J.* **2009,** *96*, 2105-2118.
16. Grimme, S.; Ehrlich, S.; Goerigk, L., Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011,** *32*, 1456-1465.
17. Mardirossian, N.; Head-Gordon, M., ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **2016,** *144*, 214110.
18. Rappoport, D.; Furche, F., Property-optimized Gaussian basis sets for molecular response calculations. *J. Chem. Phys.* **2010,** *133*, 134105.
19. Weigend, F.; Ahlrichs, R., Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005,** *7*, 3297-3305.
20. Scalmani, G.; Frisch, M. J., Continuous surface charge polarizable continuum models of solvation. I. General formalism. *J. Chem. Phys.* **2010,** *132*, 114110.