Supplementary Information for

Modelling Phosphate Hydration with a Polarizable Bond-Dipole Framework: Parameter Optimization and Benchmark Testing

Shan-Shan Gao, Meng-Yao Bai, Xiao-Nan Jiang, Lei Wang, Xiao-Lei Li, Qiang Hao*, Chang-Sheng Wang*

School of Chemistry and Chemical Engineering, Liaoning Normal University, Dalian 116029, China.

Contents

1. Model details

The total potential energy computed using the Polarizable Bond-Dipole-Based model (PBFF)[1-3] is formally expressed as

$$E_{\text{total}} = E_{\text{bonded}} + E_{\text{nonbonded}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{es}} + E_{\text{orb}} + E_{\text{vdW}} \tag{S1}$$

The bonded terms include E_{bond}, E_{angle} and E_{dihedral} terms, which represent short-range bond stretching, angle bending, bond-angle coupling, and torsion potential, respectively, and they are defined as follows.

$$E_{\text{bond}} = \sum_{\text{bond}} K_b (b - b_{\text{eq}})^2 \tag{S2}$$

$$E_{\text{angle}} = \sum_{\text{angle}} K_a (a - a_{\text{eq}})^2 \tag{S3}$$

$$E_{\text{dihedral}} = \sum_{\text{dihedral}} \sum_{n} \frac{V_n}{2}\left[1 + \cos(n\varphi - \gamma)\right] \tag{S4}$$

where, K_b is the force constant for bond stretching, where b denotes the actual bond length and b_{eq} its equilibrium value. Similarly, K_a represents the force constant for angle bending, with a being the actual bond angle and a_{eq} the equilibrium bond angle. For torsional potentials, V_n defines the barrier height, n specifies the periodicity (typically 1, 2, or 3), φ is the instantaneous dihedral angle, and γ adjusts the symmetry of the potential energy curve via phase modulation.

The nonbonded terms comprises electrostatic interaction term E_{es}, van der Waals interaction term E_{vdW}, and orbital overlap term E_{orb}.

1.1 Electrostatic interaction

The phosphate-water cluster is composed of one phosphate molecule and several water molecules. The O-H bonds of water molecules, the polar chemical bonds of phosphate molecules (P=O, P–O, C–O, O–H and C–H bonds) are treated as bond dipoles. The dipole direction is oriented from atoms with high electronegativity towards atoms with low electronegativity. The electrostatic interaction between molecules can be expressed as:

$$E_{\text{es}} = \sum_{n \neq j} \frac{Q \cdot \mu_j}{R_{nj}^2} \cdot \cos\varphi + \sum_{i \neq j} \frac{\mu_i \mu_j}{r_{ij}^3} \left(2\cos\theta\cos\theta' + \sin\theta\sin\theta'\cos\zeta\right) \tag{S5}$$

where, Q denotes the net partial charge at an interaction site, typically originating from the formal charge difference between ionized and neutral moieties. μ_i and μ_j represent the i^{th} and j^{th} bond dipole moments, while R_{nj}, φ define the relative distance and orientation angle between the point charge and dipole vector (see Fig. S1a). For dipole-dipole interactions, r_{ij}, θ, θ', and ζ specify the spatial configuration of dipole pairs (see Fig. S1b). Bond dipole moments are expressed as $\mu = \mu_0 + \delta\mu$, where μ_0 represents the permanent bond dipole moment of a stable, isolated molecule, and $\delta\mu$ denotes the induced dipole resulting from the surrounding electrostatic environment. When $\delta\mu = 0$, only the permanent dipole contribution to the electrostatic energy is considered (i.e., $E_{\text{es}} = E_{\text{perm}}$). Conversely, when $\delta\mu \neq 0$, polarization effects are incorporated, and the polarization energy can be expressed as:

$$\begin{aligned} E_{\text{pol}} &= E_{\text{es}} - E_{\text{perm}} \\ &= \sum_{n \neq j} \frac{Q \cdot \delta\mu}{R_{nj}^2} + \sum_{i \neq j} \frac{\mu_i \delta\mu_j + \delta\mu_i \mu_j + \delta\mu_i \delta\mu_j}{r_{ij}^3} \left(2\cos\theta\cos\theta' + \sin\theta\sin\theta'\cos\zeta\right) \end{aligned} \tag{S6}$$

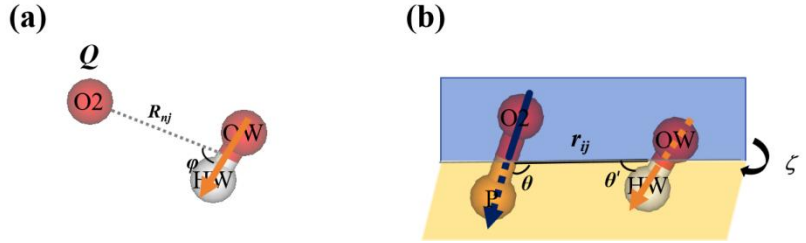


Fig. S1 Schematic representation of electrostatic interactions in the PBFF model: (a) charge–dipole interaction; (b) dipole–dipole interaction.

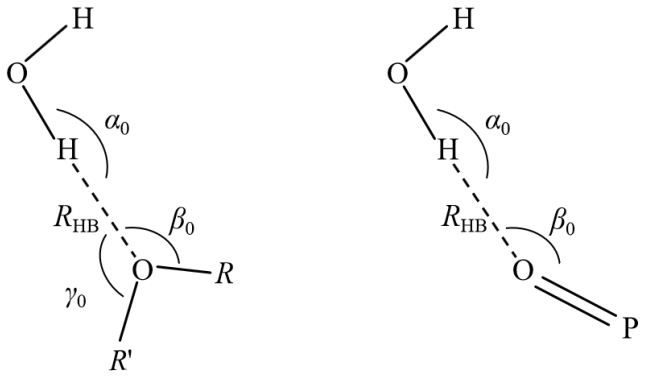
1.2 Orbital interaction

To simplify the treatment of polarization in hydrogen-bonded systems, the direction of the induced bond dipole in PBFF is constrained to align along the chemical bond axis—either in the same or opposite direction as the intrinsic dipole. However, this approximation neglects the full angular flexibility of polarization, particularly in short-range, anisotropic environments such as hydrogen bonds. To partially compensate for this, a correction term—referred to as the orbital term—is introduced to account for directional preferences and orbital overlap effects in hydrogen bonding. The orbital interaction energy is modeled as a function of both the donor–acceptor distance and the geometric arrangement of the hydrogen bond:

$$E_{\mathrm{orb}(sp^3)} = \sum_m D\left[e^{-2a(R_{\mathrm{HB},m}-R_0)} - 2e^{-a(R_{\mathrm{HB},m}-R_0)}\right]^5 \cdot \cos(\alpha_m - \alpha_0)\cos(\beta_m - \beta_0)\cos(\gamma_m - \gamma_0)$$

$$E_{\mathrm{orb}(sp^2)} = \sum_m D\left[e^{-2a(R_{\mathrm{HB},m}-R_0)} - 2e^{-a(R_{\mathrm{HB},m}-R_0)}\right]^5 \cdot \cos(\alpha_m - \alpha_0)\cos(\beta_m - \beta_0) \quad \text{(S7)}$$

here, $R_{\mathrm{HB},m}$ denotes the distance between the proton donor and proton acceptor atoms involved in the m^{th} hydrogen bond within the cluster. α_m, β_m, and γ_m are the angles associated with the m^{th} hydrogen bond. D, a, and R_0 are the action parameters for the type of hydrogen bond; and α_0, β_0, and γ_0 correspond to the ideal values of the angles of α_m, β_m, and γ_m, respectively. As illustrated in Figure S2, hydrogen bonds involving sp³-hybridized lone pairs [e.g., O–H···O(sp³)] adopt idealized angles of α_0 = 180.00°, β_0 = 109.47°, and γ_0 = 109.47°, consistent with tetrahedral geometry. In contrast, systems involving sp²-hybridized acceptor orbitals exhibit different preferred geometries, typically with α_0 = 180.00° and β_0 = 120.00°, reflecting the planar coordination environment. This angular correction enhances the model's ability to represent short-range polarization anisotropy and directionally dependent hydrogen-bonding interactions without increasing computational cost significantly.



(a) O(sp³)···H–O (b) O(sp²)···H–O

Fig. S2 Schematic representation of orbital interactions in hydrogen bonding: (a) O(sp³)···H–O; (b) O(sp²)···H–O.

1.3 Van der Waals interaction

The van der Waals interactions were investigated through the utilisation of the Lennard-Jones 12-6 potential, as represented by equation (S8).

$$E_{\mathrm{vdW}} = \sum_{i \neq j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}\right] \quad \text{(S8)}$$

here, $A_{ij}=\varepsilon_{ij}\cdot(R_{ij}^*)^{12}$, $B_{ij}=2\cdot\varepsilon_{ij}\cdot(R_{ij}^*)^6$, $R_{ij}^*=R_i^*+R_j^*$, $\varepsilon_{ij}=(\varepsilon_i\cdot\varepsilon_j)^{1/2}$, in which R^* is van der Waals radius parameters and ε is van der Waals well depth parameters, and R_{ij} is the distance between atoms i and j.

1.4 Cation parameters for phosphate interactions

To extend the PBFF framework to ion–phosphate complexes, explicit metal cations (Mg^{2+} and Na^+) were introduced to probe their electrostatic interactions with the anionic phosphate group. In contrast to the phosphate esters studied in the main text, where electrostatics are governed primarily by charge–dipole and dipole–dipole interactions, the inclusion of metal cations necessitates activation of the charge–charge term to correctly describe the dominant long-range Coulombic attraction. The complete form of the electrostatic potential employed in PBFF, encompassing charge–charge, charge–dipole, and dipole–dipole interactions, is given by:

$$E_{\mathrm{es}} = \sum_{i \neq j}\frac{Q_i \cdot Q_j}{R_{ij}} + \sum_{n \neq j}\frac{Q \cdot \mu_j}{R_{nj}^2}\cdot\cos\varphi + \sum_{i \neq j}\frac{\mu_i\mu_j}{r_{ij}^3}\left(2\cos\theta\cos\theta' + \sin\theta\sin\theta'\cos\zeta\right) \quad \text{(S9)}$$

For this study, Na^+ and Mg^{2+} ions were parameterized by optimizing their Lennard-Jones (LJ) parameters and point charges to reproduce DLPNO-CCSD(T)/CBS benchmark interaction energy curves for $DMPA^-\cdot M^{n+}$ (M = Na, Mg) complexes. The optimized LJ radius, well depth, and charge are 1.73 Å, 0.00277 kcal·mol^{-1}, and +1.10 e for Na^+, and 0.92 Å, 0.8947 kcal·mol^{-1}, and +2.72 e for Mg^{2+}. All other parameters for the phosphate fragment were adopted from the standard PBFF parameter set (Table S5).

These ionic parameters are designed specifically for assessing the capability of PBFF in modeling strongly polarizing ion–phosphate interactions and are not yet intended for general condensed-phase simulations. Further refinement incorporating polarization damping and solvent screening effects is planned for future work.

2. Many-body expansion of the interaction energy

The Many-body Expansion Method (MBE) represents a significant approach to investigating the interactions between multiple monomers,[4, 5] that divides the cluster into several smaller units for calculation purposes. Consider a cluster system containing N molecules. The total interaction energy of this system is given by:

$$IE_{\mathrm{tot}} = E(1,2,3...n) - \sum_{i=1}^{n} E(i) \tag{S10}$$

where $E(1,2,3...n)$ denoting the total energy of the system and $E(i)$ denoting the one-body clusters (monomers). The idea of the many-body expansion is to decompose the total interaction energy into terms arising from the two-, three-body, etc. Where the total two-body interaction energy is given by:

$$IE_{\mathrm{2B}} = \sum_{i<j}^{n} V_{\mathrm{2B}}(i,j) \tag{S11}$$

with V_{2B} represent respectively the two-body interaction energy between monomers i and j. The total interaction energy of the system is predominantly determined by two-body and three-body interactions, with four-body and higher multibody interactions being relatively insignificant.[6] In light of the aforementioned considerations, the total many-body interaction energy of the cluster system is calculated in this paper as:

$$IE_{\mathrm{mB}} = IE_{\mathrm{tot}} - IE_{\mathrm{2B}} \tag{S12}$$

The individual two-body interaction energy (V_{2B}) of dimer and the individual three-body interaction energy (V_{3B}) of trimer are defined as:

$$V_{\mathrm{2B}}(i,j) = E(i,j) - E(i) - E(j) \tag{S13}$$

$$V_{\mathrm{3B}}(i,j,k) = E(i,j,k) - [E(i) + E(j) + E(k)] - [V_{\mathrm{2B}}(i,j) + V_{\mathrm{2B}}(i,k) + V_{\mathrm{2B}}(j,k)] \tag{S14}$$

where, $E(i, j, k)$ represents the single-point energy of the trimer formed by the i^{th}, j^{th}, and k^{th} molecules within the cluster. $E(i, j)$, $E(i, k)$, and $E(j, k)$ represent the single-point energies of the dimers formed between monomers i and j, monomers i and k, and monomers j and k, respectively. $E(i)$, $E(j)$, and $E(k)$ denote the single-point energies of the i^{th}, j^{th}, and k^{th} monomer, respectively.

3. Computational details

High-precision energy calculations were performed using the ORCA 5.0 software package.[9] The molecular system energies were calculated using the DLPNO-CCSD(T)/CBS method, which integrates the Domain-based Local Pair Natural Orbital (DLPNO) approximation with Complete Basis Set (CBS) extrapolation techniques.[10-13] The total energy consists of the Hartree–Fock (HF) energy and the electron correlation energy, both extrapolated to the CBS limit:

$$E_{\mathrm{CBS}} = E_{\mathrm{SCF}}^{\mathrm{CBS}} + E_{\mathrm{corr}}^{\mathrm{CBS}} \tag{S15}$$

The convergence of the HF energy to the basis set limit is assumed to be given by:[14]

$$E_{\mathrm{SCF}}^{(X)} = E_{\mathrm{SCF}}^{\mathrm{CBS}} + A\exp(-\alpha\sqrt{X}) \tag{S16}$$

where $E_{\mathrm{SCF}}^{(X)}$ is the SCF energy calculated with the basis set with cardinal number X, $E_{\mathrm{SCF}}^{\mathrm{CBS}}$ is the basis set limit SCF energy and A and α are constants. The approach is to do a two-point extrapolation. This means that either A or α have to be known. Here, we take A as to be determined and α as a basis set specific constant.

The correlation energy is supposed to converge as:[10, 15]

$$E_{\mathrm{corr}}^{\mathrm{CBS}} = \frac{X^{\beta}E_{\mathrm{corr}}^{(X)} - Y^{\beta}E_{\mathrm{corr}}^{(Y)}}{X^{\beta} - Y^{\beta}} \tag{S17}$$

For the benchmark calculations in this work, the def2-TZVPP and def2-QZVPP basis sets were used for extrapolation, corresponding to cardinal numbers X=3 and Y=4, respectively. The values of α and β are 7.88 and 2.97 respectively.[15] Integral evaluations were accelerated using the RIJCOSX approximation. The natural orbital truncation thresholds were set to normalPNO ($T_{\mathrm{CutPNO}} = 3.33\times10^{-7}$, $T_{\mathrm{CutPairs}} = 10^{-4}$), and the SCF convergence criterion is tightSCF. Hereafter, the notation "DLPNO-CCSD(T)/CBS" specifically refers to this well-defined protocol.

To validate the computational reliability and convergence of the DLPNO-CCSD(T)/CBS method employed in this study — which utilizes def2-TZVPP/def2-QZVPP basis set extrapolation with NormalPNO settings — systematic verification was carried out on phosphate-water dimer and trimer structures. First, calculations

denoted as DLPNO-CCSD(T)/CBS1 were performed using extrapolation based on aug-cc-pVTZ (X=3) and aug-cc-pVQZ (X=4) basis sets with NormalPNO settings, where the HF and correlation energy extrapolations employed exponents α = 5.79 and β = 3.05, respectively.[15] These results were rigorously compared against reference benchmark values. The observed maximum absolute errors (MAE) were 0.22 kcal/mol for dimer interaction energies and 0.20 kcal/mol for trimer three-body interaction energies (see Tables S2 and S3), which are within acceptable tolerances for high-level quantum chemical methods, confirming the accuracy of the chosen protocol. Furthermore, to assess the sensitivity to pair natural orbital truncation, additional calculations labeled DLPNO-CCSD(T)/CBS2 were conducted using def2-TZVPP/def2-QZVPP extrapolation with TightPNO settings. For dimer interaction energies, the MAE between NormalPNO and TightPNO results was 0.36 kcal/mol, with a root-mean-square error (RMSE) of 0.20 kcal/mol. For trimer three-body interactions, the MAE was reduced to 0.23 kcal/mol, with an RMSE of 0.07 kcal/mol (See Tables S2 and S3). This trend indicates that the PNO method maintains high numerical convergence and computational stability even as system complexity increases. Additionally, MP2/aug-cc-pVQZ level calculations were performed for DMPA-nW, MPA-nW, DEP-nW, TMP-nW, and DMPH-nW complexes. These results showed excellent agreement with DLPNO-CCSD(T)/CBS values, with Pearson correlation coefficients exceeding 0.9999 (See Tables S11). This strong correlation further affirms the reliability of the DLPNO-CCSD(T)/CBS approach, even for anionic systems. Together, these validation efforts provide a solid theoretical foundation for subsequent force field parameterization and development.

