Supplementary Material for "An efficient, fragment-based electronic structure method for molecular systems: Self-consistent polarization with perturbative two-body exchange and dispersion"

Leif D. Jacobson and John M. Herbert*
Department of Chemistry, The Ohio State University, Columbus, OH 43210

This document provides some additional theoretical details and benchmark calculations. Section S-I is an expanded version of Section II B in the paper, and provides detailed equations for implementing the XPS method with either Mulliken, Löwdin, or CHELPG embedding charges. For CHELPG embedding charges, the original CHELPG algorithm[2] must be modified in order to avoid discontinuities as the nuclei are displaced (since CHELPG uses atom-centered Cartesian grids to discretize the electrostatic potential). The weighted least-squares procedure that is used to ensure continuity is discussed in Section S-II. Finally, in Section S-III, we present expanded versions of some of the figures and tables from the paper. These provide additional benchmark data, using a wider variety of basis sets and density functionals.

S-I. DETAILS OF THE CHARGE SCHEMES

We investigate three different charge schemes: Mulliken, Löwdin and CHELPG. The Mulliken and Löwdin charges stem from two different partitions of the electron density,[1]

$$q_J^{\text{Mull}} = Z_J - \sum_{\substack{\mu\in J\\ \nu}} S_{\mu\nu} P_{\nu\mu} \tag{S1a}$$

$$q_J^{\text{Lowd}} = Z_J - \sum_{\substack{\mu\in J\\ \nu,\lambda}} (\mathbf{S}^{1/2})_{\mu\nu} P_{\nu\lambda} (\mathbf{S}^{1/2})_{\lambda\mu} \; . \tag{S1b}$$

Derivatives of these charges with respect to density matrix elements are quite simple:

$$(\Lambda_J^{\text{Mull}})_{\mu\nu} = -\frac{1}{2}\left(S_{\mu\nu}\delta_{\mu\in J} + S_{\nu\mu}\delta_{\nu\in J}\right) \tag{S2a}$$

$$(\Lambda_J^{\text{Lowd}})_{\mu\nu} = -\sum_{\lambda\in J} (\mathbf{S}^{1/2})_{\lambda\mu} (\mathbf{S}^{1/2})_{\nu\lambda} \; . \tag{S2b}$$

The quantity $\delta_{\mu\in J} = 1$ if the basis function μ is centered on atom J, and is zero otherwise.

CHELPG charges[2] are designed to minimize the difference between the electrostatic potential, $\Phi(\vec{r})$, that is generated by the electron density, and the electrostatic potential, $\phi(\vec{r})$, that is generated by a set of atom-centered point charges. This minimization is performed subject to the constraint that the sum of the point charges equals the overall system charge, Q, and the potentials are evaluated at a set of grid points $\vec{R}_k$. The CHELPG charges are defined as those that minimize the quantity

$$\mathcal{L} = \sum_{k}^{N_{grid}} (\Phi_k - \phi_k)^2 + \lambda \left(\sum_{J}^{N_{atoms}} q_J - Q \right) \; , \tag{S3}$$

where λ is a Lagrange multiplier. The true electrostatic potential at the ith grid point, $\Phi_k = \Phi(\vec{R}_k)$, is

$$\Phi_k = \sum_{J}^{N_{atoms}} \frac{Z_J}{\left|\vec{R}_k - \vec{R}_J\right|} - \sum_{\mu\nu} (I_k)_{\mu\nu} P_{\mu\nu} \; . \tag{S4}$$

---

* `herbert@chemistry.ohio-state.edu`

The electrostatic potential $\phi_k = \phi(\vec{R}_k)$ that is generated by the point charges is

$$\phi_k = \sum_{J}^{N_{atoms}} \frac{q_J}{\big|\vec{R}_k - \vec{R}_J\big|} \; . \tag{S5}$$

Minimization of $\mathcal{L}$ with respect to variation of q_J defines the CHELPG charges:

$$q_J^{\text{CHELPG}} = (\mathbf{G}^{-1}\mathbf{e})_J - \left(\frac{\text{tr}(\mathbf{G}^{-1}\mathbf{e}) - Q}{\text{tr}(\mathbf{G}^{-1})} \right) \sum_K (\mathbf{G}^{-1})_{JK} \; . \tag{S6}$$

