

Accuracy of RPA for Noncovalent Interactions for Halogenated Molecules

Huynh Thanh, Nguyen Brian

October 6, 2020

Introduction

Exploration and investigation of natural products lay the foundation for the development of new medicine.¹ Upon traversing nature, the sea squirt was discovered to contain a potent cytotoxic chemical, called chlorolissoclimide, which can be useful in cancer therapeutics. It targets the production of proteins inhibiting the functions of ribosomes and induces apoptosis, or cell death.¹ X-ray crystal structure data reveals an interesting ribosome-drug interaction. When chlorolissoclimide interacts with the 80S ribosome, the chlorine appears to form a halogen- π bond with the guanine residue, G2793 and G2794, of the 25S rRNA.¹ Previous computational work revealed a significant 1.8 kcal/mol contribution to the overall interaction.¹ Halogens form halogen bonds by working as electron acceptors and can interact with π -systems to form halogen- π bonds. However, there has been little research conducted on this specific bond such as specificity, orientation, and direction. The understanding of these may help design new and effective cancer therapeutics.

To begin investigating this interaction, it was necessary to determine the available methods that can accurately describe the halogen- π bond. The random phase approximation (RPA), and the semilocal Perdew-Burke-Ernzerhof (PBE) functional with and without Grimme's dispersion corrections, denoted as D3 and D4, were chosen to be studied.²⁻⁴ RPA has been shown to accurately capture noncovalent interactions (NIs) and this study will expand on the performance of RPA for NIs involving halogenated molecules.⁵ Meanwhile, PBE, PBE-D3, and PBE-D4 were chosen to explore Grimme's known dispersion corrections for halogenated complexes.^{3,4} This will provide a comparison for the approximate treatment of NIs involving halogenated molecules between dispersion corrected semilocal density functional approximations (DFA) and RPA.

For this study, the X40 test set serves as a benchmark for these calculations and contains 40 complexes with different NIs involving halogenated compounds, see Fig. 1.⁶ Binding energies for these complexes range between -0.49 kcal/mol to -15.00 kcal/mol. Based on these calculations, a set of recommendations are provided to determine which method and basis sets are most appropriate for researching the halogen- π bond.