Table S1 Calculated dipole moments (μ, in Debye) for monomers in Training Set I

label \ method	B3LYP/aug-cc-pVTZ				PBFF				error			
	μ_X	μ_Y	μ_Z	μ_{Tot}	μ_X	μ_Y	μ_Z	μ_{Tot}	$\Delta\mu_X$	$\Delta\mu_Y$	$\Delta\mu_Z$	$\Delta\mu_{Tot}$
DMPAgg	0.00	0.00	-5.29	5.29	0.00	0.00	-5.31	5.31	0.00	0.00	-0.02	0.02
DMPAgt	-0.34	-4.73	-0.64	4.78	-0.19	-4.51	-0.92	4.61	0.15	0.22	-0.28	-0.17
DMPAtt	0.00	0.00	-2.90	2.90	0.00	0.00	-2.92	2.92	0.00	0.00	-0.02	0.02
DEPAgg	0.00	-6.70	0.00	6.70	0.00	-7.03	0.00	7.03	0.00	-0.33	0.00	0.33
DEPAgt	0.09	-6.09	0.29	6.10	0.22	-6.20	-0.03	6.21	0.13	-0.11	-0.32	0.11
DEPAtt	0.00	-4.38	0.00	4.38	0.00	-4.65	0.00	4.65	0.00	-0.27	0.00	0.27
MPAgg	3.45	-2.90	-0.40	4.53	3.20	-2.96	-0.14	4.37	-0.25	-0.06	0.26	-0.16
MPAgt	2.39	-2.55	-1.11	3.66	2.41	-2.53	-0.91	3.61	0.02	0.02	0.20	-0.05
TMPC1	0.41	0.42	-3.70	3.74	0.51	0.27	-3.69	3.74	0.10	-0.15	0.01	0.00
TMP C3	0.00	0.01	-1.11	1.11	0.00	0.01	-1.28	1.28	0.00	0.00	-0.17	0.17
TMP Cs	-3.64	-0.17	0.00	3.64	-3.81	-0.35	0.00	3.83	-0.17	-0.18	0.00	0.19
TEP C1	-0.75	-0.11	-3.41	3.50	-0.62	0.13	-3.84	3.89	0.13	0.24	-0.43	0.39
TEP C3	0.00	0.01	-1.38	1.38	0.01	0.02	-1.38	1.38	0.01	0.01	0.00	0.00
TEP Cs	-3.24	-1.85	0.00	3.74	-3.22	-2.26	0.00	3.93	0.02	-0.41	0.00	0.19
DMPH gg	0.34	2.48	2.60	3.61	0.10	2.19	2.70	3.48	-0.24	-0.29	0.10	-0.13
DMPH gt	-0.27	-0.84	-0.86	1.23	-0.36	-0.52	-0.89	1.09	-0.09	0.32	-0.03	-0.14
DMPH tt	0.00	0.66	-1.33	1.49	0.00	1.22	-1.46	1.90	0.00	0.56	-0.13	0.41
RMSE									0.11	0.25	0.18	0.20

Table S2 The individual three-body interaction energies (V_{3B}, in kcal·mol^{-1}) for phosphate-water trimers

label \ method	DLPNO-CCSD(T)/CBS1[a]	DLPNO-CCSD(T)/CBS2[b]	DLPNO-CCSD(T)/CBS
DMPA-2W$_1$	0.95	0.82	0.80
DMPA-2W$_2$	0.59	0.63	0.54
DMPA-2W$_3$	0.86	0.84	0.90
DMPA-2W$_4$	-0.43	-0.28	-0.35
DMPA-2W$_5$	-4.22	-4.23	-4.20
DMPA-2W$_6$	-4.33	-4.34	-4.29
DEPA-2W$_1$	0.63	0.57	0.48
DEPA-2W$_2$	-0.39	-0.36	-0.43
DEPA-2W$_3$	-3.66	-3.69	-3.66
DEPA-2W$_4$	-4.14	-4.19	-4.16
MPA-2W$_1$	0.44	0.69	0.46
MPA-2W$_2$	-0.21	-0.20	-0.23
MPA-2W$_3$	-3.72	-3.75	-3.70
MPA-2W$_4$	-3.29	-3.33	-3.35
TMP-2W$_1$	-2.45	-2.44	-2.47
TMP-2W$_2$	-1.36	-1.26	-1.32
TEP-2W$_1$	-2.52	-2.43	-2.55
TEP-2W$_2$	0.55	0.56	0.57
DMPH-2W$_1$	-5.44	-5.48	-5.53
DMPH-2W$_2$	-2.39	-2.39	-2.45
DMPH-2W$_3$	-1.33	-1.24	-1.28
DMPH-2W$_4$	-1.58	-1.59	-1.54
2DMPH-W	-7.78	-7.89	-7.98
3DMPH	-2.66	-2.60	-2.63

[a]DLPNO-CCSD(T)/CBS1: CBS extrapolation using aug-cc-pVTZ and aug-cc-pVQZ basis sets (α = 5.79, β = 3.05) with NormalPNO settings.
[b]DLPNO-CCSD(T)/CBS2: CBS extrapolation using def2-TZVPP and def2-QZVPP basis sets with TightPNO settings.

Table S3 The individual two-body interaction energies (V_{2B}, in kcal·mol^{-1}) for phosphate-water dimers

label \ method	DLPNO-CCSD(T)/CBS1[a]	DLPNO-CCSD(T)/CBS2[b]	DLPNO-CCSD(T)/CBS
DMPAgg-W	-16.04	-16.43	-16.26
DMPAgt-W	-17.55	-17.75	-17.55
DMPAtt-W_1	-13.20	-13.49	-13.29
DMPAtt-W_2	-17.78	-18.20	-17.92
DEPAgg-W	-16.07	-16.42	-16.26
DEPAgt-W	-17.42	-17.61	-17.41
DEPAtt-W_1	-13.26	-13.59	-13.43
DEPAtt-W_2	-17.76	-18.07	-17.71
MPAgg-W_1	-16.87	-17.04	-16.90
MPAgg-W_2	-17.82	-17.97	-17.65
MPAgt-W_1	-15.83	-15.93	-15.66
MPAgt-W_2	-13.10	-13.33	-13.11
TMPC1-W	-8.90	-9.01	-8.90
TMPC3-W	-8.62	-8.79	-8.66
TMPCs-W	-5.57	-5.71	-5.59
TEPC1-W	-8.99	-9.07	-9.00
TEPC3-W	-8.84	-9.03	-8.89
TEPCs-W	-5.73	-5.98	-5.79
DMPHgg-W	-8.60	-8.77	-8.68
DMPHgt-W	-14.00	-14.14	-14.00
DMPHtt-W	-4.62	-4.75	-4.61

[a]DLPNO-CCSD(T)/CBS1: CBS extrapolation using aug-cc-pVTZ and aug-cc-pVQZ basis sets (α = 5.79, β = 3.05) with NormalPNO settings.
[b]DLPNO-CCSD(T)/CBS2: CBS extrapolation using def2-TZVPP and def2-QZVPP basis sets with TightPNO settings.

Table S4. Error statistics of Training Set calculations

Training Subset	Properties [a]	RMSE
Training set I (Monomer)	μ	0.20 Debye
	ΔE_{conf}	1.41 kcal·mol^{-1}
Training set II (Trimer)	V_{3B}	0.39 kcal·mol^{-1}
Training set III & IV (Dimer)	V_{2B}	1.64 kcal·mol^{-1}

[a]μ: Dipole moment; ΔE_{conf} : Conformational energy; V_{3B}: Three-body interaction energy; V_{2B}: Two-body interaction energy.

Table S5 Force field parameters utilized in the Polarizable Bond-dipole-based Force Field (PBFF)

Parameters for van der Waals interactions terms

atom	description	R^* (Å)	ε (kcal·mol^{-1})
OW	O in water	1.7683[b]	0.1520[b]
HW	H in water	0.0000[b]	0.0000[b]
P/PN	phosphorus in anionic or neutral phosphate species	2.1000[b]	0.2000[b]
O2/O2N	sp^2 oxygen in anionic or neutral phosphate species	1.6612[b]	0.2100[b]
OH/OHN	sp^3 oxygen attached to H in anionic or neutral phosphate species	1.7210[b]	0.2104[b]
OS/OSN	sp^3 oxygen attached to C in anionic or neutral phosphate species	1.6837[b]	0.1700[b]
HO/HON	H attached to O in anionic or neutral phosphate species	0.0000[b]	0.0000[b]
H1/H1N	H attached to alpha carbon in anionic or neutral phosphate species	1.3870[b]	0.0157[b]
HC/HCN	H attached to beta carbon in anionic or neutral phosphate species	1.4870[b]	0.0157[b]
CT/CTN	sp^3 carbon in anionic or neutral phosphate species	1.9080[b]	0.1094[b]

Parameters for electrostatic terms

bond	μ_0 (Debye)	c	atom	Q (e)
OW–HW	1.51[a]	1.74	P	-0.5640
P=O2	2.70[a]	2.00	O2	-0.1126
P–OS/OH	0.30[a]	–	OS	-0.1054
CT–OS	0.70[a]	–	OH	-0.1054
OH–HO	1.51[a]	0.80		
CT–H1/HC	0.30[a]	-		
PN=O2N	3.07	3.00		
PN–OSN/OHN	0.00	–		
CTN–OSN	1.11	–		
OHN–HON	1.70	2.40		
CTN–H1N/HCN	0.70	–		

Parameters for orbital term	D (kcal·mol^{-1})	a (Å^{-1})	R_0 (Å)
O(sp^3)···H–O	1.48	1.21	1.85
O(sp^2)···H–O	2.10	1.50	1.85

Parameters for bond strething

bond	b_{eq} (Å)	K_b (kcal·mol^{-1}·Å^{-2})	bond	b_{eq} (Å)	K_b (kcal·mol^{-1}·Å^{-2})
P=O2(N)*	1.4800[b]	525.0[b]	CT–OS(N)	1.4100[b]	320.0[b]
P–OH(N)	1.6100[b]	230.0[b]	CT–H1(N)	1.0900[b]	340.0[b]
P–OS(N)	1.6100[b]	230.0[b]	CT–CT(N)	1.5260[b]	310.0[b]
OH–HO(N)	0.9600[b]	553.0[b]	CT–CT(N)	1.0900[b]	340.0[b]
OW–HW	0.9572[b]	553.0[b]			

Table S5 (Continued)

Parameters for angle bending

angle	a_{eq} (º)	K_a (kcal·mol^{-1}·rad^{-2})	angle	a_{eq} (º)	K_a (kcal·mol^{-1}·rad^{-2})
O2N–PN–OHN	113.40	100.0	O2–P–OH	108.23[b]	90.0[b]
O2N–PN–OSN	113.40	100.0	O2–P–OS	108.23[b]	100.0[b]
PN–OHN–HON	113.28	45.0	P–OH–HO	108.50	90.0
P–OS–CT(N)	120.50[b]	100.0[b]	O2–P–O2	119.90[b]	140.00[b]
OH–P–OS(N)	102.60[b]	45.0[b]	CT–CT–HC(N)	109.50[b]	50.0[b]
OS–P–OS(N)	102.60[b]	45.0[b]	H1–CT–CT(N)	109.50[b]	50.0[b]
OS–CT–H1(N)	109.50[b]	50.0[b]	H1–CT–H1(N)	109.50[b]	35.0[b]
OS–CT–CT(N)	109.50[b]	50.0[b]	HC–CT–HC(N)	109.50[b]	35.0[b]
HW–OW–HW	104.25[b]	100.0[b]			

Parameters for torsion

torsion	*n*=1		*n*=2		*n*=3	
	$V_1/2$ (kcal·mol^{-1})	γ(º)	$V_2/2$(kcal·mol^{-1})	γ(º)	$V_3/2$(kcal·mol^{-1})	γ(º)
O2–P–OS–CT	0.00	0	0.00	0	0.700	0
O2–P–OH–HO	0.00	0	0.00	0	0.150	180
OS–P–OH–HO	0.00	0	1.20	0	0.250	180
OS–P–OS–CT	0.00[b]	0[b]	1.20[b]	0[b]	0.250[b]	0[b]
O2N–PN–OSN–CTN	0.00[b]	0[b]	0.00[b]	0[b]	0.250[b]	0[b]
O2N–PN–OHN–HON	0.00	0	0.00	0	0.400	0
OSN–PN–OHN–HON	0.00[b]	0[b]	0.00[b]	0[b]	0.250[b]	0[b]
OSN–PN–OSN–CTN	0.00	0	0.00	0	0.250	0
P–OS–CT–H1(N)	0.00	0	0.00	0	0.150	0
P–OS–CT–CT(N)	0.00[b]	0[b]	0.00[b]	0[b]	0.383[b]	0[b]
OH–P–OS–CT(N)	0.00[b]	0[b]	1.20[b]	0[b]	0.250[b]	0[b]
OS–CT–CT–HC(N)	0.00[b]	0[b]	0.25[b]	0[b]	0.156[b]	0[b]
H1–CT–CT–HC(N)	0.00[b]	0[b]	0.00[b]	0[b]	0.156[b]	0[b]

*(N) indicates applicability to neutral phosphate systems. For example, P=O2(N) collectively represents P=O2 and PN=O2N;
[a] values taken from literature;[7]
[b] values taken from AMBER99sb force field;[8]

Table S6 Summary of species and descriptions in Training and Testing Sets

database	subset	species[a]	data point	brief description[b]
Training Set I	Monomer	all	17	
	Monomer-Angle	DMPA	18	angle (O2–P–OS, P–OS–CT)×9
		TMP	18	angle (O2N–PN–OSN, PN–OSN–CTN)×9
		DMPH	33	angle (PN–OHN–HON, O2N–PN–OHN, O2N–PN–OSN)×11
	Monomer-Dihedral	DMPA	19	dihedral (P–OS–CT–H1)×19, (0–360°, 20° increments)
		TMP	49	dihedral (O2N–PN–OSN–CTN, PN–OSN–CTN–H1N)×19, (OSN–PN–OSN–CTN)×11
		DMPH	38	dihedral (O2N–PN–OSN–CTN, PN–OSN–CTN–H1N)×19
Training Set II	Trimer	all	24	
	Trimer-Interdistance×12	all	576	inter-distance×12 (0.9t, 1.0t, 1.1t, 1.25t, 1.5t, 3Å, 3.5Å, 4Å, 4.5Å, 5Å, 10Å, 15Å)
	Trimer-Interangle×9	all	297	inter-angle×8(0, +5, +10, +15, +20, +25, +30, +40, +50)
TrainingSet III & IV	Dimer	all	21	
	Dimer-Interdistance×14	all	294	inter-distance×14(0.85t, 0.9t, 0.95t, 1t, 1.05t , 1.1t, 1.15t, 1.2t, 1.4t, 1.6t, 1.8t, 2t, 3t, 4t)
	Dimer-Interangle×12	DMPA, TMP	84	inter-angle×12(-10, -5, 0, +5, +10, +15, +20, +30, +40, +50, +60, +70)
Testing Set I	Testing monomers	all	703	flexible scanning dihedral angle (0–360°, 10° increments)
		DMPA	148	dihedral (O2–P–OS–CT, P–OS–CT–H1, OS–P–OS–CT, CT–OS–OS–CT)×37
		DEPA	111	dihedral (P–OS–CT-CT, OS–CT–CT–HC, O2-P-OS-CT)×37
		MPA	111	dihedral (O2–P–OH–HO, O2–P–OS–CT, P–OS–CT–H1)×37
		TMP	111	dihedral (OSN–PN–OSN–CTN, O2N–PN–OSN–CTN, P–OSN–CTN–H1N)×37
		TEP	74	dihedral (OSN–PN–OSN–CTN, OSN–CTN–CTN–HCN)×37
		DMPH	148	dihedral (OHN–PN–OSN–CTN, OSN–PN–OSN–CTN, O2N–PN–OSN–CTN, PN–OSN–CTN–H1N)×37
Testing Set II	Testing clusters	all+water	46	DMPA-*n*W (*n* = 4-16, 18, 20, 22, 24), DEPA-*n*W (*n* = 10, 20), MPA-*n*W (*n* = 10, 16, 18, 20, 22), TMP-*n*W (n = 4, 6, 8, 10, 15, 20), TEP-*n*W (*n* =10, 20), DMPH-*n*W (*n* = 4-11, 13, 16, 18, 20, 22, 24)

[a] "all" denotes all types of phosphate compounds, including DMPA, DEPA, MPA, TMP, TEP, and DMPH.
[b] t is a multiple of the equilibrium distance. e.g. 0.9t = 0.9 x equilibrium distance; + or - = degrees added or subtracted from the equilibrium angle. +5 is five degrees above the equilibrium;