Here, the matrix $\mathbf{G}$ is defined by

$$G_{IJ} = \sum_{k}^{N_{grid}} \big|\vec{R}_I - \vec{R}_k\big|^{-1} \big|\vec{R}_k - \vec{R}_J\big|^{-1} \tag{S7}$$

and the vector $\mathbf{e}$ has elements

$$e_J = \sum_{k}^{N_{grid}} \frac{\Phi_k}{\big|\vec{R}_k - \vec{R}_J\big|} \; . \tag{S8}$$

It is straightforward to take the derivative of Eq. (S6) to obtain

$$\begin{aligned}(\Lambda_J^{\text{CHELPG}})_{\mu\nu} = &- \sum_K (\mathbf{G}^{-1})_{JK} \sum_{k}^{N_{grid}} \frac{(I_k)_{\mu\nu}}{\big|\vec{R}_k - \vec{R}_K\big|} \\ &+ \left(\frac{\sum_{KL} (\mathbf{G}^{-1})_{KL} \sum_k^{N_{grid}} (I_k)^{-1}_{\mu\nu} \big|\vec{R}_k - \vec{R}_K\big|^{-1}}{\text{tr}(\mathbf{G}^{-1})} \right) \sum_K (\mathbf{G}^{-1})_{JK}\end{aligned} \tag{S9}$$

In what follows, we will require an operator, $\hat{\Lambda}_J$, whose matrix elements are equal to $(\Lambda_J)_{\mu\nu}$ as given by Eq. (S2) or Eq. (S9). For CHELPG charges, this operator is trivially constructed by making the replacement $\mathbf{I}_k \rightarrow \hat{I}_k$ in Eq. (S9). The operators corresponding to the matrices in Eq. (S2) are less obvious, but it can be verified that the appropriate choice is

$$\hat{\Lambda}_J^{\text{Mull}} = -\tfrac{1}{2} \sum_{\mu,\nu \in J} \Big(|\mu\rangle (\mathbf{S}_J^{-1})_{\mu\nu} \langle\nu| + |\nu\rangle (\mathbf{S}_J^{-1})_{\nu\mu} \langle\mu| \Big) \tag{S10a}$$

$$\hat{\Lambda}_J^{\text{Lowd}} = \sum_{\mu,\lambda,\sigma \in J} |\lambda\rangle (\mathbf{S}_J^{-1/2})_{\lambda\mu} (\mathbf{S}_J^{-1/2})_{\mu\sigma} \langle\sigma| \; . \tag{S10b}$$

The quantities $\mathbf{S}_J^{-1}$ and $\mathbf{S}_J^{-1/2}$ in this equation refer to the inverse and inverse square-root of the *fragment* overlap matrix, for the fragment that contains atom J. At no point is it necessary to invert the supersystem's overlap matrix.

S-II. WEIGHTED LEAST SQUARES CHELPG CHARGES

We have not yet implemented analytic gradients for the XPS method, so the geometry optimizations reported here were performed using a three-point finite difference of the total energy, with atomic displacements of 10^{-3} bohr. The number of CHELPG grid points may change as the nuclei are displaced, so to avoid discontinuities we replace the Lagrangian in Eq. (S3) with a weighted Lagrangian,

$$\mathcal{L} = \sum_{k}^{N_{grid}} w_k \left(\Phi_k - \phi_k\right)^2 + \lambda \left(\sum_{J}^{N_{atoms}} q_J - Q \right) . \tag{S11}$$

The weight w_k for grid point k is taken to be a product of short- and long-range weighting functions, $w_k = w_k^{short}\, w_k^{long}$. The short-range weight is itself a product over atomic weights, $w_k^{short} = \Pi_J A_k^J$, where

$$A_k^J = \begin{cases} 0 & \text{if } |\vec{R}_k - \vec{R}_J| < R_{cut,J}^{short} \\ \tau(|\vec{R}_k - \vec{R}_J|, R_{cut,J}^{short}, R_{on,J}) & \text{if } R_{cut,J}^{short} \le |\vec{R}_k - \vec{R}_J| < R_{on,J} \\ 1 & \text{otherwise} \end{cases} . \tag{S12}$$

