**Supporting Information: Does water relay play an important role in phosphoryl transfer reactions? Insights from theoretical study of a model reaction in water and *tert*-butanol**

Yang Yang, Qiang Cui

Department of Chemistry and Theoretical Chemistry Institute,

University of Wisconsin, Madison, 1101 University Ave, Madison, WI 53706

## 1. Test for the MM model of *tert*-butanol

As shown in Table S1, a few key properties (density, diffusion constant and heat of vaporization) are well reproduced by the MM model for the *tert*-butanol.

## 2. Benchmark calculations for SCC-DFTBPR for pseudo-rotation barriers

As pointed by York et al.,[1] conventional semiempirical methods (MNDO, AM1, PM3 and MNDO/d methods are tested) are not reliable for pseudo-rotation and that generally the difference between the semi-empirical and DFT results exceed 50% of the actual barrier at the DFT level. Therefore, it is crucial to test the performance of SCC-DFTBPR in this context to show that the corresponding PMFs that we calculate are meaningful. Accordingly, we have selected representative pseudo-rotation processes from the QCRNA database[2] (10 neutral plus 10 negatively charged cases) and compared the SCC-DFTBPR barriers to high-level DFT data in that database; for further comparison, we have also included results from the second-order SCC-DFTB and a

number of standard semi-empirical methods (see Table S2, S3 and S4). in the **Supporting Materials**).

If the B3LYP/6-31++G(d,p) structures are used and only single point energy calculations are performed, SCC-DFTBPR gives consistently better results than the popular semi-empirical methods. For example, for the 10 neutral systems (5 acyclic and 5 cyclic), SCC-DFTBPR has a RMSE of 2.09 kcal/mol, while the values for AM1, PM3, MNDO are 3.20, 5.50, 3.85 kcal/mol, respectively; the second order SCC-DFTB has a similar RMSE of 2.50 kcal/mol. For the 10 negatively charged cases, the performance becomes worse for all methods; the RMSE for SCC-DFTBPR is 3.72 kcal/mol, and the values for AM1, PM3 and MDNO are 7.51, 5.01 and 5.73 kcal/mol, respectively. The exception is that the second-order SCC-DFTB remains close to the high-level DFT results with a RMSE of 1.6 kcal/mol, even smaller than for the neutral cases.

Besides the single-point calculations, geometry optimizations also indicate that SCC-DFTBPR is a reasonable method for describing pseudo-rotation. For example, for the P(O)(OH)(-O-sugar-O-)(OCH3)_PR case from QCRNA data base, which is structurally similar to the model UNP studied here, the SCC-DFTBPR barrier after geometry optimization is 5.76 kcal/mol, which is only 0.10 kcal/mol higher than the B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p) reference. For more test cases, see Table S4.

These gas-phase benchmark calculations indicate that SCC-DFTBPR can be used to meaningfully compare the energetics of pseudo-rotation for the UNP system in different solutions.

**3. van der Waals parameters for the SCC-DFTBPR atoms**

As shown in Fig. S1, the radial distributions of the solvent (water) near the solute (model UNP) do not change in any major way when the QM region expands from solute-only to include all water molecules within 6 Å from the phosphorus atom or the oxygen in the ribose ring; the results also do not change when different van der Waals parameters (i.e., using the CHARMM parameters vs. optimized values for the key oxygen atoms: $\varepsilon$=-0.1397 kcal/mol; $r_0$=1.698 Å) are used for the solute-QM simulations. This suggests that the QM/MM results reported in the main text are not expected to depend sensitively on the van der Waals parameters for the QM atoms.

**4. Reaction coordinate for the pseudo-rotation PMF calculations**

As shown in Fig. S2, there is clear linear correlations between the approximate reaction coordinate ($d(O_7\text{-}O_{12})$-$d(O_8\text{-}O_{11})$) and the relevant angles that characterize the pseudo-rotation process in both the aqueous and *tert*-butanol simulations. Therefore, the computed PMF profiles can be meaningful compared to characterize the energetic differences between the two solution conditions.

**5. Full Reference for CHARMM 22 force field (Ref. 40 in main text)**

MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, 102, 3586–3616.

