Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Polarization Consistent Basis Sets with Projector Augmented Wave Method: A Renovation Brought by PAW into Gaussian Basis Sets

Quan Manh Phung,*,†,‡ Masaya Hagai,† Xiao-Gen Xiong,¶ and Takeshi Yanai*,†,‡

†*Department of Chemistry, Graduate School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8602, Japan*

‡*Institute of Transformative Bio-Molecules (WPI-ITbM), Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8602, Japan*

¶*Sino-French Institute of Nuclear Engineering and Technology, Sun Yat-Sen University, Zhuhai, 519082, China*

E-mail: quan.phung@chem.nagoya-u.ac.jp; yanait@chem.nagoya-u.ac.jp

Electronic Supplementary Information (ESI)

List of Figures

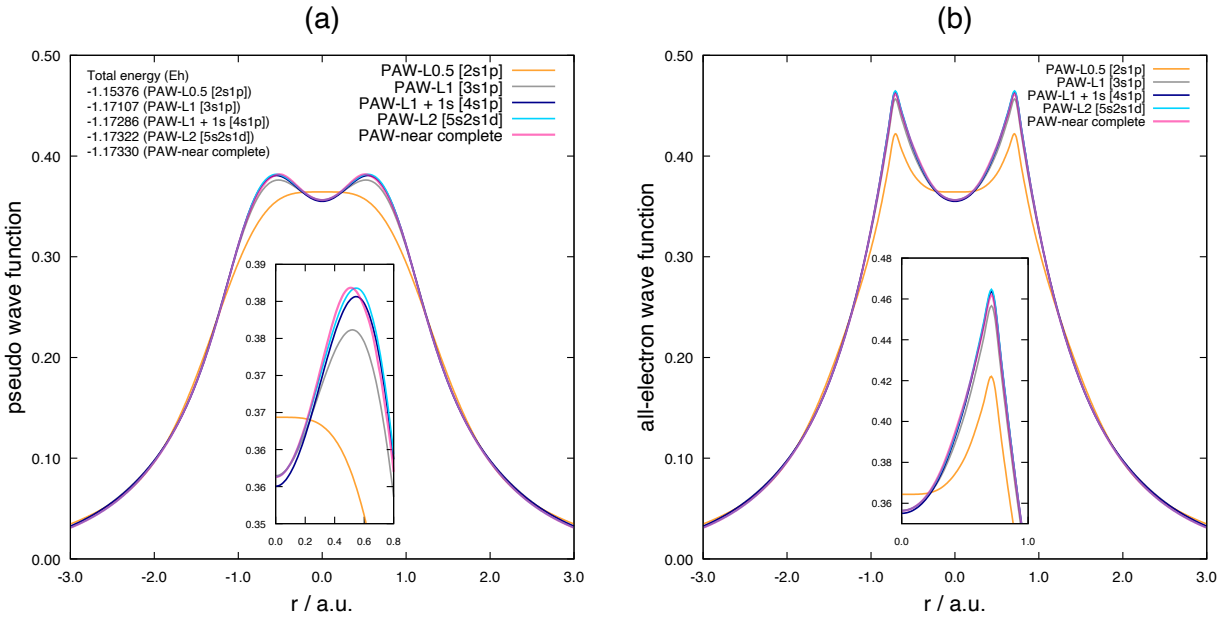

Figure S1: (a) Pseudo wave functions $\tilde{\Psi}(r)$ obtained with GTF-PAW method using uncontracted PAW-Ln basis for H_2 with d(H–H) = 0.76 Å. (b) The corresponding all-electron wave functions $\Psi(r) = \tilde{\Psi}(r) + \sum_a (\phi^a(r) - \tilde{\phi}^a(r))\langle p^a|\tilde{\Psi}\rangle$ where $\phi^a(r)$, $\tilde{\phi}^a(r)$, and $p^a(r)$ are partial all-electron waves, partial pseudo waves, and partial projectors, respectively.