The cutoff parameters $R_{cut,J}^{short}$ and $R_{on,J}$ are given below. The tapering function, τ, is taken from Ref. 3:

$$\tau(R, R_{cut}, R_{\mathit{off}}) = \frac{(R - R_{cut})^2 (3R_{\mathit{off}} - R_{cut} - 2R)}{(R_{\mathit{off}} - R_{cut})^3} . \tag{S13}$$

To determine the long-range weight, we first find the minimum distance from the grid point $\vec{R}_k$ to any nucleus,

$$R_k^{min} = \min_J |\vec{R}_k - \vec{R}_J| . \tag{S14}$$

Then

$$w_k^{long} = \begin{cases} 1 & \text{if } R_k^{min} < R_{cut}^{long} \\ 0 & \text{if } R_k^{min} > R_{\mathit{off}} \\ 1 - \tau(R_k^{min}, R_{cut}^{long}, R_{\mathit{off}}) & \text{otherwise} \end{cases} \tag{S15}$$

To evaluate the weights, we set $R_{cut,J}^{short}$ equal to the Bondi radius for atom J. We set $R_{\mathit{off}} = 3.0$ Å, $R_{on,J} = R_{cut,J}^{short} + \Delta r$, and $R_{cut}^{long} = R_{\mathit{off}} - \Delta r$, where the quantity Δr controls how rapidly a grid point's weight is scaled to zero by the tapering function. We use a fairly small value, $\Delta r = 0.1$ bohr, because we were concerned about possible discontinuities arising on the length scale of the finite-difference steps. Although it may be necessary to increase this to ensure smoothness in molecular dynamics applications, we have not encountered difficulties in geometry optimizations.

S-III. EXPANDED VERSIONS OF DATA TABLES AND FIGURES

Table S1 is an expanded version of Table I from the paper, and provides a statistical summary of how XPS performs for the S22 database, when Mulliken embedding charges are employed. Table S2 is an expanded version of Table II from the paper, and shows the XPS results for the S22 database using either Löwdin or CHELPG embedding charges. Figure S1 expands upon Fig. 1 from the paper, and shows errors over the S22 database for several other density functionals, including B3LYP. (B3LYP results were not discussed in the paper because, in the context of XPS, they tend to be similar to—but of slightly lower quality than—PBE0 results.) Figures S2 and S3 are analogous to Figs. 3 and 4 in the paper, and show benzene dimer potential energy curves computed at the XPS(0) level, but for a wider variety of basis sets than were considered in the paper. Figure S4 expands upon Fig. 7 from the paper and shows the performance of XPS(0) for water clusters, using a wider variety of basis sets. Numerical results for this wider variety of basis sets are provided in Table S3.