Table S7 Total interaction energies (IE_{tot}, in kcal·mol^{-1}) for DMPA-*n*W, $n \leq 10$

method / label	DLPNO-CCSD(T)/CBS	ωB97X-D/ aug-cc-pVQZ	ωB97X-D/ aug-cc-pVTZ	ωB97X-D/ aug-cc-pVDZ	B3LYP-D3/ aug-cc-pVQZ	B3LYP-D3/ aug-cc-pVTZ	B3LYP-D3/ aug-cc-pVDZ
DMPA-4W	-57.90	-59.17	-59.30	-58.72	-61.22	-61.01	-60.35
DMPA-5W	-70.63	-72.14	-72.29	-71.68	-74.81	-74.53	-73.79
DMPA-6W	-81.12	-83.10	-83.28	-82.62	-86.19	-85.87	-85.10
DMPA-7W	-92.56	-95.13	-95.34	-94.39	-98.64	-98.23	-97.02
DMPA-8W	-104.39	-106.91	-107.13	-106.34	-111.19	-110.79	-109.73
DMPA-9W	-121.72	-125.69	-125.97	-124.75	-130.29	-129.76	-128.26
DMPA-10W	-122.59	-125.77	-126.02	-125.15	-130.69	-130.22	-128.93

Table S8 Total two-body interaction energies (IE_{2B}, in kcal·mol^{-1}) for DMPA-*n*W, $n \leq 10$

label \ method	DLPNO-CCSD(T)/CBS	ωB97X-D/ aug-cc-pVQZ	ωB97X-D/ aug-cc-pVTZ	ωB97X-D/ aug-cc-pVDZ	B3LYP-D3/ aug-cc-pVQZ	B3LYP-D3/ aug-cc-pVTZ	B3LYP-D3/ aug-cc-pVDZ
DMPA-4W	-60.09	-62.04	-62.19	-61.51	-64.22	-64.02	-63.35
DMPA-5W	-72.67	-75.03	-75.21	-74.46	-77.82	-77.54	-76.82
DMPA-6W	-80.80	-83.56	-83.76	-82.99	-86.64	-86.34	-85.58
DMPA-7W	-91.33	-94.12	-94.47	-93.25	-98.05	-97.65	-96.47
DMPA-8W	-100.13	-102.89	-103.16	-102.18	-107.06	-106.67	-105.55
DMPA-9W	-118.68	-123.08	-123.55	-121.87	-128.02	-127.52	-126.06
DMPA-10W	-112.14	-115.27	-115.56	-114.53	-119.87	-118.69	-118.04

Table S9 Maximum absolute error (MAE, in kcal·mol^{-1}) of interaction energy and its decomposition, and the proportion of two-body contribution

	Methods	IE_{tot}	IE_{2B}	IE_{mB}	IE_{2B}/IE_{tot}
DMPA-*n*W	DLPNO-CCSD(T)/CBS	0.00	0.00	0.00	92.67%
	PBFF	3.34	2.23	1.87	92.66%
	AMOEBA	3.52	6.24	5.11	94.81%
	ωB97X-D/aug-cc-pVTZ	10.34	8.21	2.24	92.86%
	B3LYP-D3/aug-cc-pVTZ	22.13	18.52	3.61	92.78%
DEPA-*n*W	DLPNO-CCSD(T)/CBS	0.00	0.00	0.00	94.33%
	PBFF	0.57	0.30	0.87	93.94%
	AMOEBA	–	–	–	–
	ωB97X-D/aug-cc-pVTZ	8.70	7.34	1.36	94.39%
	B3LYP-D3/aug-cc-pVTZ	18.25	15.48	2.77	94.30%
MPA-*n*W	DLPNO-CCSD(T)/CBS	0.00	0.00	0.00	91.82%
	PBFF	4.13	1.71	2.42	91.74%
	AMOEBA	–	–	–	–
	ωB97X-D/aug-cc-pVTZ	11.67	10.00	1.79	91.71%
	B3LYP-D3/aug-cc-pVTZ	22.88	19.88	3.00	91.77%
TMP-*n*W	DLPNO-CCSD(T)/CBS	0.00	0.00	0.00	83.38%
	PBFF	3.70	4.19	3.11	82.55%
	AMOEBA	9.14	8.19	2.01	83.75%
	ωB97X-D/aug-cc-pVTZ	7.02	4.93	2.09	83.05%
	B3LYP-D3/aug-cc-pVTZ	17.44	12.56	4.88	83.24%
TEP-*n*W	DLPNO-CCSD(T)/CBS	0.00	0.00	0.00	82.67%
	PBFF	2.19	2.65	2.50	81.65%
	AMOEBA	–	–	–	–
	ωB97X-D/aug-cc-pVTZ	7.54	5.52	2.02	82.45%
	B3LYP-D3/aug-cc-pVTZ	17.91	13.04	4.87	82.32%
DMPH-*n*W	DLPNO-CCSD(T)/CBS	0.00	0.00	0.00	82.37%
	PBFF	6.18	4.24	5.00	82.85%
	AMOEBA	–	–	–	–
	ωB97X-D/aug-cc-pVTZ	9.55	5.83	3.74	81.77%
	B3LYP-D3/aug-cc-pVTZ	22.66	16.60	6.06	81.91%

Table S10 Total interaction energies (IE_{tot}, in kcal·mol^{-1}) for phosphate-water clusters in Testing Set II

label \ method	DLPNO-CCSD(T)/CBS	ωB97X-D /aug-cc-pVTZ	ωB97X-D /aug-cc-pVDZ	B3LYP-D3 /aug-cc-pVTZ	B3LYP-D3 /aug-cc-pVDZ	PBFF	AMOEBA	AMBER99SB	CHARMM36
DMPA-4W	-57.90	-59.30	-58.72	-61.01	-60.35	-58.74	-59.31	-59.48	-55.06
DMPA-5W	-70.63	-72.29	-71.68	-74.53	-73.79	-72.74	-71.76	-72.88	-67.68
DMPA-6W	-81.12	-83.28	-82.62	-85.87	-85.10	-81.97	-81.48	-81.64	-76.25
DMPA-7W	-92.56	-95.34	-94.39	-98.23	-97.02	-92.63	-94.11	-93.73	-89.88
DMPA-8W	-104.39	-107.13	-106.34	-110.79	-109.73	-104.88	-104.56	-104.39	-98.40
DMPA-9W	-121.72	-125.97	-124.75	-129.76	-128.26	-121.40	-123.28	-121.48	-116.24
DMPA-10W	-122.59	-126.02	-125.15	-130.22	-128.93	-123.22	-122.07	-122.50	-116.16
DMPA-11W	-140.13	-144.92	-143.58	-149.80	-147.96	-138.09	-138.95	-139.14	-134.58
DMPA-12W	-150.80	-155.63	-154.35	-161.10	-159.28	-147.86	-148.48	-149.65	-144.31
DMPA-13W	-166.58	-172.35	-171.02	-178.50	-176.60	-164.51	-164.06	-164.36	-159.33
DMPA-14W	-176.22	-182.39	-180.90	-188.88	-186.72	-174.76	-173.35	-173.81	-169.12
DMPA-15W	-186.16	-192.92	-191.14	-199.94	-197.62	-183.99	-182.64	-183.58	-179.07
DMPA-16W	-204.97	-213.81	-212.11	-221.51	-219.35	-204.62	-204.00	-201.13	-196.58
DMPA-18W	-226.34	-235.78	-234.01	-245.14	-242.69	-225.72	-224.99	-223.54	-219.91
DMPA-20W	-233.85	-242.88	-241.04	-252.41	-249.73	-233.10	-230.89	-230.36	-226.97
DMPA-22W	-269.81	-279.73	-277.56	-290.74	-287.65	-272.18	-267.33	-269.45	-262.47
DMPA-24W	-287.84	-298.18	-295.93	-309.97	-306.65	-291.18	-284.63	-286.14	-279.76
DEPA-10W	-130.00	-134.71	-133.52	-139.42	-138.07	-130.57	–	–	–
DEPA-20W	-239.13	-247.83	-245.97	-257.38	-254.69	-239.20	–	–	–
MPA-10W	-131.48	-135.01	-133.87	-139.64	-138.13	-132.71	–	–	–
MPA-16W	-205.31	-212.74	-211.20	-220.31	-218.26	-206.66	–	–	–
MPA-18W	-230.54	-239.18	-237.35	-247.80	-245.19	-229.74	–	–	–
MPA-20W	-250.71	-260.86	-258.87	-270.92	-268.34	-250.51	–	–	–
MPA-22W	-283.21	-294.88	-292.37	-306.09	-302.87	-279.07	–	–	–

TableS10 (Continued)

label \ method	DLPNO-CCSD(T)/CBS	ωB97X-D /aug-cc-pVTZ	ωB97X-D /aug-cc-pVDZ	B3LYP-D3 /aug-cc-pVTZ	B3LYP-D3 /aug-cc-pVDZ	PBFF	AMOEBA	AMBER99SB	CHARMM36
TMP-4W	-38.23	-38.97	-38.62	-41.06	-40.52	-34.53	-36.64	–	–
TMP-6W	-54.42	-56.29	-55.63	-59.65	-58.8	-51.43	-52.7	–	–
TMP-8W	-78.77	-81.76	-80.89	-85.94	-84.74	-77.14	-76.04	–	–
TMP-10W	-101.00	-105.00	-103.82	-110.19	-108.68	-98.86	-97.2	–	–
TMP-15W	-142.90	-147.66	-146.45	-155.21	-153.25	-141.21	-135.03	–	–
TMP-20W	-201.35	-208.37	-206.92	-218.79	-216.13	-202.33	-192.21	–	–
TEP-10W	-102.28	-106.58	-105.43	-111.71	-110.24	-100.10	–	–	–
TEP-20W	-202.45	-209.99	-208.56	-220.36	-217.72	-202.56	–	–	–
DMPH-4W	-40.03	-40.31	-39.85	-42.26	-41.58	-38.33	–	–	–
DMPH-5W	-43.66	-44.44	-44.14	-46.88	-46.23	-41.87	–	–	–
DMPH-6W	-58.74	-59.49	-58.87	-62.72	-61.76	-54.25	–	–	–
DMPH-7W	-66.27	-68.10	-67.50	-71.85	-70.81	-63.15	–	–	–
DMPH-8W	-80.14	-81.76	-81.03	-85.65	-84.47	-76.40	–	–	–
DMPH-9W	-81.41	-83.87	-83.18	-88.62	-87.50	-80.91	–	–	–
DMPH-10W	-93.91	-96.57	-95.66	-101.39	-100.13	-91.53	–	–	–
DMPH-11W	-126.58	-130.65	-129.27	-135.86	-133.89	-120.40	–	–	–
DMPH-13W	-127.00	-130.99	-130.04	-137.65	-135.88	-123.20	–	–	–
DMPH-16W	-174.12	-179.91	-178.24	-187.64	-185.11	-169.40	–	–	–
DMPH-18W	-193.16	-199.07	-197.34	-207.76	-204.96	-189.42	–	–	–
DMPH-20W	-215.77	-224.53	-222.60	-235.37	-232.57	-220.59	–	–	–
DMPH-22W	-238.92	-248.06	-246.08	-260.18	-257.14	-240.89	–	–	–
DMPH-24W	-259.93	-269.48	-267.27	-282.59	-279.2	-261.04	–	–	–

Table S11 Total two-body interaction energies (IE_{2B}, in kcal·mol^{-1}) for phosphate-water clusters in Testing Set II

label \ method	DLPNO-CCSD(T)/CBS	ωB97X-D/ aug-cc-pVTZ	ωB97X-D/ aug-cc-pVDZ	B3LYP-D3/ aug-cc-pVTZ	B3LYP-D3/ aug-cc-pVDZ	PBFF	AMOEBA	MP2/aug-cc-pVQZ
DMPA-4W	-60.09	-62.19	-61.51	-64.02	-63.35	-60.34	-64.18	-60.88
DMPA-5W	-72.67	-75.21	-74.46	-77.54	-76.82	-74.07	-76.88	-73.56
DMPA-6W	-80.80	-83.76	-82.99	-86.34	-85.58	-81.80	-83.85	-81.95
DMPA-7W	-91.33	-94.47	-93.25	-97.65	-96.47	-90.46	-95.56	-91.80
DMPA-8W	-100.13	-103.16	-102.18	-106.67	-105.55	-100.97	-101.99	-101.26
DMPA-9W	-118.68	-123.55	-121.87	-127.52	-126.06	-117.56	-124.92	-119.56
DMPA-10W	-112.14	-115.56	-114.53	-118.69	-118.04	-114.38	-112.97	-113.51
DMPA-11W	-127.20	-131.67	-130.19	-134.94	-133.99	-125.08	-128.76	-128.37
DMPA-12W	-136.29	-141.03	-139.60	-144.70	-143.85	-134.22	-137.23	-138.01
DMPA-13W	-146.85	-151.89	-150.54	-156.78	-154.81	-146.28	-146.50	-148.81
DMPA-14W	-154.11	-159.47	-157.91	-164.65	-162.44	-154.12	-154.00	-156.12
DMPA-15W	-164.14	-169.83	-167.83	-176.05	-173.66	-162.92	-163.99	-165.88
DMPA-16W	-187.56	-194.78	-192.56	-202.45	-200.12	-187.00	-190.42	-189.19
DMPA-18W	-195.81	-203.07	-200.75	-211.76	-209.07	-196.03	-196.95	-198.14
DMPA-20W	-204.58	-211.37	-209.02	-220.27	-217.28	-203.87	-203.79	-206.64
DMPA-22W	-235.88	-243.99	-241.12	-253.55	-250.04	-237.91	-238.32	-238.53
DMPA-24W	-252.32	-260.53	-257.56	-270.84	-267.09	-253.80	-254.22	-254.96
DEPA-10W	-129.70	-134.79	-133.06	-139.43	-138.00	-129.40	–	-130.38
DEPA-20W	-212.56	-219.90	-217.43	-228.04	-225.03	-212.36	–	-214.06
MPA-10W	-129.83	-133.45	-131.91	-138.16	-136.55	-128.94	–	-130.47
MPA-16W	-188.61	-195.15	-193.01	-202.40	-200.18	-190.07	–	-190.41
MPA-18W	-206.01	-213.56	-211.32	-220.58	-217.65	-207.04	–	-208.52
MPA-20W	-227.42	-235.78	-232.91	-245.81	-242.95	-226.26	–	-229.36
MPA-22W	-250.45	-260.45	-257.19	-270.33	-266.79	-248.75	–	-254.25

Table S11 (Continued)

label \ method	DLPNO-CCSD(T)/CBS	ωB97X-D/ aug-cc-pVTZ	ωB97X-D/ aug-cc-pVDZ	B3LYP-D3/ aug-cc-pVTZ	B3LYP-D3/ aug-cc-pVDZ	PBFF	AMOEBA	MP2/aug-cc-pVQZ
TMP-4W	-31.51	-31.93	-31.55	-33.85	-33.31	-28.71	-30.03	-31.51
TMP-6W	-49.08	-50.73	-50.00	-53.99	-53.24	-44.90	-48.28	-49.45
TMP-8W	-66.31	-68.63	-67.72	-72.30	-71.22	-63.90	-64.74	-66.91
TMP-10W	-84.31	-87.46	-86.13	-92.01	-90.70	-81.42	-82.52	-85.18
TMP-15W	-112.19	-114.93	-113.68	-121.24	-119.25	-112.52	-105.00	-113.90
TMP-20W	-164.15	-169.08	-167.63	-176.71	-173.95	-162.03	-155.96	-166.45
TEP-10W	-85.67	-89.12	-87.79	-93.59	-92.29	-83.03	–	–
TEP-20W	-165.16	-170.68	-169.22	-178.20	-175.44	-162.79	–	–
DMPH-4W	-37.17	-37.17	-36.73	-38.90	-38.23	-35.21	–	-36.87
DMPH-5W	-34.70	-35.09	-34.87	-37.07	-36.43	-35.20	–	-34.99
DMPH-6W	-53.34	-53.79	-53.12	-56.72	-55.76	-49.10	–	-53.30
DMPH-7W	-56.23	-57.48	-56.87	-60.80	-59.80	-53.25	–	-56.70
DMPH-8W	-66.43	-67.26	-66.50	-70.54	-69.37	-64.51	–	-66.74
DMPH-9W	-69.03	-70.83	-70.11	-75.03	-73.91	-68.77	–	-69.67
DMPH-10W	-80.05	-81.45	-80.51	-85.78	-84.57	-78.86	–	-80.65
DMPH-11W	-95.57	-97.87	-96.50	-101.86	-99.81	-94.39	–	-96.66
DMPH-13W	-101.71	-104.50	-103.52	-109.63	-107.81	-99.60	–	-102.94
DMPH-16W	-136.64	-139.82	-137.96	-146.53	-143.81	-133.84	–	-138.10
DMPH-18W	-152.81	-155.83	-153.83	-163.56	-160.55	-148.86	–	-154.19
DMPH-20W	-171.70	-177.10	-174.70	-185.96	-183.05	-173.43	–	-173.36
DMPH-22W	-190.68	-196.51	-194.24	-206.41	-203.24	-191.05	–	-193.43
DMPH-24W	-207.06	-212.87	-210.34	-223.66	-220.10	-205.80	–	-209.71