Methods

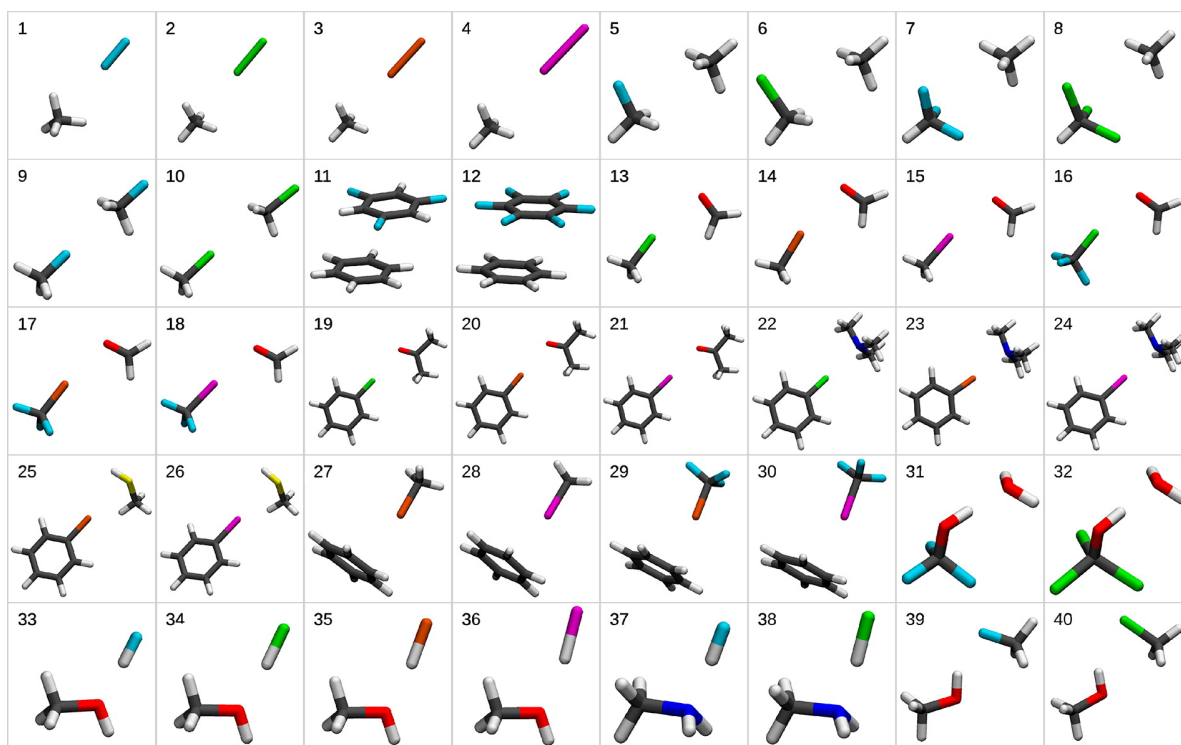


Figure 1. Noncovalent complexes containing halogens based on the X40 test set.⁶ Color scheme of the atoms is as follows F = Turquoise, Cl = Green, Br = Red, I = Purple, C = Black, and H = White.

The X40 Test Set was used to benchmark the methods' accuracy for noncovalent interactions involving halogens. Calculations were computed on greenplanet and the basis sets def2-QZVP, cc-pVTZ, and cc-pVQZ were used.⁷⁻⁹ Complexes containing iodine were run using the cc-pVTZ-PP, and cc-pVQZ-PP basis sets with the corresponding small-core (28-electron) quasirelativistic effective core potentials.^{10,11} The RPA energies were computed based on converged Kohn-Sham orbitals from PBE.² The resolution of identity (RI) was included to improve efficiency. Once all the calculations were run, the Hartree Exact Exchange (HXX) total energy, and RIRPA correlation energy were collected from the supermolecule and monomers. Due to the basis set superposition error (BSSE), counterpoise corrections (CP) were only applied to the RIRPA correlation energy. For semilocal DFAs with and without dispersion corrections,²⁻⁴ the interaction energies were computed with def2-QZVP basis sets and counterpoise corrected as recommended by Grimme.³

Binding energies (ΔE) were computed using the supramolecular approach where it is defined as $\Delta E = E_{ab} - E_a - E_b$. The variables E_{ab} , E_a , and E_b correspond to the total supermolecule energy, fragment A, and fragment B, respectively. For the RPA(PBE) interaction energies, these were extrapolated using the two-point $1/X^3$ extrapolation, where $X = 3$ (triple- ζ), $X = 4$ (quadruple- ζ), etc.^{12,13} The error was investigated by comparing the semilocal DFAs and RPA(PBE) to the coupled-cluster singles, doubles, and perturbative triples (CCSD(T))

reference.⁶ Throughout the report, signed errors are defined as differences between calculated and reference values; for example, a positive error in binding energies signifies underbinding.

RPA Basis Sets Convergence

Since the behavior for dispersion interactions of halogenated complexes is not known for RPA at the complete basis set limit (CBS),^{3,12,13} the choice of basis sets were also investigated for RPA(PBE) binding energies. The larger basis sets cc-pV5Z and cc-pV6Z⁷⁻⁹ were run for 11 complexes to explore the basis set convergence for these complexes which were chosen based on the different types of NIs and halogen atoms. Similarly, the HXX total energy and RIRPA correlation energy were recorded including counterpoise correction to generate a convergence plot of the interaction energy.

Additional calculations were run with different subvalence basis sets for the complexes containing Cl, Br, and I atoms. It has been reported before that subvalence basis sets were needed for well behaved convergences of dissociation energies.¹⁴ For iodine and bromine complexes, the basis set for the halogens were replaced with cc-pV(Q+2f)Z-PP.^{8,14} For chlorine complexes, the basis set was replaced with cc-pV(T+d)Z, cc-pV(Q+d)Z, cc-pV(5+d)Z, cc-pV(6+d)Z.^{8,14} For the subvalence basis sets, RI was not included for these calculations. A total of 9 complexes were investigated, each different halogens and interactions. Using the subvalence basis set data, a convergence plot was generated for each complex.