Method	SCF method, X =					
	HF	B3LYP	BOP	PBE0	LRC-μBOP	LRC-ωPBEh
SAPT(X)/3-21G*	0.84 (2.25)	0.97 (2.92)	1.32 (4.96)	1.01 (2.86)	0.82 (4.24)	0.68 (3.62)
SAPT(X)-*resp*/3-21G*	0.91 (2.72)	1.04 (5.23)	1.41 (5.67)	0.98 (3.89)	0.87 (4.61)	0.74 (4.53)
XPS(X)/3-21G*	1.04 (3.06)	0.92 (3.76)	1.31 (4.70)	0.84 (2.43)	0.81 (3.14)	0.65 (2.97)
XPS(X)-*resp*/3-21G*	1.10 (3.43)	0.91 (3.71)	1.28 (4.61)	0.84 (2.38)	0.80 (2.95)	0.65 (2.83)
SAPT(X)/6-31G*	0.79 (3.27)	1.76 (6.82)	2.02 (9.55)	1.52 (6.47)	1.56 (7.31)	1.45 (6.75)
SAPT(X)-*resp*/6-31G*	0.65 (2.48)	2.03 (7.88)	2.50 (9.21)	1.73 (6.50)	1.52 (7.33)	1.51 (7.36)
XPS(X)/6-31G*	0.56 (1.46)	1.43 (6.70)	2.00 (9.66)	1.33 (6.24)	1.10 (3.79)	0.99 (3.57)
XPS(X)-*resp*/6-31G*	0.90 (3.16)	1.39 (6.37)	1.87 (9.01)	1.28 (5.99)	0.80 (3.14)	0.82 (3.11)
SAPT(X)/6-311G*	1.24 (6.40)	2.27 (8.43)	2.58 (9.80)	2.07 (7.48)	1.88 (9.63)	1.90 (9.51)
SAPT(X)-*resp*/6-311G*	1.09 (5.51)	2.54 (10.44)	3.05 (11.19)	2.26 (9.20)	1.83 (9.55)	1.96 (10.04)
XPS(X)/6-311G*	1.11 (4.91)	2.20 (8.15)	2.99 (12.76)	1.98 (7.39)	1.71 (6.70)	1.72 (6.95)
XPS(X)-*resp*/6-311G*	0.57 (2.22)	2.05 (6.82)	2.54 (8.14)	1.81 (5.65)	1.16 (5.91)	1.41 (6.37)
SAPT(X)/cc-pVDZ	1.06 (4.45)	2.14 (6.79)	2.45 (9.58)	1.83 (6.27)	1.89 (8.30)	1.81 (7.76)
SAPT(X)-*resp*/cc-pVDZ	0.91 (3.72)	2.41 (8.75)	2.92 (9.41)	2.06 (7.46)	1.87 (8.34)	1.88 (8.37)
XPS(X)/cc-pVDZ	0.71 (1.94)	2.09 (7.44)	2.58 (10.11)	1.73 (6.81)	1.62 (5.93)	1.59 (5.83)
XPS(X)-*resp*/cc-pVDZ	0.50 (1.81)	2.02 (6.45)	2.52 (9.00)	1.66 (5.90)	1.38 (5.43)	1.42 (5.45)

TABLE S1: Mean absolute errors and (in parentheses) maximum absolute errors for the the S22 database, in kcal/mol. A variety of SAPT(X) and XPS(X) variants are considered; note that SAPT(HF) is equivalent to the method that is traditionally called SAPT(0). All XPS methods use Mulliken charges for the electrostatic embedding.

	XPS-Löwdin			XPS-CHELPG		
Basis	HF	LRC-μBOP	LRC-ωPBEh	HF	LRC-μBOP	LRC-ωPBEh
3-21G*	1.09 (3.52)	0.81 (3.01)	0.66 (2.87)	1.17 (3.95)	0.79 (2.57)	0.67 (2.46)
resp	1.11 (3.61)	0.81 (2.98)	0.66 (2.87)	1.17 (3.98)	0.79 (2.56)	0.67 (2.45)
6-31G*	0.73 (2.31)	0.91 (3.91)	0.93 (3.86)	0.87 (2.47)	0.87 (3.56)	0.90 (3.53)
resp	0.75 (2.40)	0.90 (3.88)	0.93 (3.88)	0.87 (2.55)	0.87 (3.54)	0.91 (3.55)
6-31+G*	0.90 (2.08)	2.52 (9.06)	1.94 (8.32)	0.89 (1.99)	2.56 (9.03)	1.98 (8.26)
resp	0.89 (1.99)	2.55 (9.12)	1.96 (8.40)	0.90 (2.10)	2.60 (9.05)	2.00 (8.32)
6-31+G**	1.11 (3.94)	2.76 (10.85)	2.22 (9.98)	1.03 (3.22)	2.73 (10.29)	2.20 (9.40)
resp	1.08 (3.75)	2.77 (10.84)	2.24 (10.03)	1.03 (3.08)	2.77 (10.30)	2.21 (9.46)
6-311G*	0.57 (2.76)	1.20 (6.52)	1.38 (6.92)	0.56 (2.32)	1.21 (6.20)	1.40 (6.67)
resp	0.54 (2.55)	1.18 (6.41)	1.37 (6.88)	0.54 (2.18)	1.20 (6.15)	1.40 (6.65)
6-311+G*	1.13 (6.63)	2.21 (12.31)	2.24 (12.25)	1.20 (6.73)	2.31 (12.55)	2.34 (12.48)
resp	1.11 (6.48)	2.22 (12.30)	2.25 (12.28)	1.20 (6.59)	2.34 (12.54)	2.37 (12.50)
6-311+G**	1.23 (7.12)	2.37 (12.75)	2.39 (12.54)	1.19 (6.60)	2.37 (12.39)	2.41 (12.20)
resp	1.18 (6.81)	2.36 (12.65)	2.39 (12.53)	1.19 (6.47)	2.39 (12.38)	2.43 (12.23)
cc-pVDZ	0.55 (2.04)	1.46 (6.39)	1.48 (6.30)	0.39 (1.12)	1.35 (5.72)	1.39 (5.68)
resp	0.51 (1.75)	1.44 (6.26)	1.47 (6.27)	0.38 (1.02)	1.35 (5.70)	1.39 (5.69)
aug-cc-pVDZ	1.52 (4.48)	2.46 (10.24)	2.71 (9.63)	1.26 (3.38)	2.21 (8.43)	2.49 (7.91)
resp	1.40 (3.80)	2.39 (9.91)	2.67 (9.52)	1.25 (3.39)	2.23 (8.40)	2.49 (7.91)
aug-cc-pVDZ′	—	—	—	1.31 (3.86)	—	—
aug-cc-pVDZ-proj	—	—	—	1.31 (4.42)	1.66 (4.36)	2.05 (6.51)
aug-cc-pVDZ′-proj	—	—	—	0.75 (3.38)	—	—