Table S12 Total many-body interaction energies (IE_{mB}, in kcal·mol^{-1}) for phosphate-water clusters in Testing Set II

label \ method	DLPNO-CCSD(T)/CBS	ωB97X-D/ aug-cc-pVTZ	ωB97X-D/ aug-cc-pVDZ	B3LYP-D3/ aug-cc-pVTZ	B3LYP-D3/ aug-cc-pVDZ	PBFF	AMOEBA
DMPA-4W	2.19	2.89	2.79	3.01	3.00	1.60	4.87
DMPA-5W	2.04	2.92	2.78	3.01	3.03	1.33	5.12
DMPA-6W	-0.32	0.48	0.37	0.47	0.48	-0.17	2.37
DMPA-7W	-1.23	-0.87	-1.14	-0.58	-0.55	-2.17	1.45
DMPA-8W	-4.26	-3.97	-4.16	-4.12	-4.18	-3.91	-2.57
DMPA-9W	-3.04	-2.42	-2.88	-2.24	-2.20	-3.84	1.64
DMPA-10W	-10.45	-10.46	-10.62	-11.53	-10.89	-8.84	-9.10
DMPA-11W	-12.93	-13.25	-13.39	-14.86	-13.97	-13.01	-10.19
DMPA-12W	-14.51	-14.60	-14.75	-16.40	-15.43	-13.64	-11.25
DMPA-13W	-19.73	-20.46	-20.48	-21.72	-21.79	-18.23	-17.56
DMPA-14W	-22.11	-22.92	-22.99	-24.23	-24.28	-20.64	-19.35
DMPA-15W	-22.02	-23.09	-23.31	-23.89	-23.96	-21.07	-18.65
DMPA-16W	-17.41	-19.03	-19.55	-19.06	-19.23	-17.62	-13.58
DMPA-18W	-30.53	-32.71	-33.26	-33.38	-33.62	-29.69	-28.04
DMPA-20W	-29.27	-31.51	-32.02	-32.14	-32.45	-29.23	-27.10
DMPA-22W	-33.93	-35.74	-36.44	-37.19	-37.61	-34.27	-29.01
DMPA-24W	-35.52	-37.65	-38.37	-39.13	-39.56	-37.38	-30.41
DEPA-10W	-0.30	0.08	-11.96	0.01	-0.07	-1.17	–
DEPA-20W	-26.57	-27.93	-28.54	-29.34	-29.66	-26.84	–
MPA-10W	-1.65	-1.56	-1.96	-1.48	-1.58	-3.77	–
MPA-16W	-16.70	-17.59	-18.19	-17.91	-18.08	-16.59	–
MPA-18W	-24.53	-25.62	-26.03	-27.22	-27.54	-22.70	–
MPA-20W	-23.29	-25.08	-25.96	-25.11	-25.39	-24.25	–
MPA-22W	-32.76	-34.43	-35.18	-35.76	-36.08	-30.32	–
TMP-4W	-6.72	-7.04	-7.07	-7.21	-7.21	-5.82	-6.61
TMP-6W	-5.34	-5.56	-5.63	-5.66	-5.56	-6.53	-4.42
TMP-8W	-12.46	-13.13	-13.17	-13.64	-13.52	-13.24	-11.30
TMP-10W	-16.69	-17.54	-17.69	-18.18	-17.98	-17.44	-14.68
TMP-15W	-30.71	-32.73	-32.77	-33.97	-34.00	-28.69	-30.03
TMP-20W	-37.20	-39.29	-39.29	-42.08	-42.18	-40.30	-36.25
TEP-10W	-16.61	-17.46	-17.64	-18.12	-17.95	-17.07	–
TEP-20W	-37.29	-39.31	-39.34	-42.16	-42.28	-39.77	–
DMPH-4W	-2.86	-3.14	-3.12	-3.36	-3.35	-3.12	–
DMPH-5W	-8.96	-9.35	-9.27	-9.81	-9.80	-6.67	–
DMPH-6W	-5.40	-5.70	-5.75	-6.00	-6.00	-5.15	–
DMPH-7W	-10.04	-10.62	-10.63	-11.05	-11.01	-9.90	–
DMPH-8W	-13.71	-14.50	-14.53	-15.11	-15.10	-11.89	–
DMPH-9W	-12.38	-13.04	-13.07	-13.59	-13.59	-12.14	–
DMPH-10W	-13.86	-15.12	-15.15	-15.61	-15.56	-12.67	–

Table S12 (Continued)

method / label	DLPNO-CCSD(T)/CBS	ωB97X-D/ aug-cc-pVTZ	ωB97X-D/ aug-cc-pVDZ	B3LYP-D3/ aug-cc-pVTZ	B3LYP-D3/ aug-cc-pVDZ	PBFF	AMOEBA
DMPH-11W	-31.01	-32.78	-32.77	-34.00	-34.08	-26.01	–
DMPH-13W	-25.29	-26.49	-26.52	-28.02	-28.07	-23.60	–
DMPH-16W	-37.48	-40.09	-40.28	-41.11	-41.30	-35.56	–
DMPH-18W	-40.35	-43.24	-43.51	-44.20	-44.41	-40.56	–
DMPH-20W	-44.07	-47.43	-47.90	-49.41	-49.52	-47.16	–
DMPH-22W	-48.24	-51.55	-51.84	-53.77	-53.90	-49.84	–
DMPH-24W	-52.87	-56.61	-56.93	-58.93	-59.10	-55.24	–

Table S13 Comparison of single-point energy calculation cost between PBFF and AMOEBA*

	Number of atoms	PBFF				AMOEBA			
		Number of electrotatic terms	CPU Time (ms)			Number of electrotatic terms	CPU Time (ms)		
			Initialization	Energy	**Total**		Initialization	Energy	**Total**
DMPA-4W	25	175	60	0.254	**60.3**	256	219	0.238	**219.2**
DMPA-5W	28	225	63	0.311	**63.3**	331	219	0.291	**219.3**
DMPA-6W	31	279	65	0.370	**65.4**	415	219	0.385	**219.4**
DMPA-7W	34	337	67	0.436	**67.4**	508	219	0.440	**219.4**
DMPA-8W	37	399	72	0.502	**72.5**	610	219	0.523	**219.5**
DMPA-9W	40	465	78	0.579	**78.6**	721	219	0.606	**219.6**
DMPA-10W	43	535	81	0.657	**81.7**	841	219	0.668	**219.7**
DMPA-11W	46	609	83	0.738	**83.7**	970	219	0.760	**219.8**
DMPA-12W	49	687	87	0.828	**87.8**	1108	219	0.866	**219.9**
DMPA-13W	52	769	88	0.921	**88.9**	1255	219	0.964	**220.0**
DMPA-14W	55	855	92	1.025	**93.0**	1411	219	1.049	**220.0**
DMPA-15W	58	945	93	1.127	**94.1**	1576	219	1.164	**220.2**
DMPA-16W	61	1039	102	1.233	**103.2**	1750	220	1.323	**221.3**
DMPA-18W	67	1239	108	1.471	**109.5**	2125	220	1.563	**221.6**
DMPA-20W	73	1455	122	1.673	**123.7**	2536	220	1.741	**221.7**
DMPA-22W	79	1687	141	1.736	**142.7**	2983	220	2.024	**222.0**
DMPA-24W	85	1935	155	1.903	**156.9**	3466	220	2.394	**222.4**
TMP-4W	29	162	59	0.215	**59.2**	351	219	0.314	**219.3**
TMP-6W	35	246	65	0.302	**65.3**	534	219	0.486	**219.5**
TMP-8W	41	346	67	0.403	**67.4**	753	219	0.635	**219.6**
TMP-10W	47	462	74	0.521	**74.5**	1008	219	0.818	**219.8**
TMP-15W	62	822	100	0.900	**100.9**	1803	219	1.279	**220.3**
TMP-20W	77	1282	133	1.392	**134.4**	2823	219	1.895	**220.9**

*Total time for a single-point energy calculation is defined as the sum of Initialization and Energy evaluation time.
PBFF initialization includes input setup, parameter loading, connectivity determination, and AM1 charge generation.
PBFF energy evaluation includes the computational cost of bonded and nonbonded energy terms.
AMOEBA initialization includes input setup, parameter loading, and related setup routines in TINKER.
AMOEBA energy evaluation includes the computational cost of bonded and nonbonded energy terms.

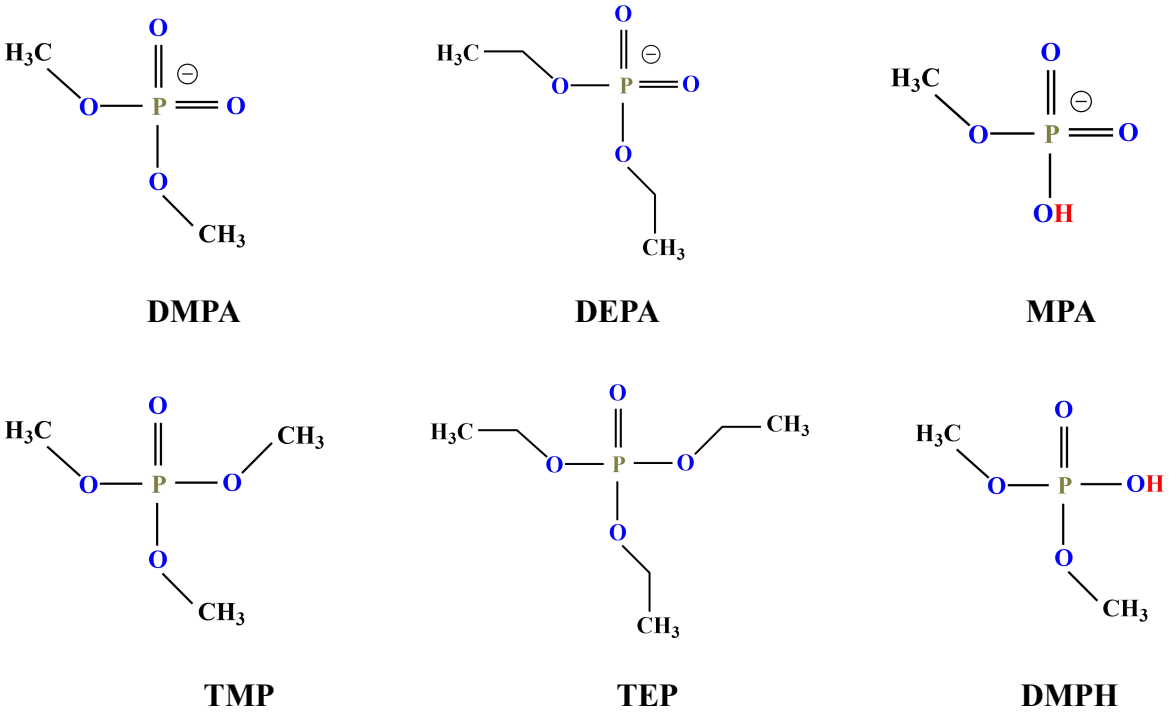


Fig. S3 Molecular structures of phosphate compounds.

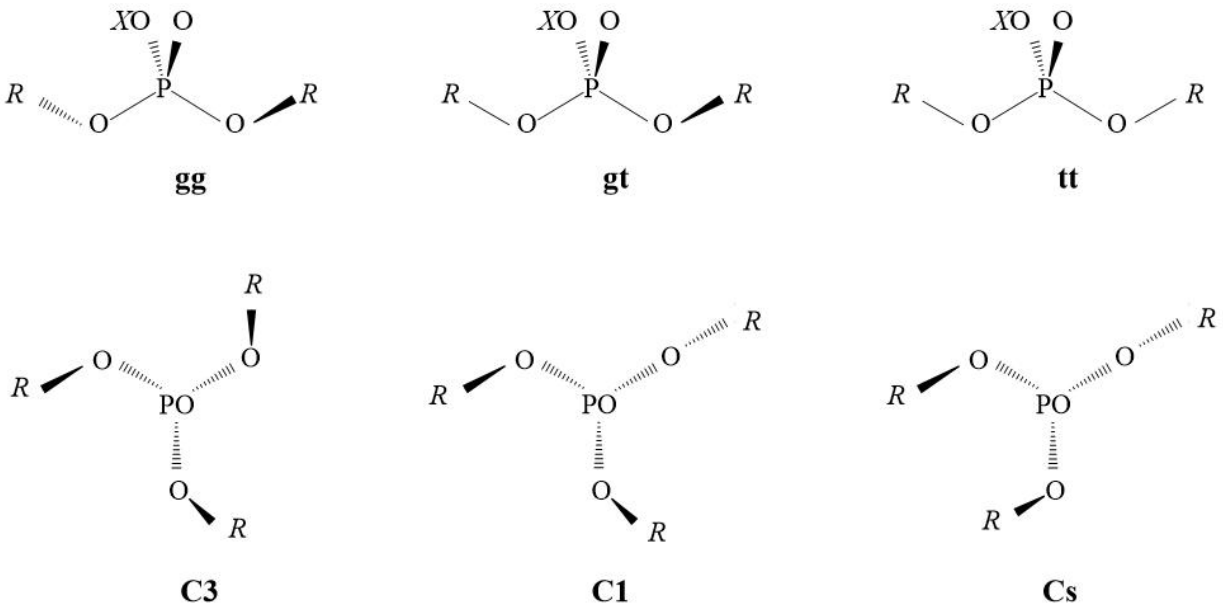


Fig. S4 Representative conformational illustrations of molecules in Training Set I.

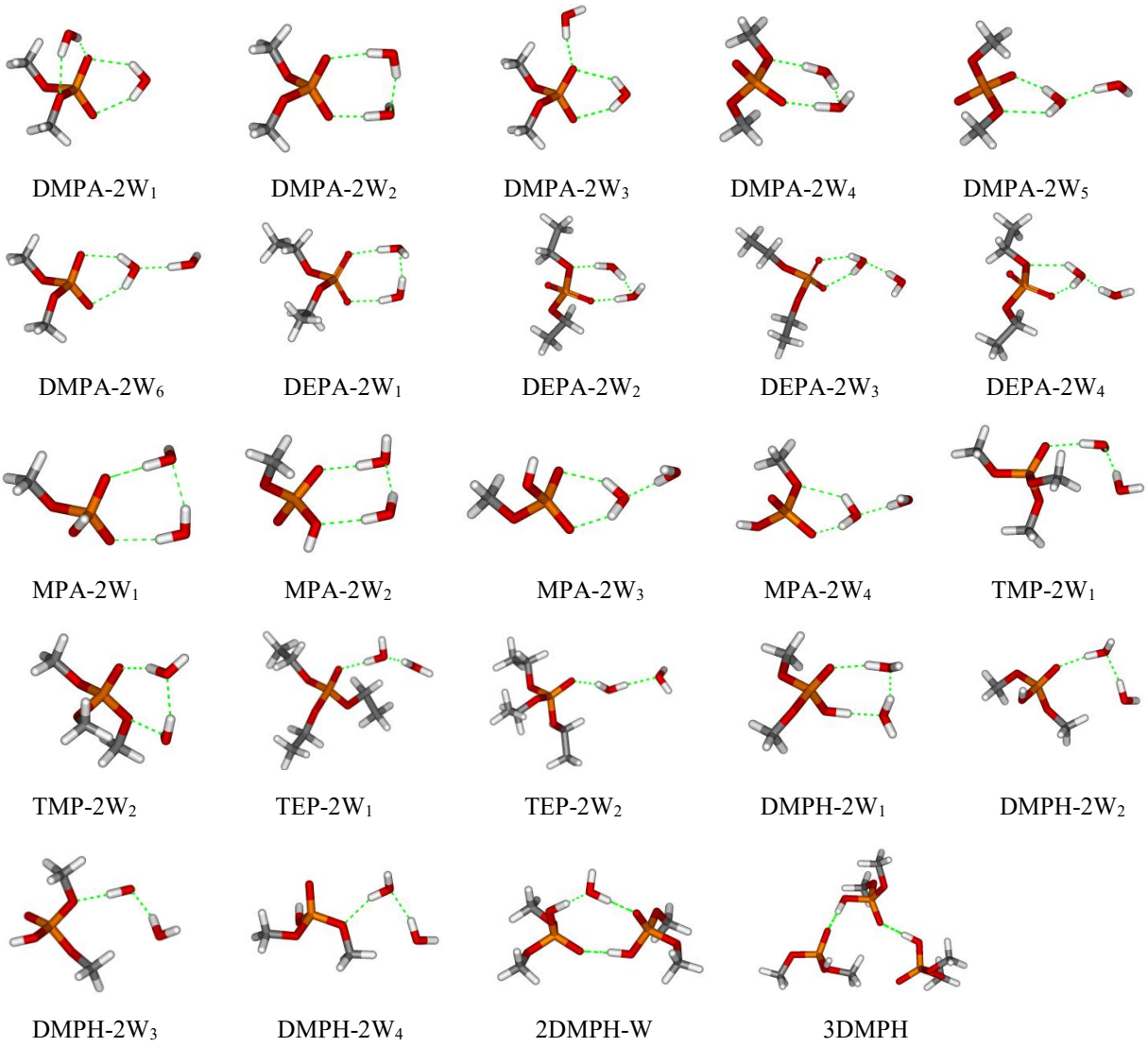


Fig. S5 Representative conformations in Training Set II.