Results and Discussion

X40 Benchmark

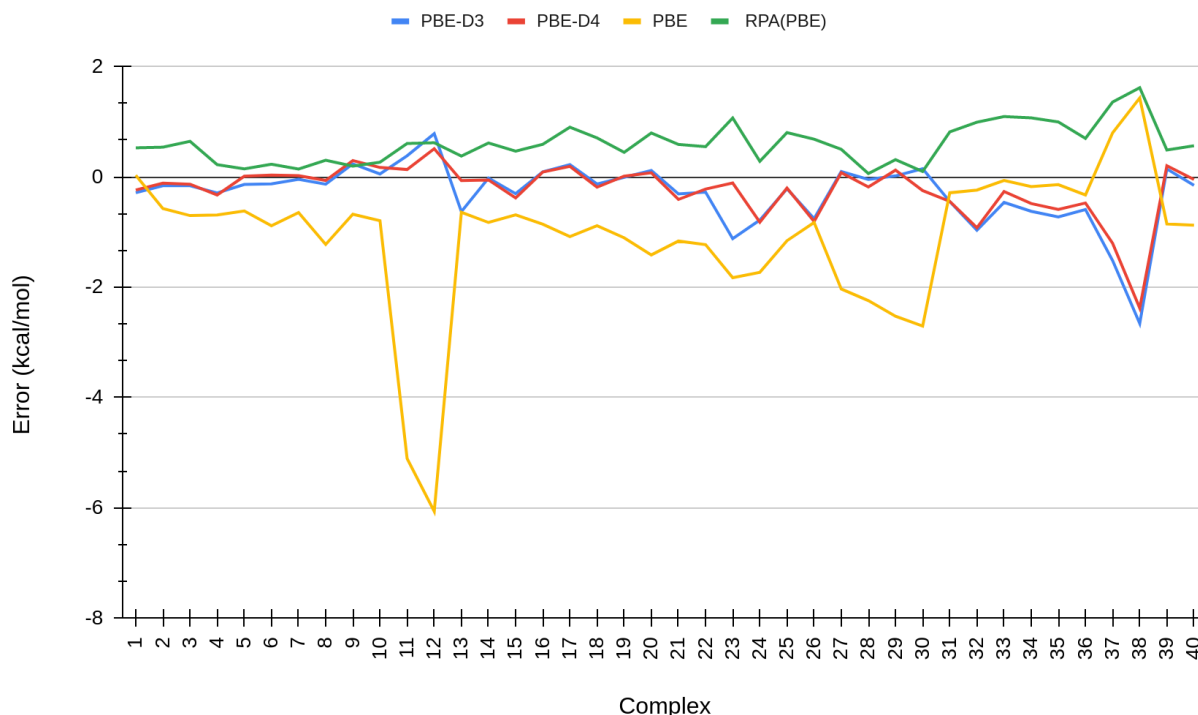


Figure 2. The binding energy errors (kcal/mol) for X40 test set computed for PBE, PBE-D3, PBE-D4, and RPA(PBE). All methods were 50% CP corrected and negative sign indicates overbinding.

From Fig. 2, the RPA(PBE) binding energy error tends to underbind consistent with previously reported NIs benchmarks.³ The mean RPA(PBE) error is 0.598 kcal/mol with a small standard deviation of 0.350 kcal/mol. In comparison, PBE significantly overbinds for most complexes in the X40 test set with a large average error of -1.037 kcal/mol and significant spread of 1.316 kcal/mol. Most noticeably, PBE significantly overbinds up to ~6 kcal/mol for complexes 11 and 12 which contain π - π stacking interactions. These structures are unique since the fluorinated benzene has the quadrupole moment reversed compared to benzene.⁶ Dispersion corrected PBE-D3 remedies this significant overbinding for complexes 11 and 12 with errors of 0.306 kcal/mol and 0.658 kcal/mol, respectively; PBE-D4 has binding errors of 0.115 kcal/mol and 0.442 kcal/mol for complexes 11 and 12, respectively. However, while the mean errors for PBE-D3 and PBE-D4 are -0.339 kcal/mol and -0.214 kcal/mol, respectively, the standard deviation is larger than the mean. Furthermore, dispersion corrected PBE does not systematically improve as seen by complexes 1, 30 to 38. These suggest some variability in the accuracy of the dispersion corrected semilocal DFAs. This observation is consistent with a previous study by Kouzuch and Martin showing that dispersion corrected DFAs do not sufficiently predict the geometries and dissociation energies of halogen bonds.¹⁵

Table 1. Statistical analysis of PBE, PBE-D3, PBE-D4, and 3-4 extrapolated RPA(PBE) for the X40 test set. The mean relative error (kcal/mol), mean absolute error (kcal/mol), mean percent error (%), and standard deviation (kcal/mol). All calculations were 50% CP corrected. Negative sign indicates overbinding.