TABLE S2: Mean absolute errors and (in parentheses) maximum absolute errors for the S22 database, in kcal/mol. A variety of XPS(X) variants are considered, using either Löwdin or CHELPG embedding charges. For several of the basis sets, the corresponding response ("*resp*") result is also listed. The primed and projected ("proj") basis sets are defined in the paper.

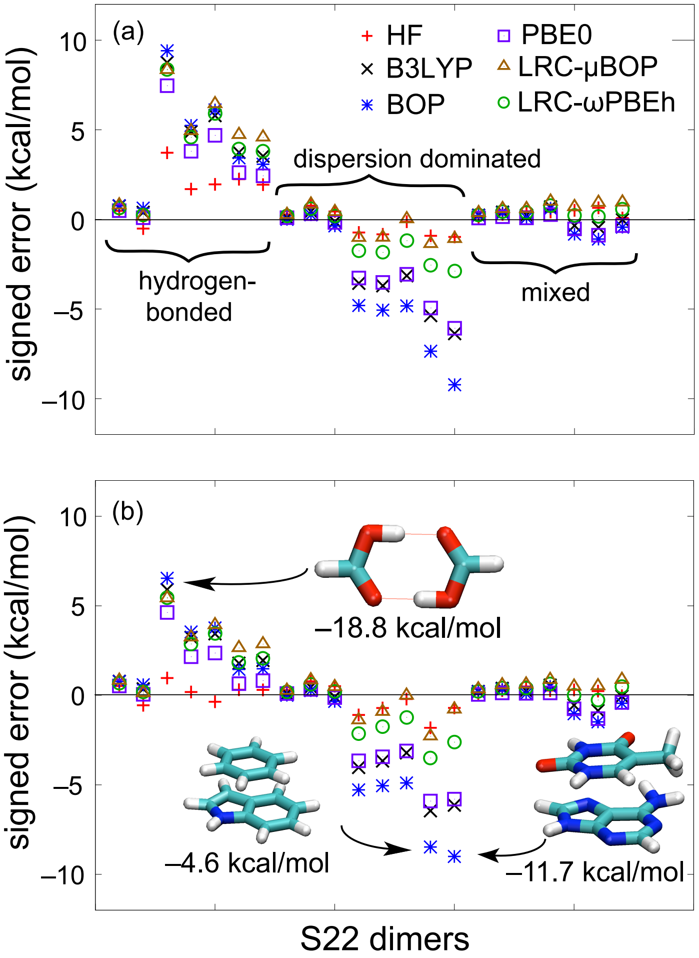

FIG. S1: Binding energy errors (in kcal/mol) across the S22 database, as computed at (a) the SAPT(X)/cc-pVDZ level (for various SCF methods, X) and (b) the XPS(X)/cc-pVDZ level, for a variety of different density functionals, X. A few difficult cases are highlighted in panel (b). Starting at the top and moving clockwise around panel (b), these are formic acid dimer, an indole-benzene π stack, and an adenine-thymine π stack.