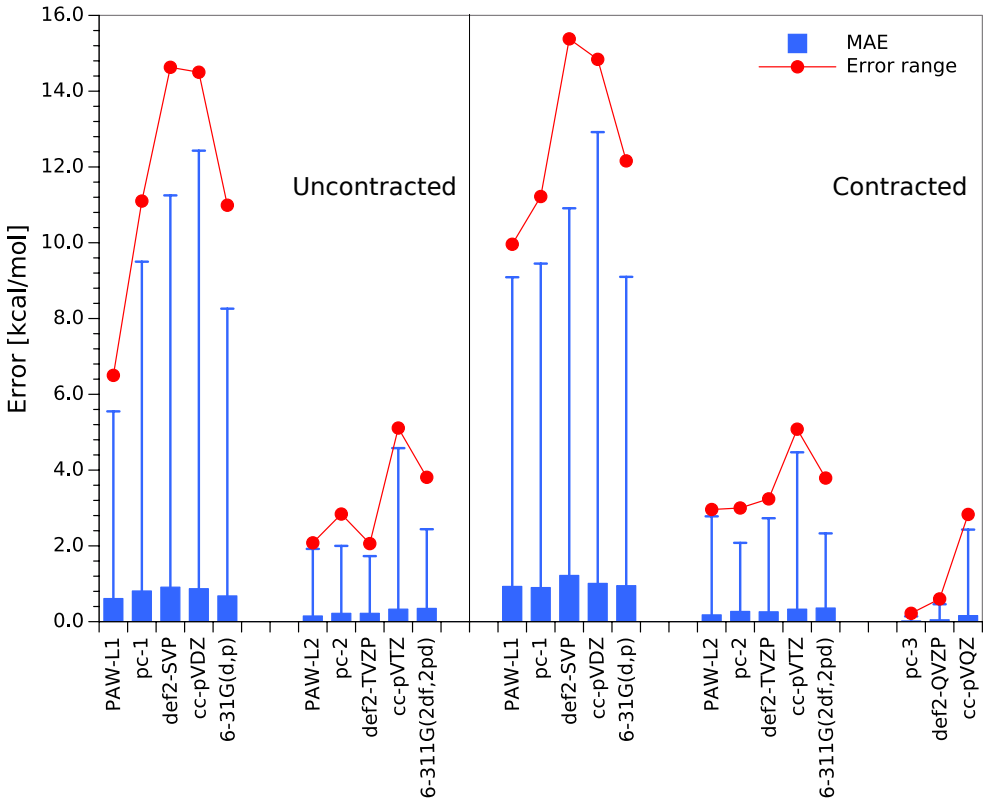

Figure S2: Mean absolute errors (blue bars), maximum absolute errors (vertical blue lines), and error ranges (red circles) of the atomization energies per atom calculated with the PAW-Ln and different all-electron basis sets and the SVWN exchange-correlation functional. The atomization energies calculated with pc-4 are used as references. The atomic energies were calculated as half the energies of the diatomic systems. Energies are in kcal/mol.

List of Tables

Table S1: Summary of the most important parameters used to construct the PAW atomic dataset.

Element	core electrons	valence electrons	# partial-waves	r_{paw} (Bohr)
H	_	$1s^1$	3	0.70
Li	_	$1s^2\ 2s^1$	5	1.60
Be	_	$1s^2\ 2s^2$	5	1.45
B	$1s^2$	$2s^2\ 2p^1$	4	1.20
C	$1s^2$	$2s^2\ 2p^2$	4	1.00
N	$1s^2$	$2s^2\ 2p^3$	4	1.00
O	$1s^2$	$2s^2\ 2p^4$	4	1.00
F	$1s^2$	$2s^2\ 2p^5$	4	1.00
Na	$1s^2$	$2s^2\ 2p^6\ 3s^1$	5	1.90
Mg	$1s^2$	$2s^2\ 2p^6\ 3s^2$	6	1.90
Al	$1s^2\ 2s^2\ 2p^6$	$3s^2\ 3p^1$	6	1.65
Si	$1s^2\ 2s^2\ 2p^6$	$3s^2\ 3p^2$	6	1.65
P	$1s^2\ 2s^2\ 2p^6$	$3s^2\ 3p^3$	6	1.65
S	$1s^2\ 2s^2\ 2p^6$	$3s^2\ 3p^4$	6	1.65
Cl	$1s^2\ 2s^2\ 2p^6$	$3s^2\ 3p^5$	6	1.65

Table S2: Systems used in the test set.