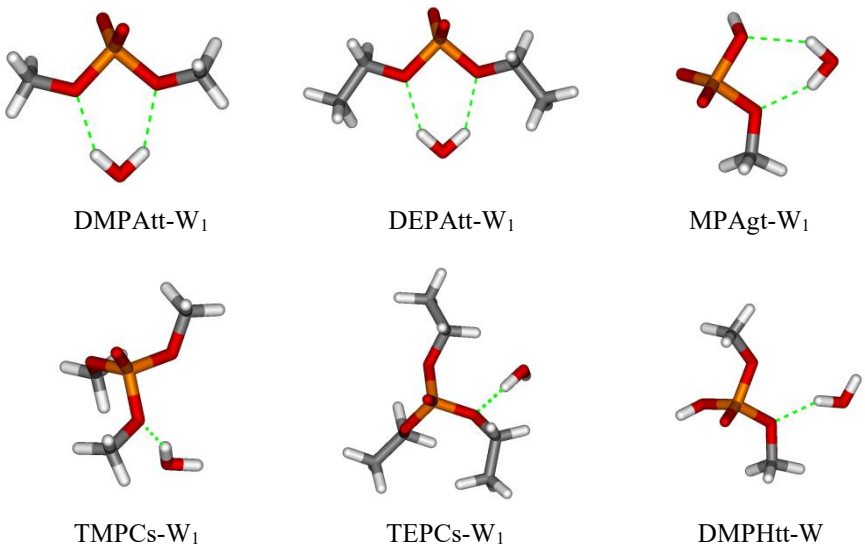


Fig. S6 Representative conformations in Training Set III.

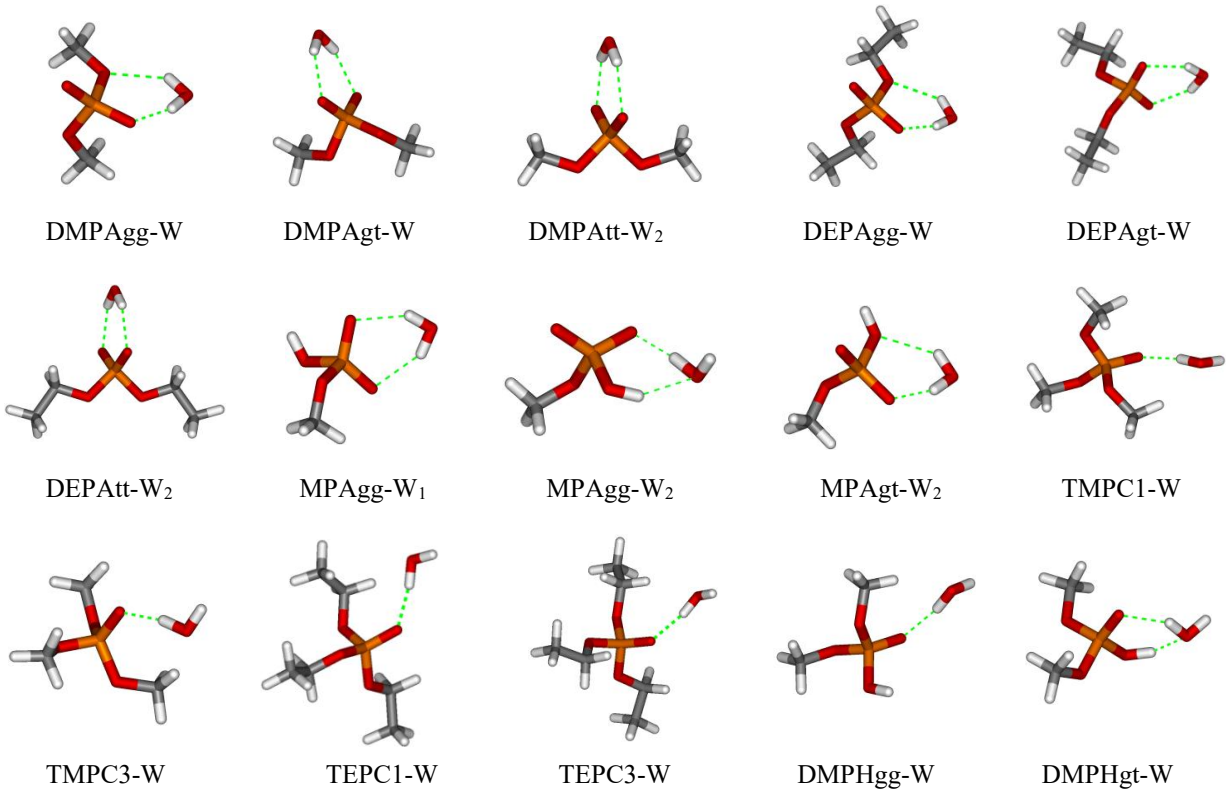


Fig. S7 Representative conformations in Training Set IV.

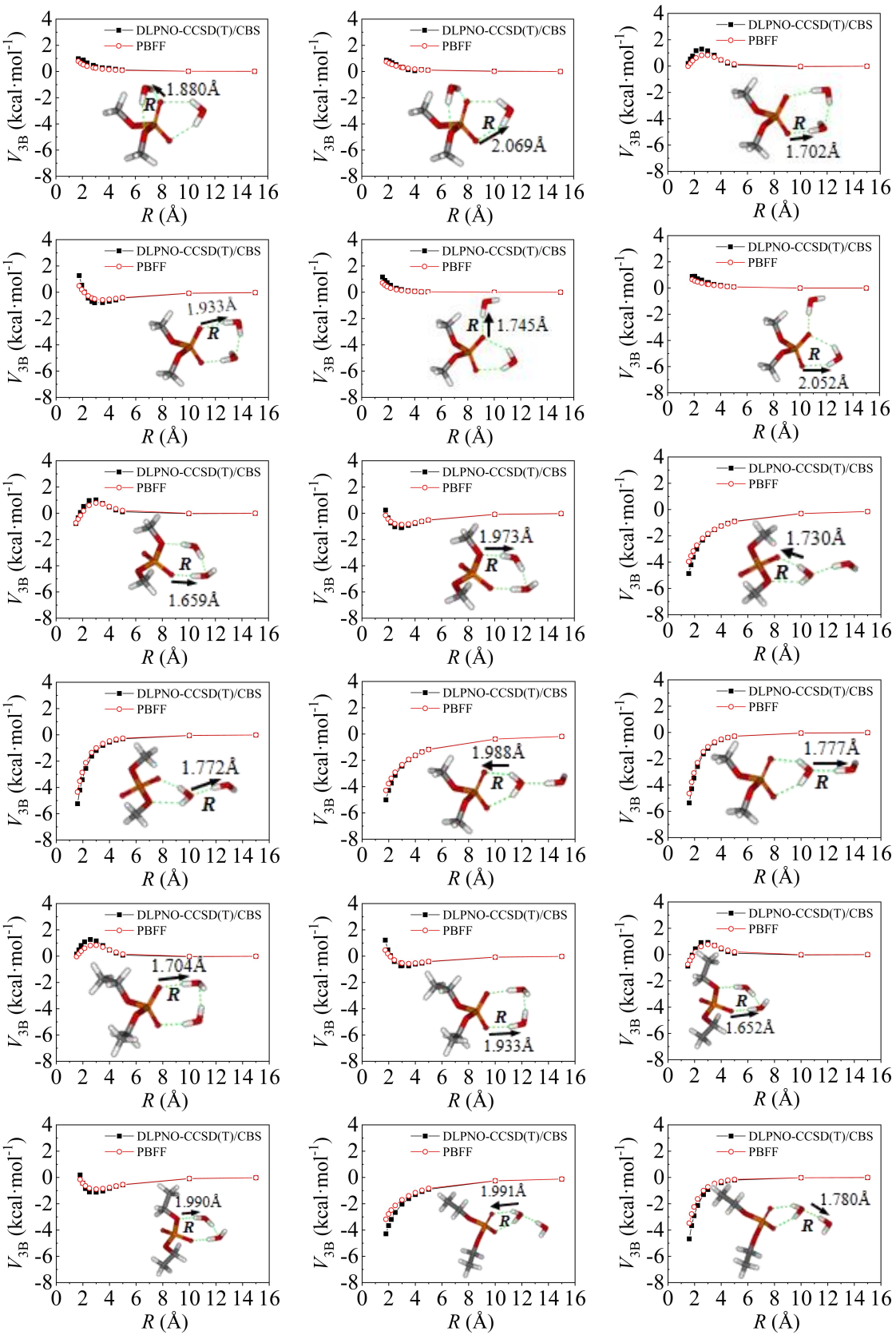


Fig. S8 Distance-dependent three-body interaction energies (V_{3B}) for Training Set II.

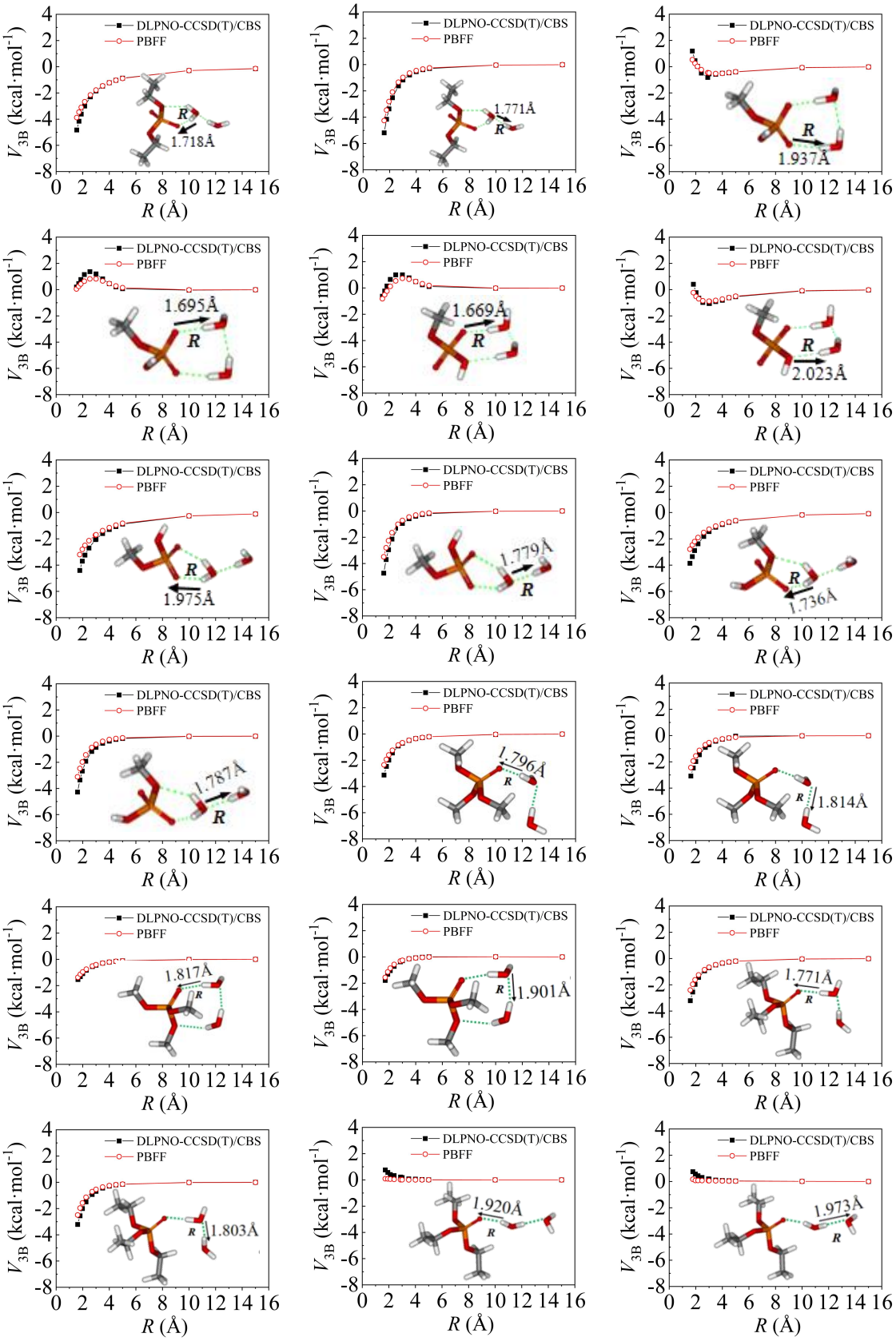


Fig. S8 (Continued)

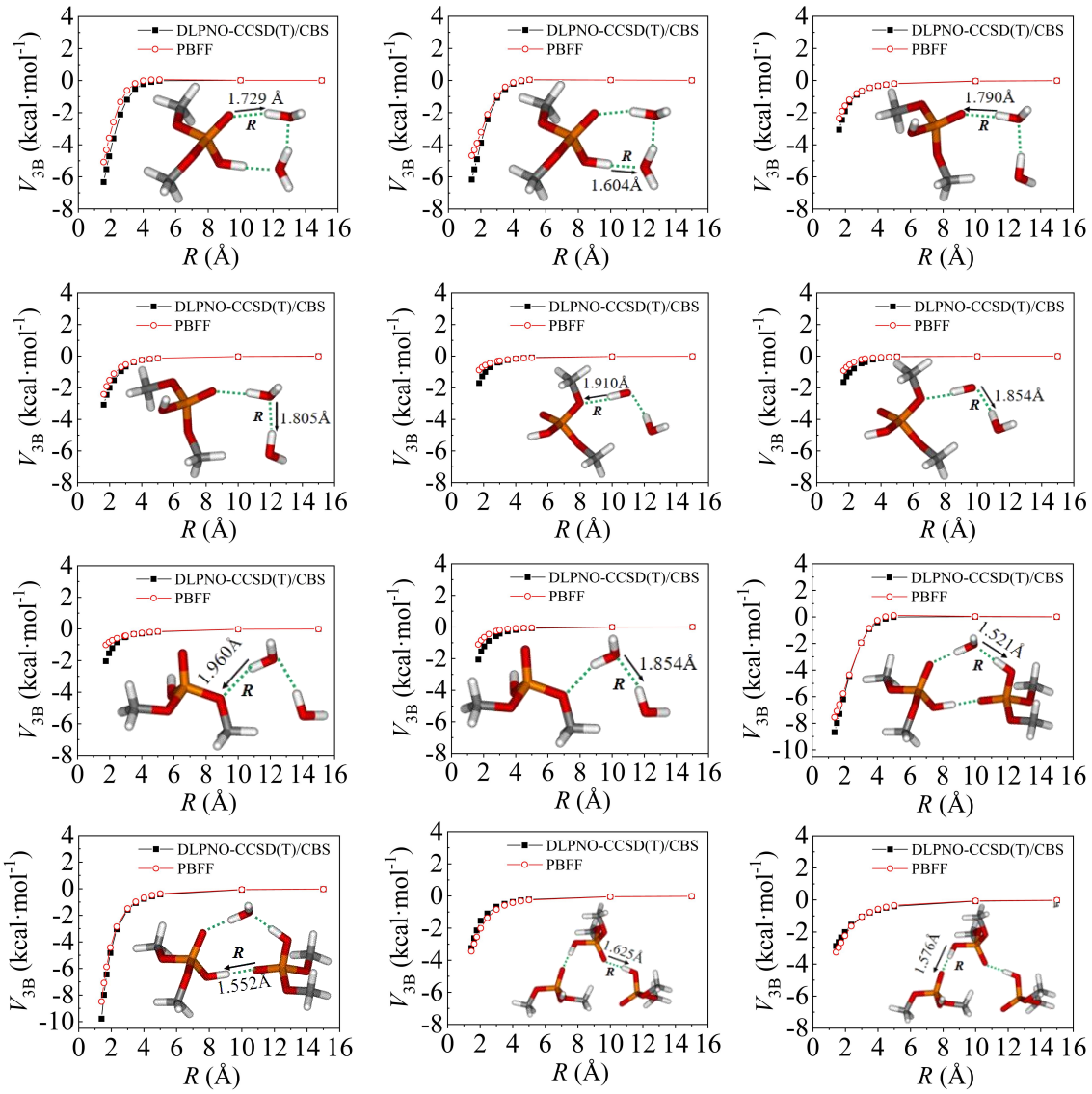


Fig. S8 (Continued)