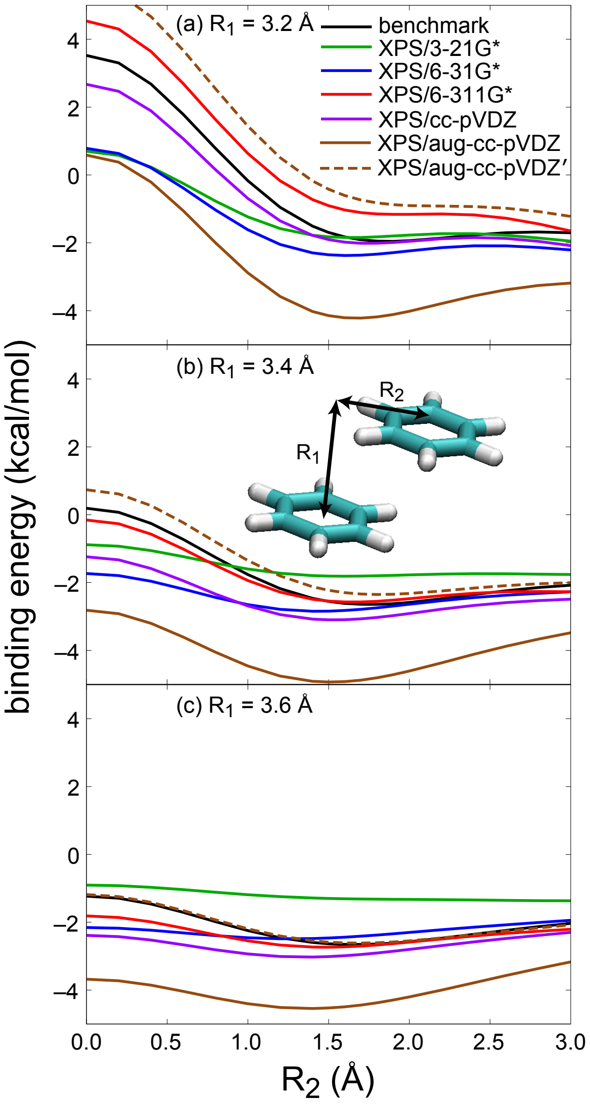

FIG. S2: Benchmark and XPS(0) binding energy curves for the parallel-displaced benzene dimer, computed at the XPS(0) level using CHELPG charges. CCSD(T)/CBS benchmarks are taken from Ref. 4.