atomization energies and vertical ionization energies (G2-97′ dataset)[a]
1,3-butadiene, 2-butyne, 2-methyl-1-propene, acetaldehyde, acetamide, acetic acid, acetone, acetonitrile, acetylene, acetyl fluoride, acrylonitrile, allene, ammonia, aziridine, B_2, diborane(4), diborane(6), Be_2, BeC_2, $Be(CH_3)_2$, BeF, BeH_2, benzene, BeO, BeOH, BeH, BF, BH, BH_3, bicyclobutane, BN, BO, boranamine, borazine, boric acid, B_2O_3, C_2, $C_2H_3^{\cdot}$, $C_2H_5^{\cdot}$, CO_2, carbonic difluoride, CO, CF_4, $CCH^{\cdot}$, $CH^{\cdot}$, $CH_3CH_2O^{\cdot}$, $CH_3CO^{\cdot}$, $CH_3O^{\cdot}$, $CN^{\cdot}$, cyanogen, cyclobutane, cyclobutene, cyclopropane, cyclopropene, OF_2, difluoro methane, dimethylamine, dimethyl ether, ethane, ethanedial, ethanol, ethylamine, ethylene, ethylene oxide, F_2, fluoro ethene, formaldehyde, formic acid, furan, H_2, $H_2COH^{\cdot}$, $HCO^{\cdot}$, hydrazine, HCN, HF, H_2O_2, isobutane, isopropanol, isopropyl radical, ketene, Li_2, Li_2C_2, Li_2NH, Li_2O, Li_2O_2, $LiCH_3$, $LiNH_2$, LiOH, LiF, LiH, methane, methoxy ethane, methyl radical, methyl alcohol, methylamine, methylene singlet, methylene triplet, methylenecyclopropane, methyl formate, methyl nitrite, N_2, $NH^{\cdot}$, $NH_2^{\cdot}$, nitric oxide, nitrogen trifluoride, nitro methane, nitrous oxide, NO_2, O_2, $OH^{\cdot}$, ozone, propane, propene, propyne, pyridine, pyrrole, spiropentane, t-butyl radical, tetrafluoro ethylene, trans butane, trifluoro acetonitrile, trifluoro borane, trifluoro methane, trimethylamine, trimethylborane, water, 1-chloropropane, acetyl chloride, Al_2, Al_2H_6, Al_2O_3, $Al_3P_3H_6$, AlCl, AlH, AlH_2PH_2, AlH_3, $Al(PH_2)_3$, AlS, $Al(SiH_3)_3$, $AlCl_3$, AlF_3, $BeCl_2$, BeS, $Be(SH)_2$, $BeSi_2$, methylsulfonylmethane, vinylsilane, CS_2, CS, CCl_4, carbonyl sulfide, CF_3Cl, CH_3MgCl, CH_3PH_2, $CH_3S^{\cdot}$, ClF, ClO, ClF_3, chloroethene, chloroform, Cl_2, sulfuryl chloride, ClN, dimethyl sulfide, dimethyl sulfoxide, disilane, ethanethiol, chloroethane, thioformaldehyde, HCl, H_2S, hypochlorous acid, Li_2PH, Li_2S_2, Li_2Si_2, LiCl, $LiPH_2$, LiSH, methanethiol, methylchloride, dichloromethane, methylsilane, Mg_2, Mg_2H_4, MgC_2, $Mg(CH_3)_2$, MgCl, MgF_2, MgH, MgH_2, MgO, MgOH, MgS, MgSH, Na_2, Na_2C_2, Na_2NH, Na_2O, Na_2O_2, Na_2PH, Na_2S_2, Na_2Si_2, $NaCH_3$, NaCl, $NaNH_2$, NaOH, $NaPH_2$, NaSH, nitrosyl chloride, P_2, P_4, PCl_3, PCl_5, PF_5, PH, $PH_2^{\cdot}$, PH_3, PF_3, PN, PO, phosphoryl chloride, PS, S_2, S_2Cl_2, SCl, SCl_2, SF_6, $SH^{\cdot}$, Si_2, disilyne, disilene, $Si_2H_5^{\cdot}$, SiCl, $SiCl_2$, SiH, silane, SiO, SiF_4, $SiH_3^{\cdot}$, SiH_2 singlet, SiH_2 triplet, SiN, SiP, SiS, SN, SO_3, NaF, NaH, SO_2, SO, tetrachloroethylene, $SiCl_4$, thiirane, thiophene, BCl_3
noncovalent binding energies (S22 dataset)[b]
2-pyridoxine–2-aminopyridine, adenine–thymine stack, adenine–thymine Watson-Crick, ammonia dimer, benzene-methane, benzene–ammonia, benzene dimer parallel displaced, benzene dimer T-shaped, benzene–HCN, benzene–water, ethene dimer, ethene–ethyne, formamide dimer, formic acid dimer, indole–benzene stack, indole–benzene T-shape, methane dimer, phenol dimer, pyrazine dimer, uracil dimer H-bonded, uracil dimer stack, water dimer

[a]The structures of the neutral molecules were optimized. [b]Only single-point calculations.