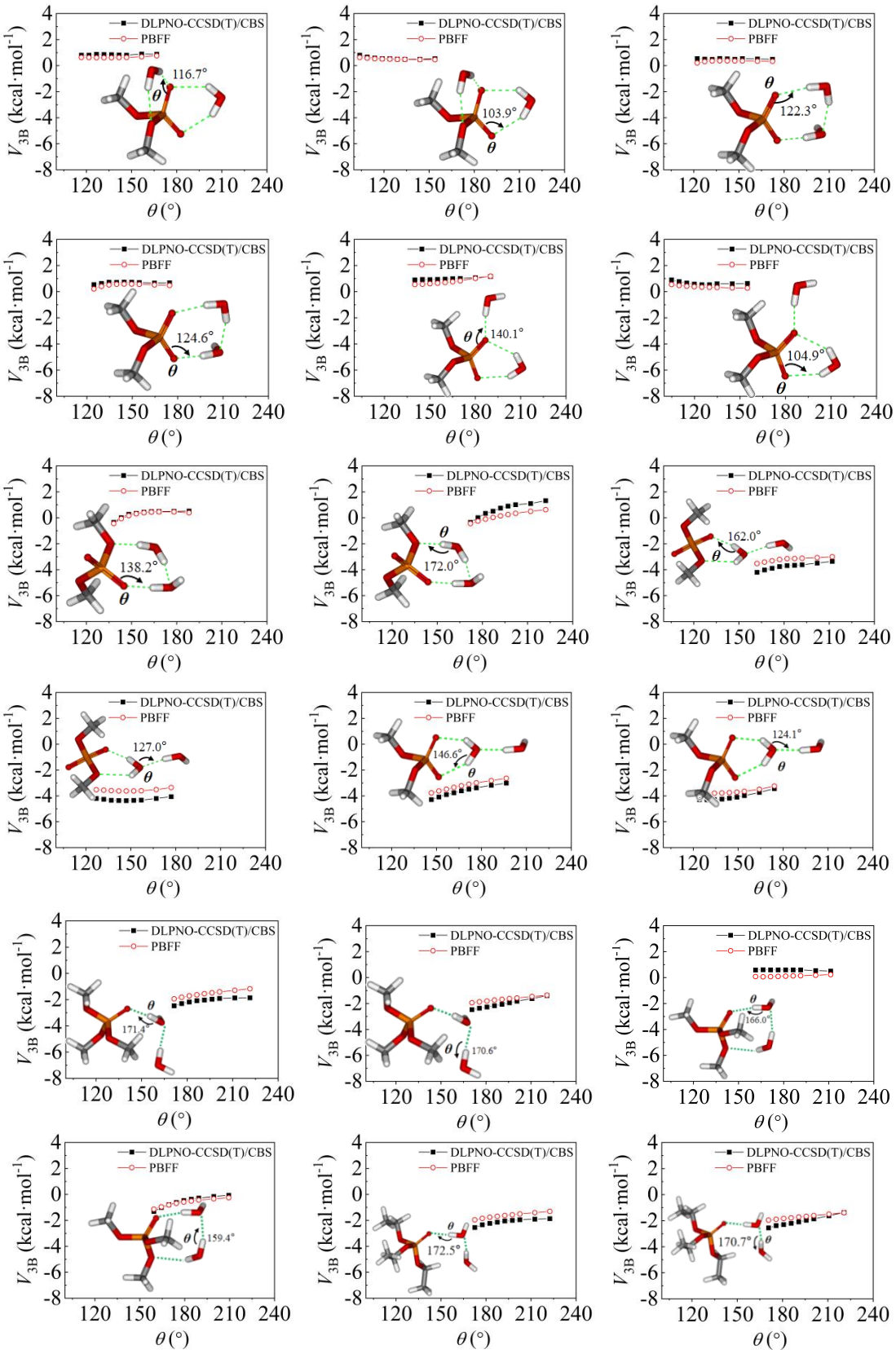


Fig. S9 Intermolecular angle-dependent three-body interaction energies (V_{3B}) for Training Set II.

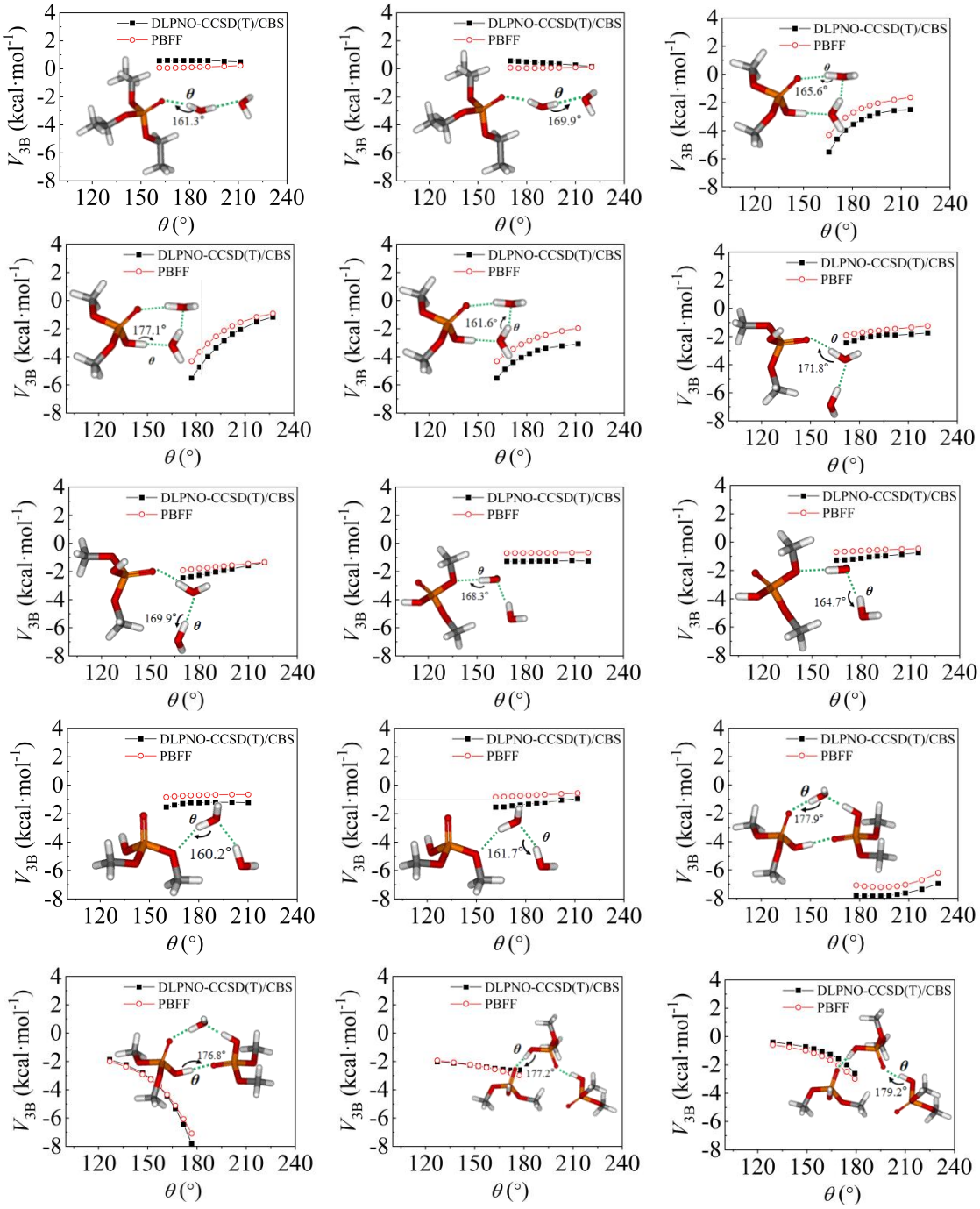


Fig. S9 (Continued)

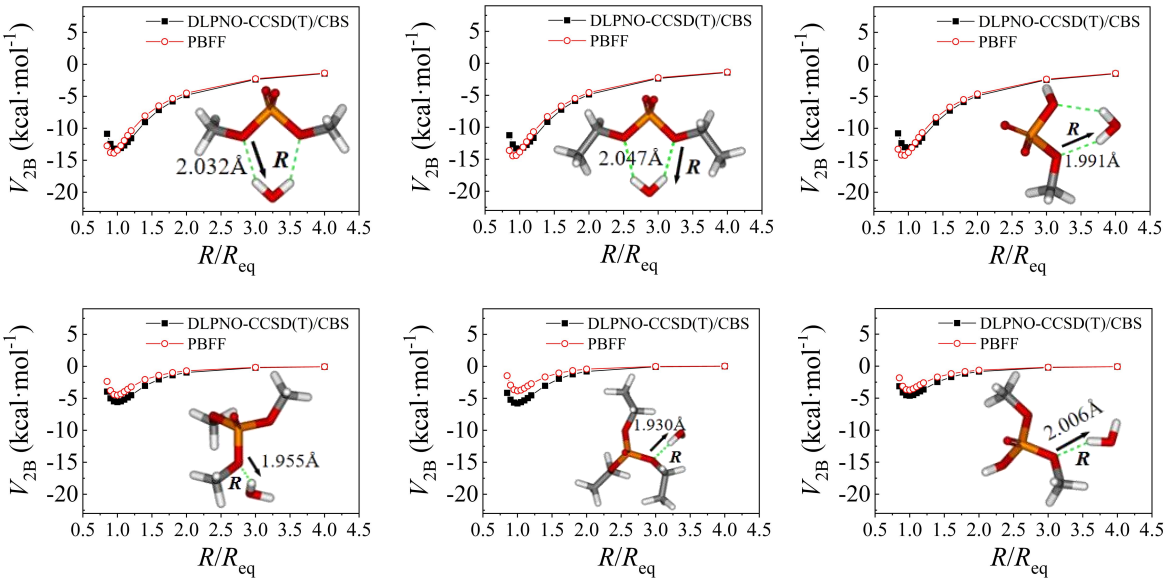


Fig. S10 Distance-dependent two-body interaction energies (V_{2B}) for Training Set III.

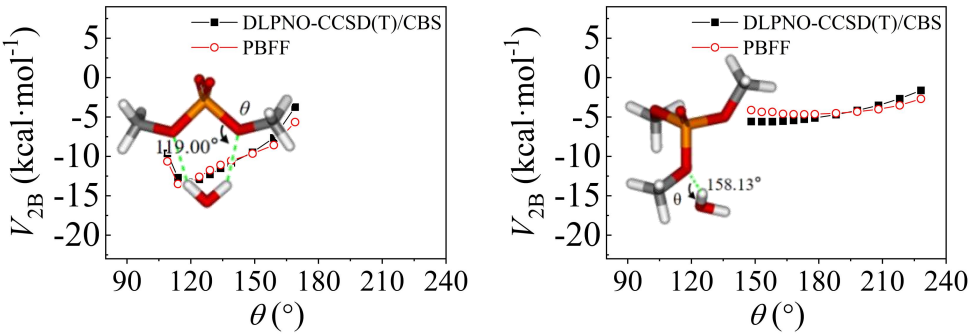


Fig. S11 Intermolecular angle-dependent two-body interaction energies (V_{2B}) for Training Set III.

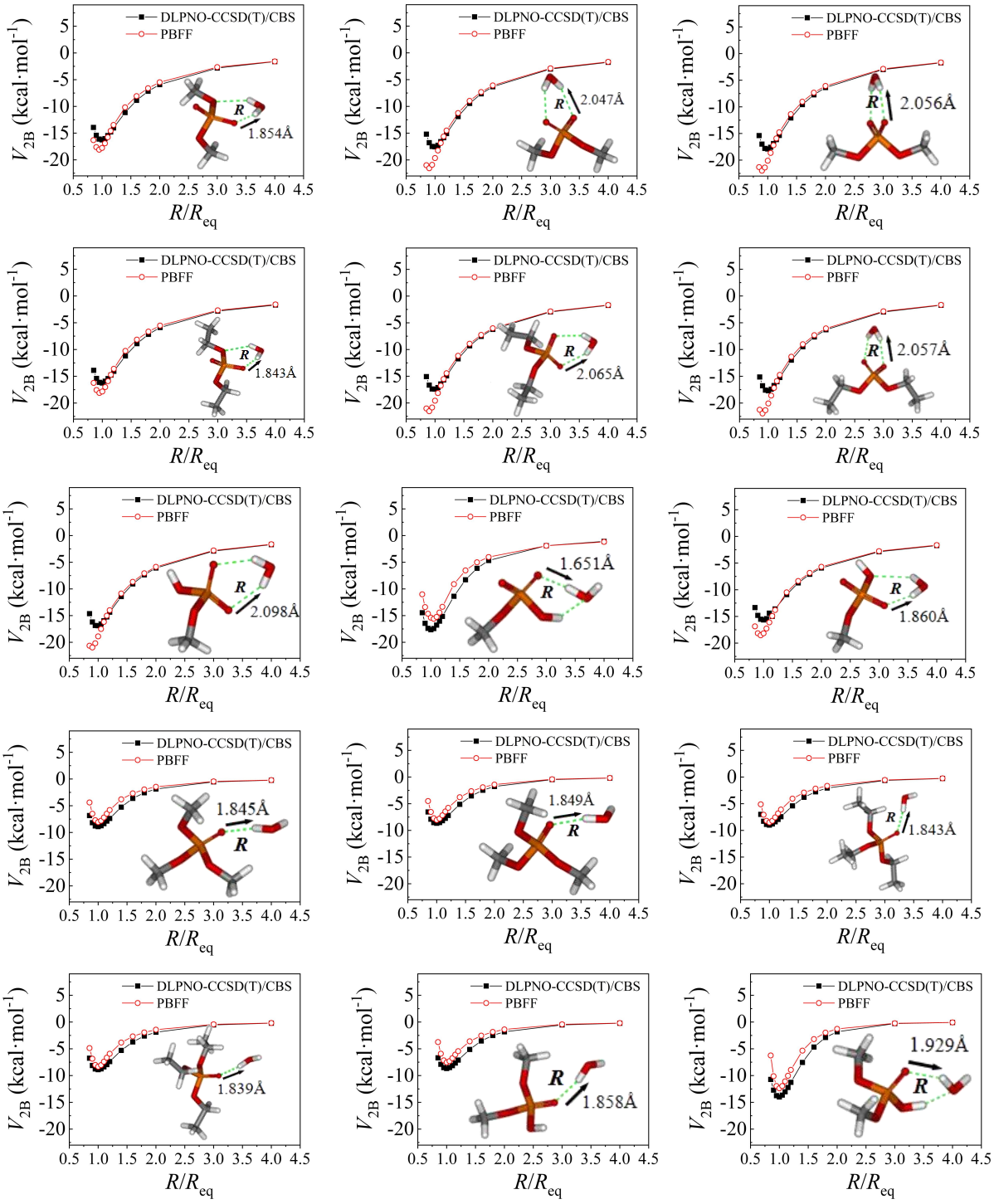


Fig. S12 Distance-dependent two-body interaction energies (V_{2B}) for Training Set IV.

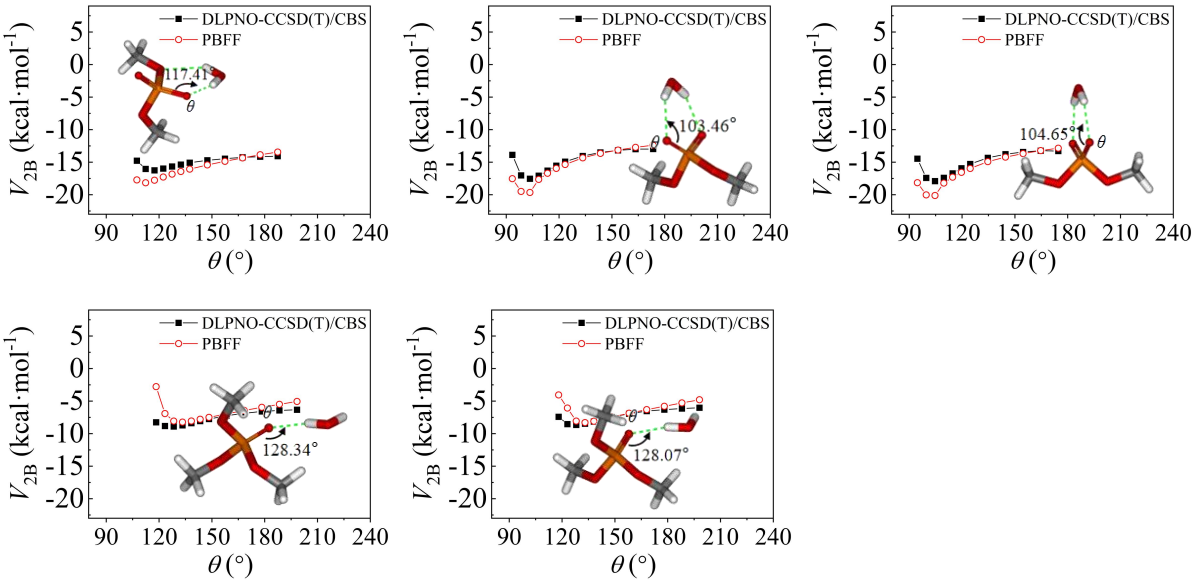


Fig. S13 Intermolecular angle-dependent two-body interaction energies (V_{2B}) for Training Set IV.