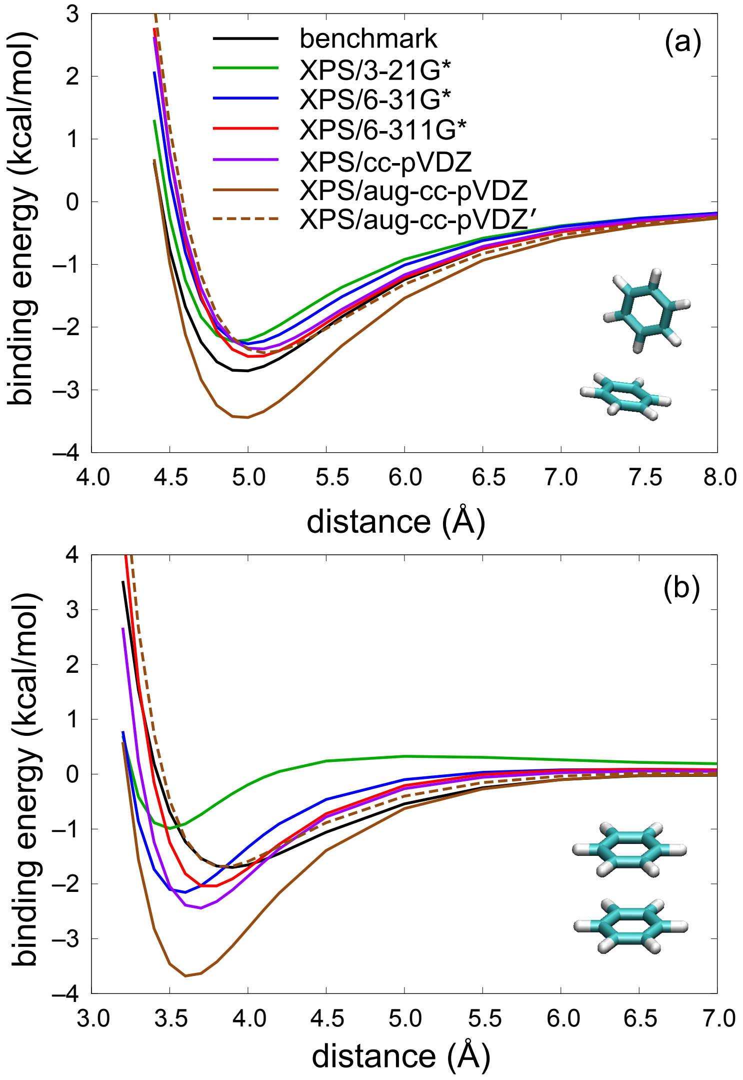

FIG. S3: Benchmark and XPS(0) binding energy curves for (a) the T-shaped and (b) the “sandwich” isomers of the benzene dimer. Benchmark CCSD(T)/CBS values are taken from Ref. 4. The distance coordinate in both panels is the center-to-center distance between the benzene rings.

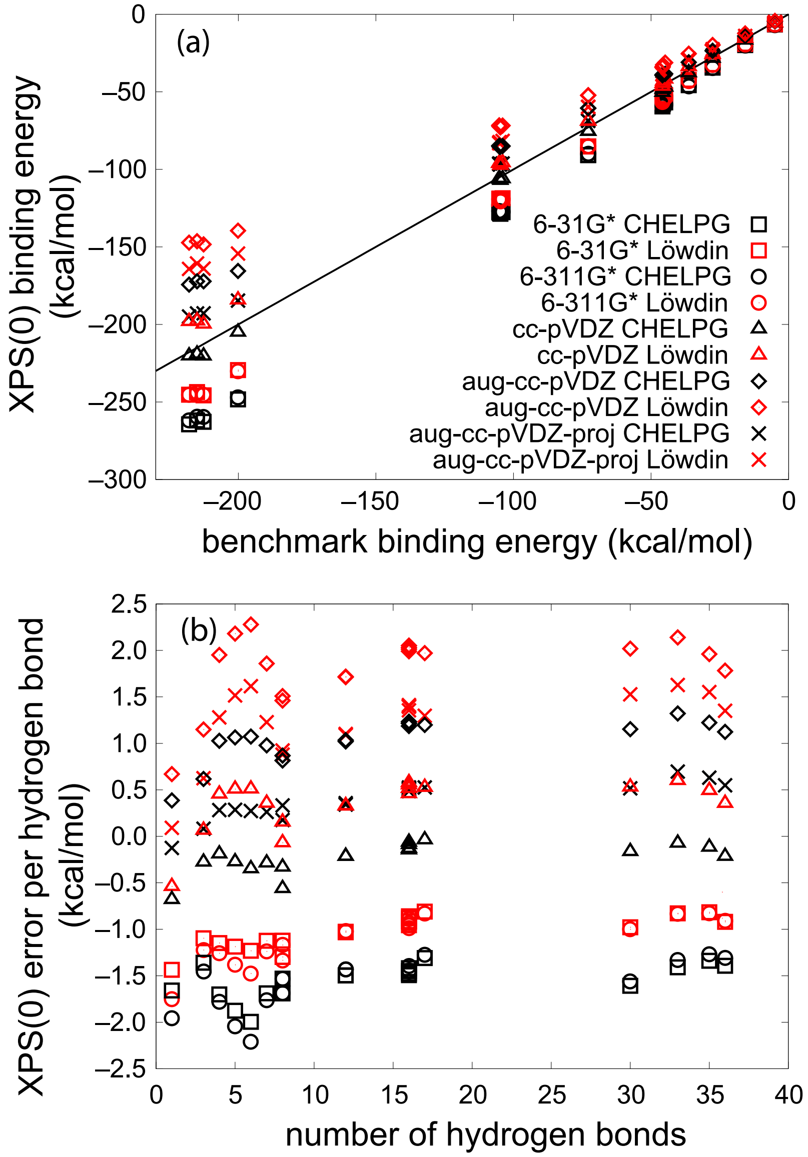

FIG. S4: (a) Correlation between XPS binding energies and MP2 benchmarks. (b) XPS error per hydrogen bond, as a function of the number of hydrogen bonds.