Table S3: Mean and maximum absolute errors (in kcal/mol) of atomization energies per atom and vertical ionization energies; mean and maximum root-mean-square deviation (RMSD) of atomic positions (in Å), calculated with different basis sets and the SVWN exchange-correlation functional. The values calculated with pc-4 are used as references. MAE = mean absolute error. MaxAE = maximum absolute error.

	Atomization energies per atom		Atomization energies per atom[b]		Ionization energies[a]		RMSD	
Basis set	MAE	MaxAE	MAE	MaxAE	MAE	MaxAE	Mean	Max
Uncontracted basis sets								
PAW-L0.5	1.74	15.22	1.49	16.09	5.13	34.58	0.021	0.276
PAW-L1	1.64	8.37	0.61	5.55	1.49	7.76	0.014	0.307
pc-1	1.94	12.36	0.81	9.50	2.00	11.30	0.015	0.366
pc-1 X2C	2.05	12.72	0.84	9.77	1.91	11.39	0.015	0.366
def2-SVP	1.15	11.13	0.91	11.25	2.37	10.92	0.013	0.283
cc-pVDZ	1.69	13.81	0.87	12.43	2.13	10.94	0.014	0.123
6-31G(d,p)	1.24	9.65	0.68	8.26	2.78	12.22	0.013	0.338
PAW-L2	0.28	2.33	0.15	1.92	0.33	2.54	0.005	0.054
pc-2	0.47	3.02	0.22	2.00	0.51	2.92	0.005	0.147
def2-TZVP	0.32	1.85	0.22	1.73	0.47	3.44	0.006	0.241
cc-pVTZ	0.50	4.82	0.33	4.58	0.80	4.24	0.007	0.082
6-311G(2df,2pd)	0.43	2.40	0.35	2.44	1.21	5.58	0.005	0.086
Contracted basis sets								
PAW-L1	2.38	10.75	0.93	9.09	1.87	8.57	0.019	0.374
pc-1	2.33	12.55	0.90	9.45	2.12	11.02	0.018	0.325
def2-SVP	1.26	9.76	1.22	10.91	2.79	11.61	0.017	0.257
cc-pVDZ	2.15	13.60	1.01	12.92	3.53	13.60	0.018	0.145
6-31G(d,p)	1.44	9.72	0.95	9.10	4.26	15.22	0.015	0.215
PAW-L2	0.39	3.09	0.18	2.78	0.40	2.73	0.005	0.057
pc-2	0.47	3.07	0.27	2.08	0.54	3.07	0.006	0.202
def2-TZVP	0.37	2.90	0.26	2.73	0.54	3.35	0.006	0.242
cc-pVTZ	0.53	4.87	0.33	4.47	0.85	4.06	0.007	0.087
6-311G(2df,2pd)	0.43	2.33	0.36	2.33	1.23	5.64	0.006	0.055
pc-3	0.03	0.18	0.02	0.14	0.08	1.11	0.002	0.033
def2-QZVP	0.12	0.62	0.05	0.46	0.19	1.36	0.003	0.041
cc-pVQZ	0.24	2.57	0.16	2.43	0.35	2.02	0.004	0.041
ECP basis sets								
STRLC	9.37	50.97	3.21	34.21	7.02	34.50	0.064	0.485
SBK	6.19	35.78	1.71	28.50	4.03	17.19	0.041	0.484

[a]Calculations of boric acid, isobutane, and trifluoroacetonitrile with PAW-Ln basis sets did not converge due to electronic degeneracy. [b]The atomic energies were calculated as half the energies of the diatomic systems.

Table S4: Mean and maximum absolute errors (in kcal/mol) of atomization energies per atom and vertical ionization energies; mean and maximum root-mean-square deviation of atomic positions (in Å), calculated with different basis sets and the BLYP exchange-correlation functional. The values calculated with pc-4 are used as references. MAE = mean absolute error. MaxAE = maximum absolute error.