**References**


1. (a) C. S. Lopez, O. N. Faza, B. A. Gregersen, X. Lopez, A. R. de Lera and D. M. York, ChemPhysChem. 5, 1045-1049 (2004) (b) C. S. Lopez, O. N. Faza, A. R. de Lera and D. M. York, Chem. Eur. J. 11, 2081-2093 (2005)
2. T. J. Giese et al., and D. M. York, J. Mol. Graph. Model. 25, 423-433 (2006)
3. D. Riccardi, G. Li and Q. Cui, J. Phys. Chem. B 108, 6467-6478 (2004)

**Figure Captions**

**Figure S1.**

Radial distribution functions, g(r) (solid lines) and running coordinate numbers, rcn (dashed lines), of water oxygen around different oxygen atoms of the model UNP in SCC-DFTBPR/MM simulations. Black indicates simulations in which the QM region includes the model UNP and the first solvation shell (within 6 Å from the phosphorus atom or the oxygen atom in the ribose ring; the partitioning of the QM/MM water molecules is updated on the fly). Red/green indicates simulations with only the solute treated as QM; the van der Waals parameters in those simulations are either the standard CHARMM parameters (red) or optimized based on standard protocols (green).[3]

**Figure S2.**

Correlation between the defined reaction coordinate ($d(O_7\text{-}O_{12})\text{-}d(O_8\text{-}O_{11})$) and the relevant angles that characterize the pseudo-rotation process in the aqueous solution (left) and *tert*-butanol (right) PMF simulations. Each data point indicates the average value from a given window. For the atomic labels, see Fig. 1 in the main text.

Table S1. Comparison of computed physical properties to experimental data for *tert*-butanol

| Properties | Experiment [a] | Computation |
|---|---|---|
| Number density ($Å^{-3}$) | 0.006342 | 0.006395 |
| Self-diffusion coefficient ($cm^2/s$) | N/A | $0.77054\times10^{-5}$ |
| Heat of vaporization (kcal/mol) | 10.20 | 11.86 |

a. The value refers to 298.15K; from CRC Handbook.

Table S2. The performance of several semi-empirical methods for the pseudo-rotation barrier of ten neutral phosphates in the QCRNA database [a]

| Reactions | Ring Type | DFT Reference[b] | SCC-DFTB [c] | SCC-DFTBPR[c] | AM1[c] | PM3[c] | MNDO[c] |
|---|---|---|---|---|---|---|---|
| P(OH)(OCH3)(-O-CH2CH2-O-)(OCH3)_PR | -CH2CH2- | 7.86 | -3.95 | -3.09 | -3.24 | -6.23 | -5.11 |
| P(OH)(OH)(-O-CH2CH2-O-)(OCH3)_PR | -CH2CH2- | 7.76 | -3.17 | -3.57 | -3.51 | -6.85 | -5.48 |
| P(OH)(OH)(-O-CH2CH2-O-)(OH)_PR | -CH2CH2- | 5.15 | -2.22 | -0.19 | -1.50 | -4.58 | -3.30 |
| P(OH)(OH)(-O-sugar-O-)(OCH3)_PR | -ribose- | 6.75 | -3.03 | -3.22 | -2.97 | -5.88 | -4.00 |
| P(OH)(OH)(-O-sugar-O-)(OH)_PR | -ribose- | 4.06 | -1.68 | 0.50 | -0.84 | -3.66 | -2.20 |
| | | | | | | | |
| P(OCH3)(OCH3)(OCH3)(OCH3)(OCH3)_PR | acyclic | 4.33 | -0.48 | 0.33 | -2.28 | -3.41 | -1.72 |
| P(OH)(OCH3)(OCH3)(OCH3)(OCH3)_PR | acyclic | 7.60 | -2.30 | -1.17 | -3.91 | -5.72 | -4.50 |
| P(OH)(OH)(OCH3)(OCH3)(OCH3)_PR | acyclic | 8.47 | -2.80 | -2.22 | -4.61 | -6.82 | -5.13 |
| P(OH)(OH)(OCH3)(OH)(OCH3)_PR | acyclic | 6.36 | -2.35 | -1.01 | -3.95 | -5.90 | -2.89 |
| P(OH)(OH)(OCH3)(OH)(OH)_PR | acyclic | 5.81 | -0.99 | -1.80 | -3.22 | -4.74 | -1.69 |
| MAXE[d] | | | -3.95/-3.95/-2.8 | -3.57/-3.57/-2.22 | -4.61/-3.51/-4.61 | -6.85/-6.85/-6.82 | -5.48/-5.48/-5.13 |
| RMSE[d] | | | 2.50/2.91/1.99 | 2.09/2.57/1.46 | 3.20/2.63/3.68 | 5.50/5.56/5.44 | 3.85/4.19/3.49 |
| MUE[d] | | | 2.30/2.81/1.78 | 1.71/2.11/1.31 | 3.00/2.41/3.59 | 5.38/5.44/5.32 | 3.60/-4.02/3.19 |
| MSE[d] | | | -2.30/-2.81/-1.78 | -1.54/-1.91/-1.17 | -3.00/-2.41/-3.59 | -5.38/-5.44/-5.32 | -3.60/4.02/-3.19 |