n	Isomer	XPS(0)-CHELPG					Benchmark
		6-31G*	6-311G*	cc-pVDZ	aug-cc-pVDZ	aug-cc-pVDZ-proj	
2		6.6 (33.4)	6.9 (39.4)	5.6 (13.7)	4.6 (7.8)	5.1 (2.5)	4.97
3		19.9 (25.8)	20.2 (27.6)	16.6 (5.3)	14.0 (11.7)	15.6 (1.6)	15.82
4		34.4 (24.6)	34.8 (25.8)	28.4 (2.7)	23.5 (14.9)	26.5 (4.1)	27.63
5		45.7 (25.8)	46.5 (28.1)	37.7 (3.8)	31.0 (14.7)	34.9 (3.9)	36.31
6	book	57.4 (25.9)	57.9 (27.0)	47.6 (4.4)	38.8 (15.0)	43.8 (4.0)	45.61
6	cage	58.0 (26.7)	58.0 (26.7)	48.4 (5.8)	38.8 (15.2)	43.8 (4.2)	45.79
6	cyclic	56.8 (26.7)	58.1 (29.6)	47.0 (4.7)	38.4 (14.3)	43.2 (3.7)	44.86
6	prism	59.4 (29.5)	59.4 (29.4)	50.4 (9.8)	39.3 (14.2)	44.5 (3.1)	45.86
8	D_{2d}	90.8 (24.6)	90.1 (23.6)	75.5 (3.5)	60.5 (17.0)	68.6 (5.9)	72.88
8	S_4	90.8 (24.7)	90.0 (23.6)	75.4 (3.5)	60.6 (16.8)	68.7 (5.6)	72.83
11	$43'4$	127.4 (21.2)	126.8 (20.6)	105.8 (0.6)	84.8 (19.4)	96.2 (8.5)	105.16
11	$44'3'$	128.7 (22.8)	128.0 (22.2)	107.1 (2.2)	85.3 (18.6)	96.6 (7.8)	104.76
11	515	127.8 (21.6)	127.4 (21.2)	106.1 (1.0)	85.4 (18.8)	96.7 (8.0)	105.09
11	551	128.2 (22.1)	128.0 (22.0)	106.4 (1.4)	85.5 (18.5)	96.8 (7.8)	104.95
11	$44'12$	127.6 (22.7)	127.2 (22.4)	106.0 (2.0)	85.0 (18.3)	96.2 (7.5)	103.97
20	dodecahedron	248.4 (24.1)	247.0 (23.4)	205.0 (2.5)	165.5 (17.3)	184.6 (7.7)	200.10
20	edge-sharing	264.4 (21.3)	261.8 (20.1)	220.3 (1.1)	174.3 (20.0)	194.9 (10.6)	217.90
20	face-sharing	261.9 (21.8)	259.4 (20.6)	219.1 (1.9)	172.2 (19.9)	192.9 (10.3)	215.00
20	fused cubes	262.8 (23.6)	259.7 (22.1)	220.4 (3.6)	172.2 (19.0)	192.8 (9.3)	212.60

TABLE S3: Negative binding energies for $(\mathrm{H_2O})_n$ cluster isomers, in kcal/mol. Percent errors in the XPS binding energies, relative to the benchmarks, are listed in parentheses.

---

[1] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry*, Macmillan, New York, 1982.
[2] C. M. Breneman and K. B. Wiberg, J. Comput. Chem. **11**, 361 (1990).
[3] O. Steinhauser, Mol. Phys. **45**, 335 (1982).
[4] C. D. Sherrill, T. Takatani, and E. G. Hohenstein, J. Phys. Chem. A **113**, 10146 (2009).