	Atomization energies per atom		Atomization energies per atom[b]		Ionization energies[a]		RMSD[c]	
Basis set	MAE	MaxAE	MAE	MaxAE	MAE	MaxAE	Mean	Max
Uncontracted basis sets								
PAW-L0.5	1.69	14.64	1.54	16.17	6.06	35.84	0.022	0.226
PAW-L1	1.47	7.36	0.53	4.75	1.61	8.68	0.013	0.182
pc-1	1.62	10.54	0.81	8.63	1.79	12.83	0.013	0.110
def2-SVP	0.94	9.10	0.81	10.07	2.85	12.96	0.012	0.132
cc-pVDZ	1.34	11.74	0.84	11.33	2.62	12.82	0.014	0.125
6-31G(d,p)	1.07	7.97	0.62	7.44	3.24	14.03	0.011	0.089
PAW-L2	0.48	1.88	0.19	1.53	0.71	3.18	0.005	0.052
pc-2	0.40	2.74	0.20	1.87	0.48	2.81	0.005	0.037
def2-TZVP	0.26	1.69	0.22	1.69	0.61	3.52	0.005	0.075
cc-pVTZ	0.42	3.99	0.30	4.00	1.03	4.74	0.007	0.053
6-311G(2df,2pd)	0.48	2.83	0.35	2.49	1.59	7.38	0.006	0.044
Contracted basis sets								
PAW-L1	2.23	8.72	0.88	7.97	2.04	9.39	0.016	0.106
pc-1	2.10	10.73	0.87	8.53	1.92	12.43	0.015	0.108
def2-SVP	1.08	8.25	1.14	9.85	3.31	13.54	0.015	0.244
cc-pVDZ	1.88	11.58	0.91	11.70	3.91	15.29	0.018	0.137
6-31G(d,p)	1.22	8.13	0.87	8.20	4.69	16.73	0.013	0.102
PAW-L2	0.57	2.51	0.22	2.09	0.73	3.40	0.005	0.052
pc-2	0.47	2.82	0.23	1.86	0.51	2.88	0.005	0.037
def2-TZVP	0.34	2.59	0.26	2.62	0.68	3.70	0.005	0.087
cc-pVTZ	0.43	4.18	0.30	3.97	1.06	4.70	0.007	0.054
6-311G(2df,2pd)	0.46	2.67	0.34	2.44	1.63	7.55	0.006	0.055
pc-3	0.04	0.19	0.02	0.13	0.07	0.80	0.003	0.040
def2-QZVP	0.11	0.53	0.05	0.41	0.22	2.20	0.003	0.047
cc-pVQZ	0.22	2.21	0.15	2.13	0.40	2.08	0.004	0.041
ECP basis sets								
STRLC	8.43	47.05	3.25	30.91	7.15	34.48	0.053	0.562
SBK	5.36	33.44	1.36	25.52	3.65	16.39	0.027	0.229

[a]Calculations of boric acid with PAW-Ln basis sets did not converge due to electronic degeneracy. [b]The atomic energies were calculated as half the energies of the diatomic systems. [c]The potential energy curve of Mg_2 is too shallow and the RMSD of Mg_2 was not included.

Table S5: Mean and maximum absolute errors (in kcal/mol) of the binding energies of S22 complexes calculated with different contracted basis sets. The values calculated with pc-4 are used as references. MAE = mean absolute error. MaxAE = maximum absolute error.

	SVWN		BLYP	
Basis set	MAE	MaxAE	MAE	MaxAE
PAW-L0.5	6.87	14.08	9.19	21.23
PAW-L1	1.47	3.49	1.39	2.68
pc-1	2.00	4.34	1.87	4.03
def2-SVP	2.34	6.07	2.83	6.73
cc-pVDZ	2.03	5.82	2.65	6.62
6-31G(d,p)	1.97	4.89	2.77	5.81
aug-pc-1	1.81	4.61	1.54	3.90
def2-SVPD	1.86	4.69	1.69	4.12
aug-cc-pVDZ	0.74	1.87	0.68	1.65
6-31++G(d,p)	0.81	1.95	0.66	1.35
PAW-L2	0.17	0.44	0.21	0.40
pc-2	0.34	0.95	0.29	0.64
def2-TZVP	0.46	1.06	0.53	1.18
cc-pVTZ	0.77	1.60	1.01	1.92
6-311G(2df,2pd)	1.63	3.24	1.49	3.82
aug-pc-2	0.27	0.74	0.09	0.30
def2-TZVPD	0.23	0.74	0.15	1.49
aug-cc-pVTZ	0.17	0.44	0.18	0.50
6-311++G(2df,2pd)	0.50	1.46	0.30	0.83
pc-3	0.02	0.05	0.02	0.05
def2-QZVP	0.12	0.33	0.16	0.40
cc-pVQZ	0.36	0.78	0.50	1.04
aug-pc-3	0.02	0.05	0.01	0.04
def2-QZVPD	0.04	0.13	0.05	0.15
aug-cc-pVQZ	0.09	0.31	0.10	0.34
aug-pc-4[a]	0.00	0.01	0.00	0.01
STRLC	2.18	6.29	2.69	7.17
SBK	3.10	7.08	3.44	7.78

[a]Due to SCF convergence problem related to aug-pc-4 basis set linear dependencies, only 18 and 15 complexes were calculated with SVWN and BLYP, respectively