a) http://theory.chem.umn.edu/QCRNA/. The notations follow the QCRNA convention. For the semi-empirical methods including SCC-DFTB(PR), the values are single point energies at the B3LYP geometries.

b) Gas phase potential energy barrier at the level of B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p) in kcal/mol.

c) Barrier difference from the DFT reference, i.e., $E_{Semiempirical}$-$E_{DFT}$; for AM1, PM3, MNDO, reaction enthalpy is used. "SCC-DFTB" is the second-order method, "SCC-DFTBPR" is the 3rd-order parameterization for phosphate chemistry (Ref. 29 in the main text).

d) MAXE: largest absolute error; RMSE: root-mean-square error; MUE: mean unsigned error; MSE: mean signed error. Before the slash refers to the overall statistics including both the cyclic and acyclic cases; between the slash refers to cyclic cases; after the slash refers to acyclic cases.

Table S3. The performance of several semi-empirical methods for the pseudo-rotation barrier of ten anionic phosphates in the QCRNA database [a]

| Reactions | Ring Type | DFT Reference | SCC-DFTB | SCC-DFTBPR | AM1 | PM3 | MNDO |
|---|---|---|---|---|---|---|---|
| P(O)(OCH3)(-O-CH2CH2-O-)(OCH3)_PR | -CH2CH2- | 0.99 | -0.47 | -0.50 | -4.07 | -3.03 | -2.09 |
| P(O)(OH)(-O-CH2CH2-O-)(OCH3)_PR | -CH2CH2- | 4.70 | -0.95 | -1.71 | -10.51 | -4.98 | -8.36 |
| P(O)(OH)(-O-CH2CH2-O-)(OH)_PR | -CH2CH2- | 2.01 | -1.59 | 1.64 | -10.96 | -7.83 | -6.68 |
| P(O)(OH)(-O-sugar-O-)(OCH3)_PR | -ribose- | 5.66 | -0.15 | -1.80 | -5.37 | -4.72 | -8.15 |
| P(O)(OH)(-O-sugar-O-)(OH)_PR | -ribose- | 1.38 | -0.44 | 2.04 | -3.78 | -3.08 | 0.00 |
| P(O)(OCH3)(OCH3)(OCH3)(OCH3)_PR | acyclic | 1.75 | 1.01 | 1.80 | -4.99 | -3.64 | -0.71 |
| P(O)(OH)(OCH3)(OCH3)(OCH3)_PR | acyclic | 6.65 | -1.32 | -3.62 | -5.61 | -2.17 | -4.50 |
| P(O)(OH)(OCH3)(OH)(OCH3)_PR | acyclic | 4.15 | -1.31 | -0.66 | -8.57 | -6.63 | -3.09 |
| P(O)(OH)(OCH3)(OH)(OH)_PR | acyclic | 8.40 | -3.49 | -6.03 | -8.69 | -7.05 | -7.07 |
| P(O)(OH)(OH)(OCH3)(OCH3)_PR | acyclic | 12.36 | -2.31 | -8.48 | -8.24 | -3.47 | -7.90 |
| MAXE | | | -3.49/-1.59/-3.49 | -8.48/2.04/-8.48 | -10.96/-10.96/-8.69 | -7.83/-7.83/-7.05 | -8.36/-8.36/-7.90 |
| RMSE | | | 1.61/0.88/2.10 | 3.72/1.63/5.00 | 7.51/7.62/7.39 | 5.01/5.04/4.97 | 5.73/6.09/5.34 |
| MUE | | | 1.30/0.72/1.89 | 2.83/1.54/4.12 | 7.08/6.94/7.22 | 4.66/4.73/4.59 | 4.86/5.06/4.65 |
| MSE | | | -1.10/-0.72/-1.48 | -1.73/-0.07/-3.40 | -7.08/-6.94/-7.22 | -4.66/-4.73/-4.59 | -4.86/-5.06/-4.65 |

a. refer to Table S2 for notations.