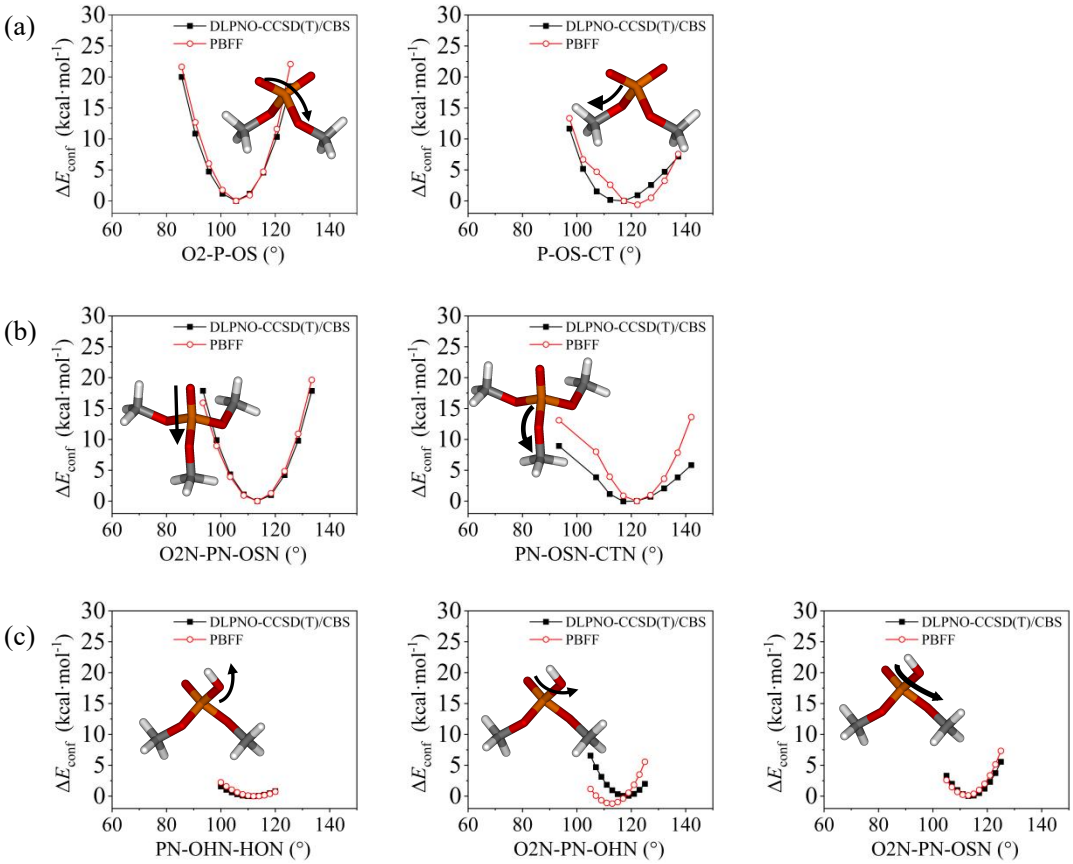


Fig. S14 Angle-dependent conformational energy for Training Set I.

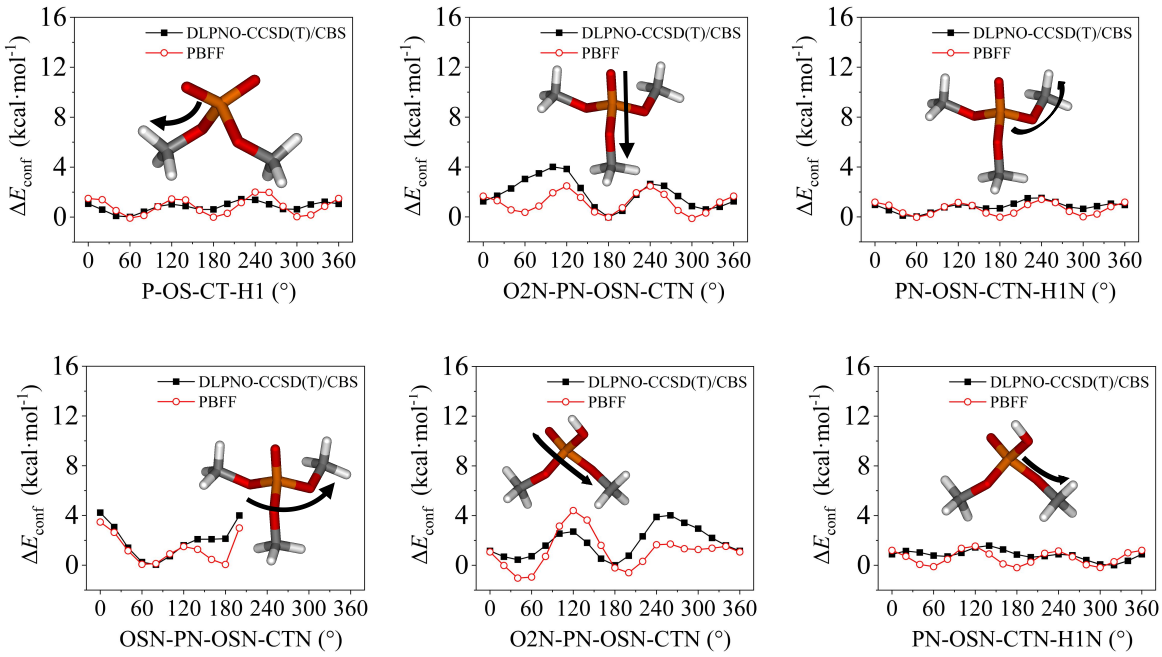
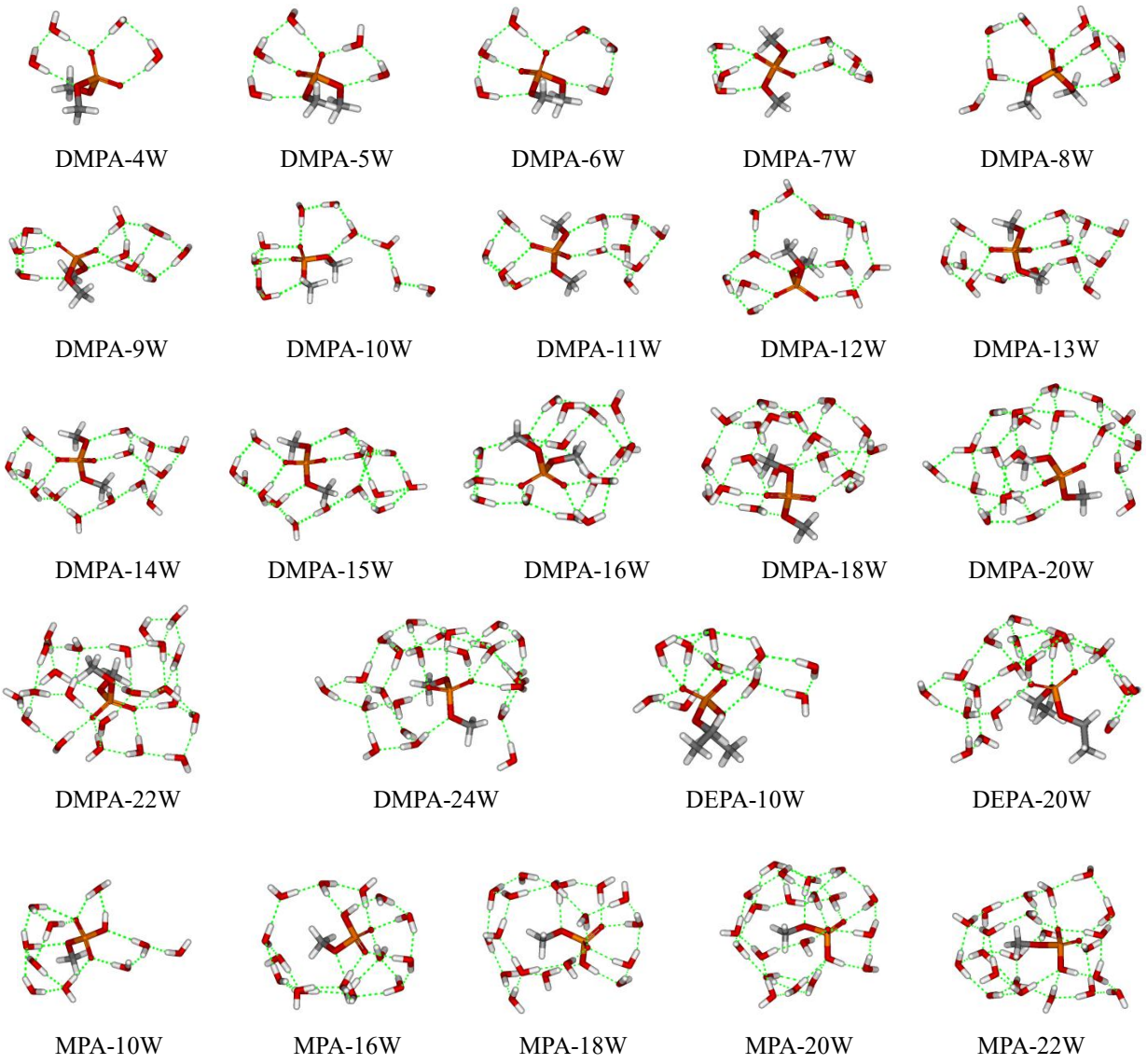
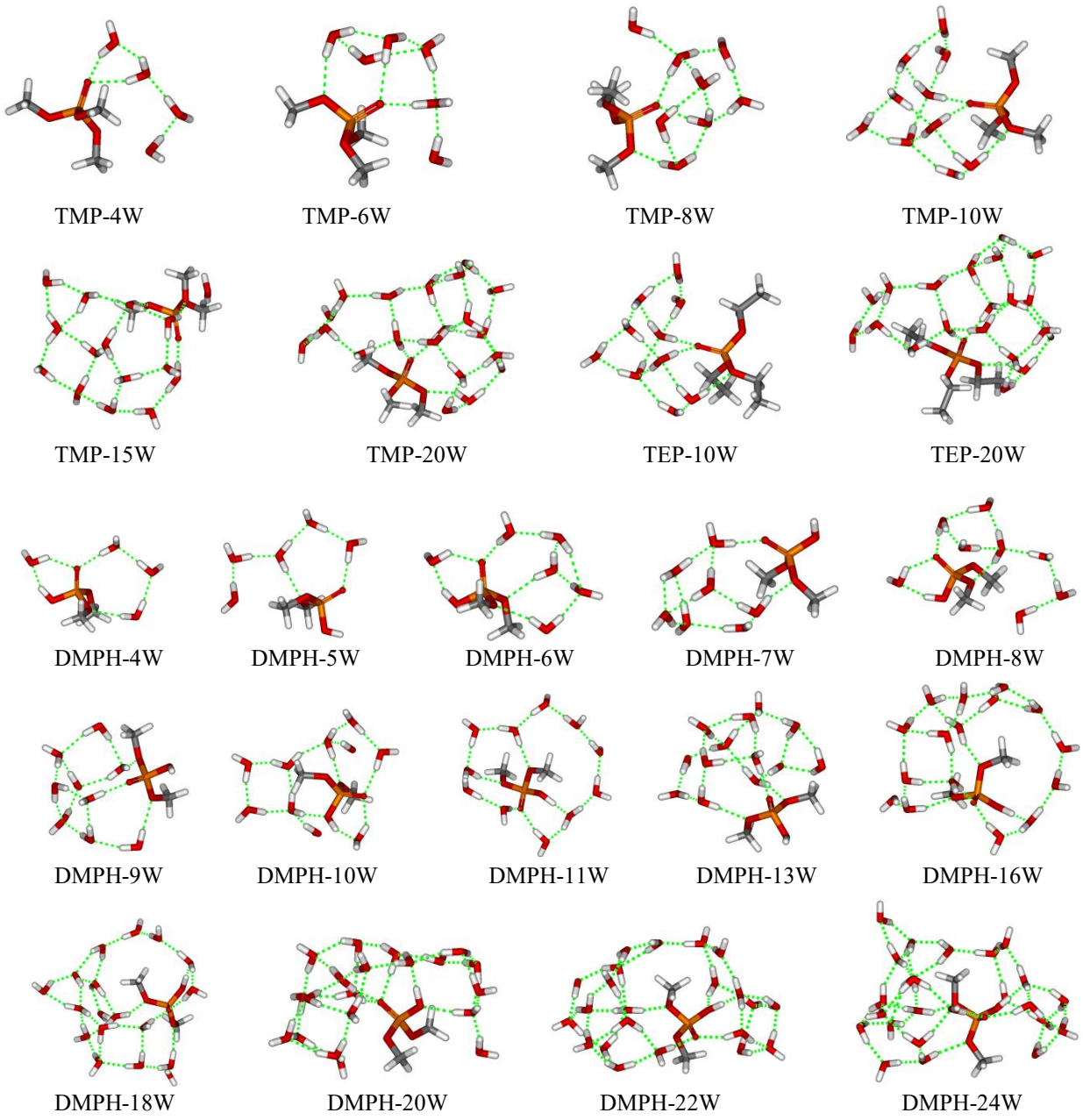


Fig. S15 Dihedral angle-dependent conformational energy for Training Set I.



(a) 24 phosphate anion-water clusters

Fig. S16 Structures of clusters comprising Testing Set II: (a) 24 phosphate anion-water clusters, (b) 22 neutral phosphate-water clusters.



(b) 22 neutral phosphate-water clusters

Fig. S16 (Continued)

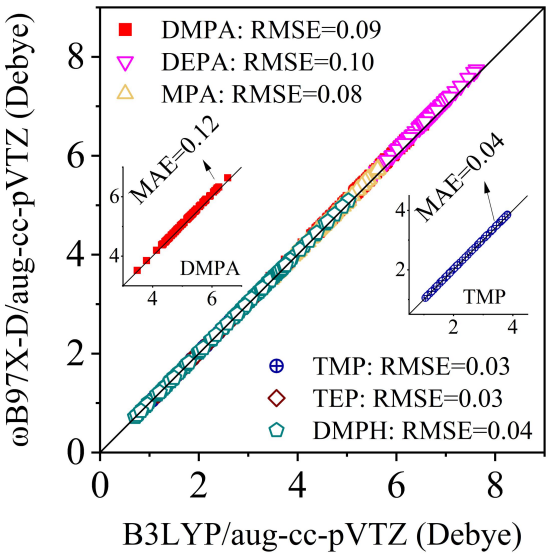


Fig. S17 Correlation of molecular dipole moments for 703 test monomers between B3LYP/aug-cc-pVTZ and ωB97X-D/aug-cc-pVTZ levels.

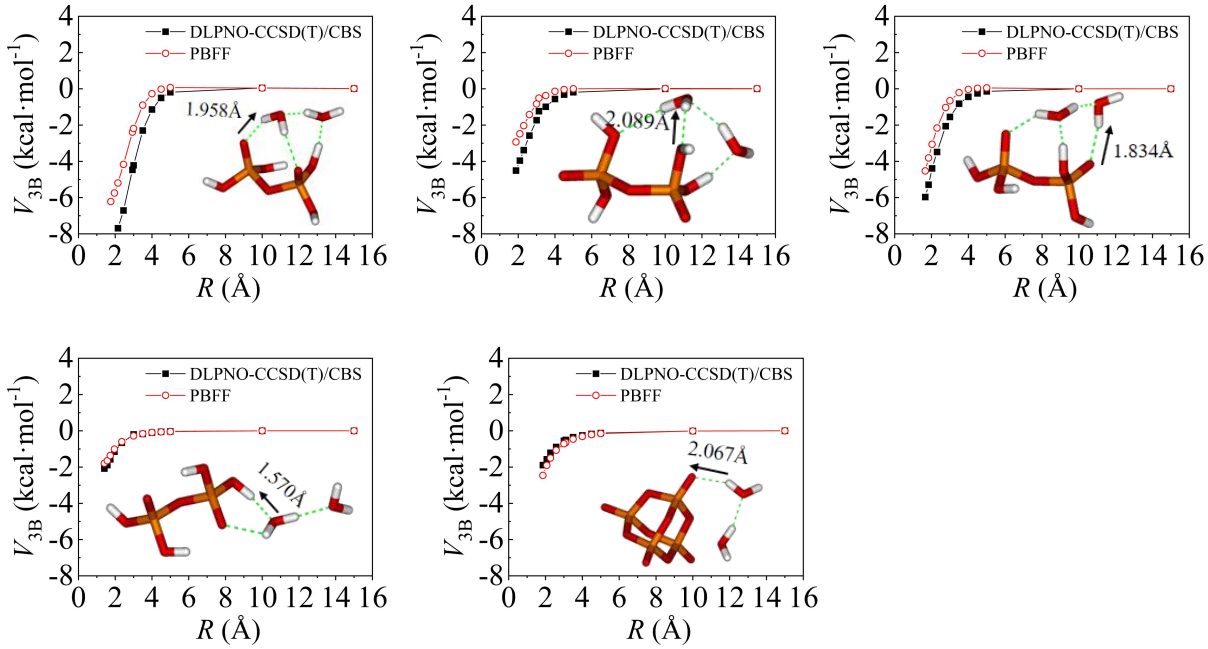


Fig. S18 Distance-dependent three-body interaction energies (V_{3B}) for phosphate anhydrides.