Table S4. The performance of several semi-empirical methods with geometry optimization for the pseudo-rotation barrier of selected phosphates in the QCRNA database [a]

| Reactions | Charge | DFT Reference | SCC-DFTB | SCC-DFTBPR | AM1 | PM3 | MNDO |
|---|---|---|---|---|---|---|---|
| P(O)(OH)(-O-sugar-O-)(OCH3)_PR | -1 | 5.66 | -0.92[e] | 0.10 | -1.77 | -1.80 | -0.77 |
| P(O)(OH)(-O-sugar-O-)(OH)_PR | -1 | 1.38 | 0.37 | 3.71 | -1.50 | 1.50 | -0.88 |
| | | | | | | | |
| P(OH)(OH)(-O-sugar-O-)(OCH3)_PR | 0 | 6.75 | -2.86 | -3.13 | -6.01 | -5.73 | -3.64 |
| P(OH)(OH)(-O-sugar-O-)(OH)_PR | 0 | 4.06 | -1.74 | -0.22 | -0.22 | -3.00 | -1.20 |

a. Please refer to Table S2 for notation. Optimizations with SCC-DFTB(PR) are done with CHARMM, while those for AM1/PM3/MNDO are done with Gaussian03.

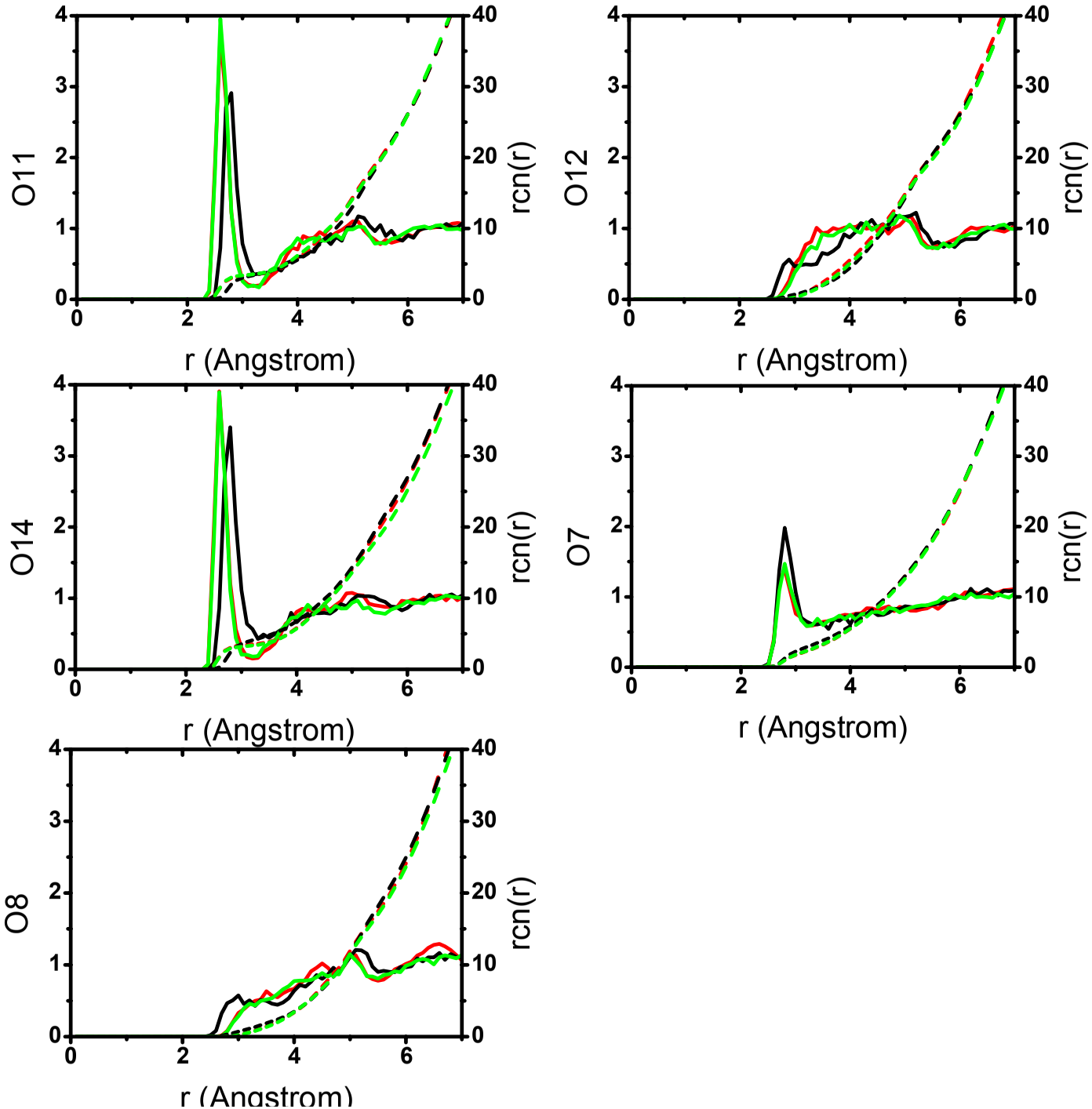


Figure S1

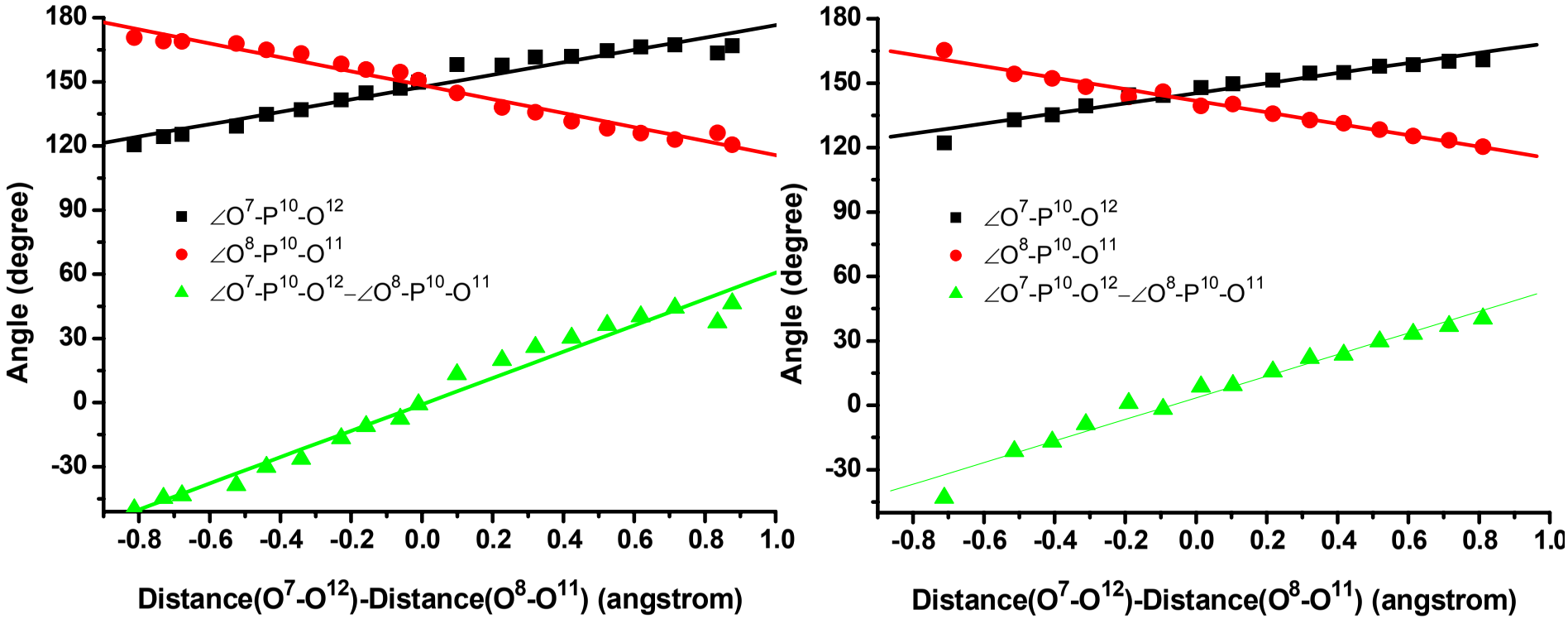
Angle (degree)
180
150
120
90
60
30
0
-30
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
∠O7-P10-O12
∠O8-P10-O11
∠O7-P10-O12−∠O8-P10-O11
Distance(O7-O12)-Distance(O8-O11) (angstrom)

Figure S2