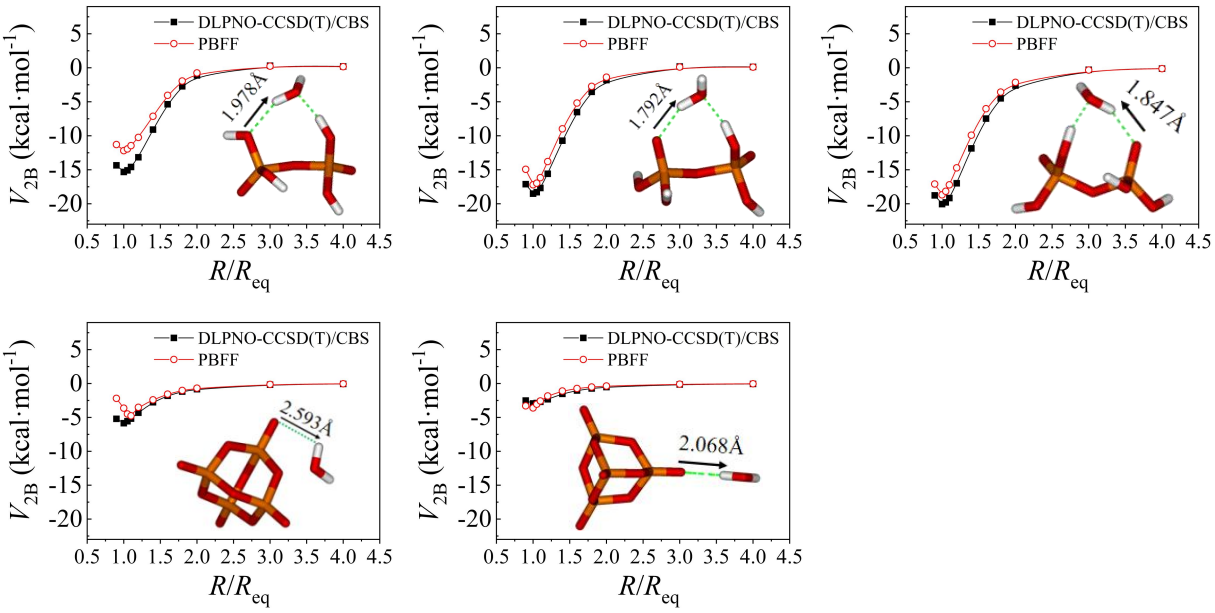


Fig. S19 Distance-dependent two-body interaction energies (V_{2B}) for phosphate anhydrides.

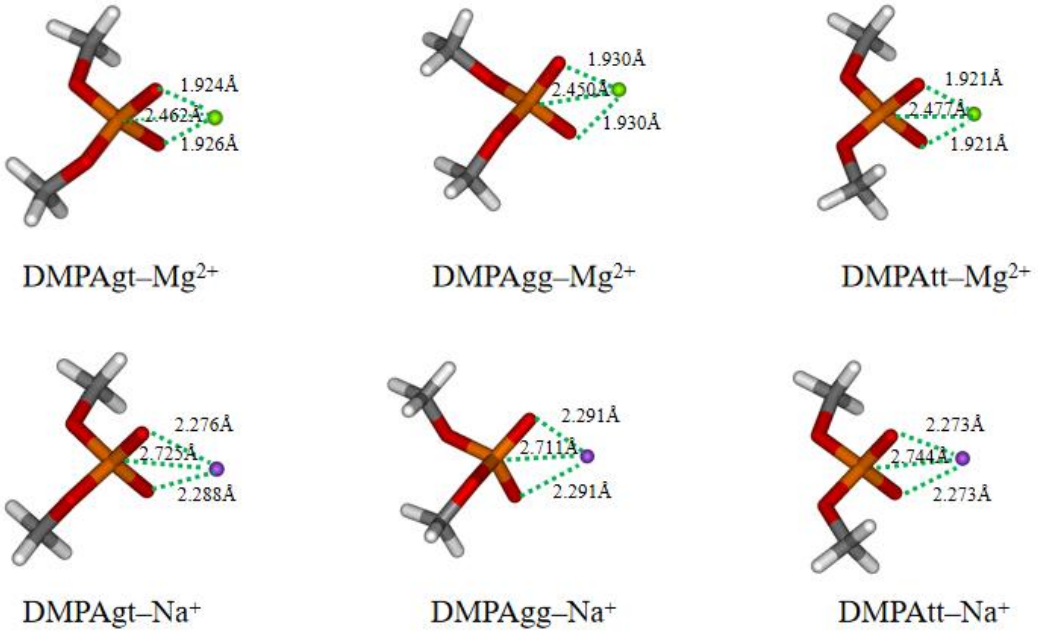


Fig. S20 Interaction geometry of DMPA with a metal ion (Na^+ or Mg^{2+}). The DMPAgt–Mg^{2+} and DMPAgt–Na^+ conformers are from the literature,[16] the remaining ones are extended structures. All optimizations were performed at the MP2/aug-cc-pVTZ level, matching the reported methodology.

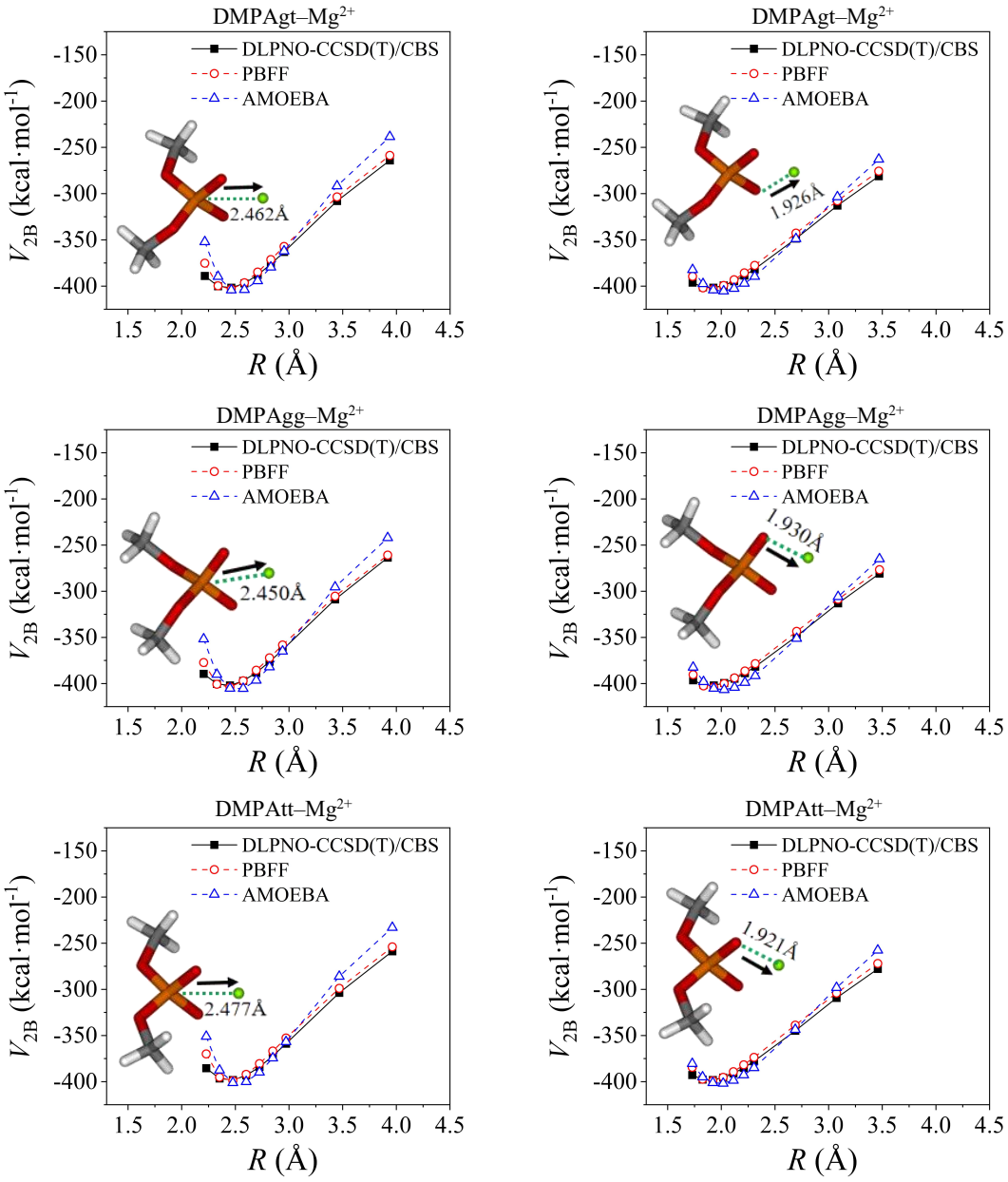


Fig. S21 Distance-dependent two-body interaction energies (V_{2B}) for DMPA–Mg^{2+}.

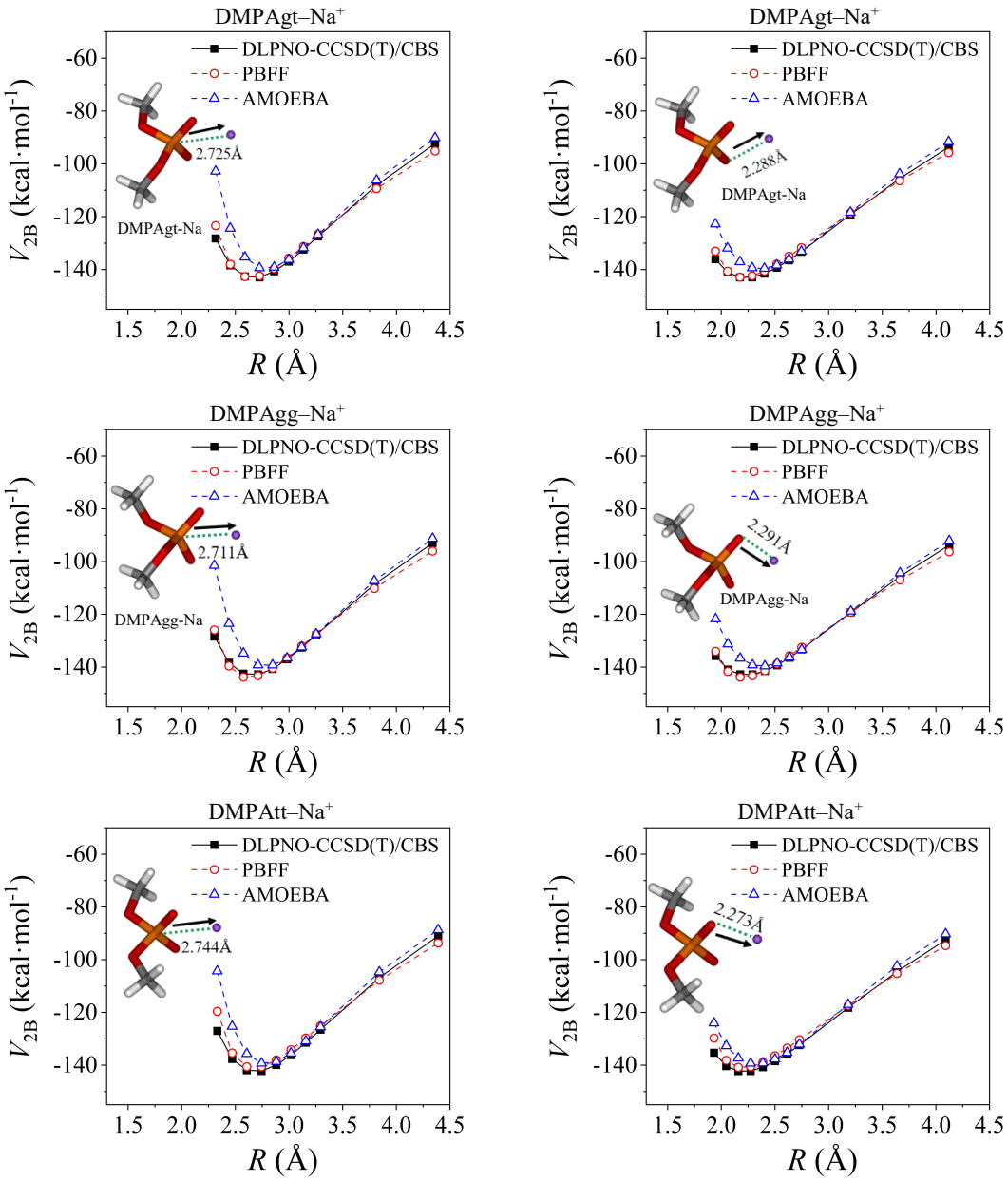


Fig. S22 Distance-dependent two-body interaction energies (V_{2B}) for DMPA–Na^+.

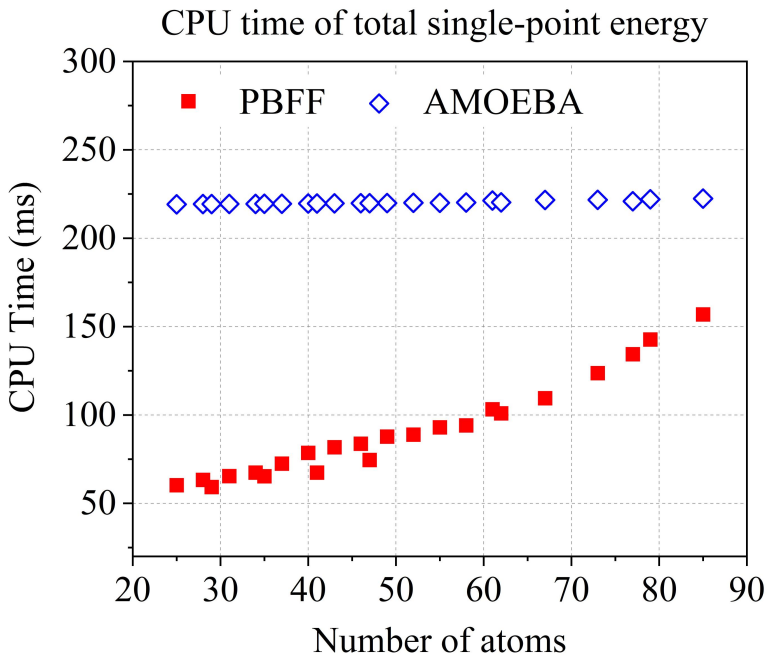


Fig. S23 Total CPU time (including initialization) per single-point evaluation for PBFF and AMOEBA.

References

1. X.-C. Gao, Q. Hao and C.-S. Wang, *J. Chem. Theory Comput*., 2017, **13**, 2730–2741.
2. X.-L. Li, C.-M. Li, J.-Y. Zhu, Z. Zhou, Q. Hao and C.-S. Wang, *J. Comput. Chem.*, 2023, **44**, 677–686.
3. J.-Y. Zhu, Q. Liu, X.-N. Jiang, X.-H. Zheng, L. Wang, Q. Hao and C.-S. Wang, *Phys. Chem. Chem. Phys*., 2023, **25**, 29867–29880.
4. U. Góra, R. Podeszwa, W. Cencek and K. Szalewicz, *J. Chem. Phys.*, 2011, **135**, 224102.
5. M. A. Collins and R. P. A. Bettens, *Chem. Rev.*, 2015, **115**, 5607–5642.
6. A. Nandi, C. Qu, P. L. Houston, R. Conte, Q. Yu and J. M. Bowman, *J. Phys. Chem. Lett.*, 2021, **12**, 10318–10324.
7. J. G. Speight and N. A. Lange, *Lange's Handbook of Chemistry*, 16th ed., McGraw-Hill Professional, New York, 2004.
8. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc*., 1995, **117**, 5179–5197.
9. F. Neese, F. Wennmohs, U. Becker and C. Riplinger, *J. Chem. Phys.*, 2020, **152**, 224108.
10. T. Helgaker, W. Klopper, H. Koch and J. *Noga, J. Chem. Phys.*, 1997, **106**, 9639–9646.
11. D. G. Truhlar, *Chem. Phys. Lett.*, 1998, **294**, 45–48.
12. A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen and A. K. Wilson, *Chem. Phys. Lett*., 1998, **286**, 243–252.
13. A. Karton and J. M. L. Martin, *Theor. Chem. Acc.*, 2006, **115**, 330–333.
14. S.-J. Zhong, E. C. Barnes and G. A. Petersson, *J. Chem. Phys.*, 2008, **129**, 184116.
15. F. Neese and E. F. Valeev, *J. Chem. Theory Comput.*, 2011, **7**, 33–43.
16. C.-S. Zhang, C. Lu, Q. Wang, J. W. Ponder and P.-Y. Ren, *J. Chem. Theory Comput.*, 2015, **11**, 5326–5339.