	PBE/def2-QZVP (kcal/mol)	PBE-D3/def2-QZVP (kcal/mol)	PBE-D4/def2-QZVP (kcal/mol)	RPA(PBE)/CBS (kcal/mol)
Mean Error	-1.037	-0.339	-0.242	0.598
Mean Abs Error	1.171	0.431	0.332	0.598
Standard Deviation	1.316	0.576	0.496	0.350

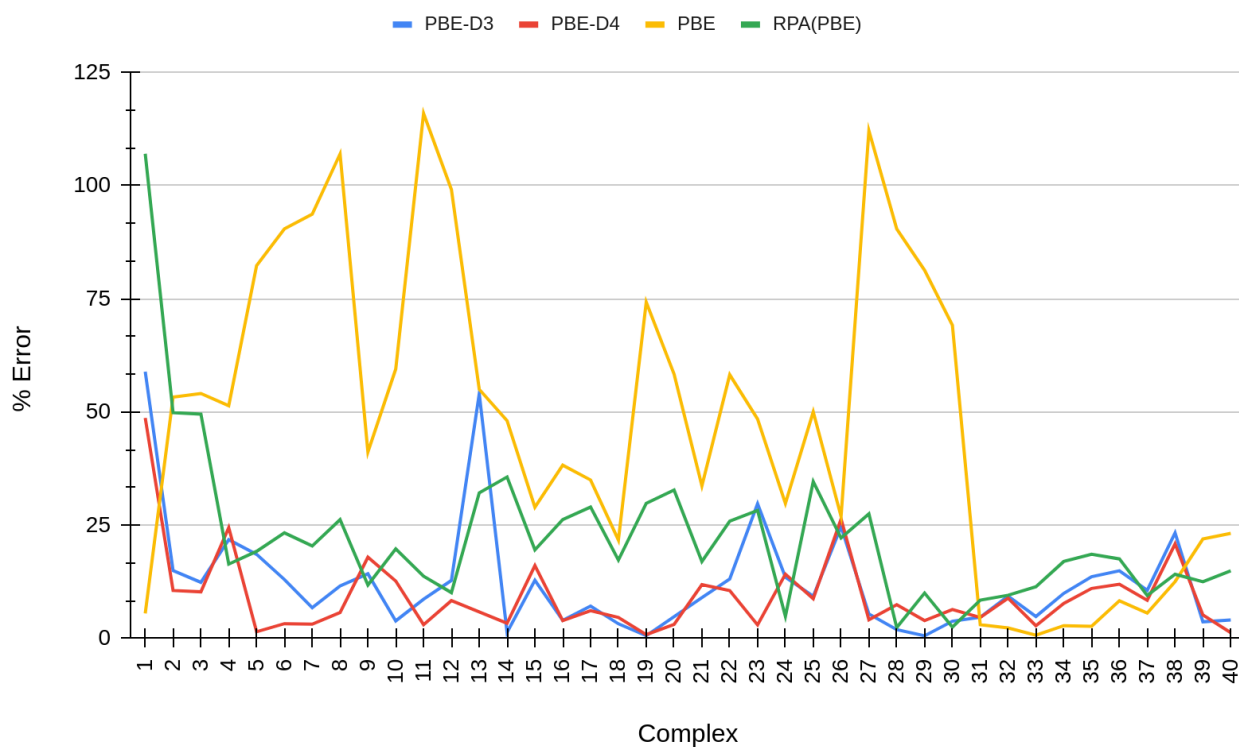


Figure 3. The binding energy percent error (%) for the X40 test set⁶ computed for PBE, PBE-D3, PBE-D4, and RPA(PBE). All methods were 50% CP corrected.

Percentage errors of the X40 test set presented a different story for the methods studied in this report. These errors were erratic for all methods. In particular, complex 1, which involves the interaction between methane and fluorine, suggested that PBE is the most accurate, see Figs. 2 and 3. PBE-D3 and PBE-D4 predict the binding energy to be -0.780 kcal/mol and -0.730 kcal/mol, respectively. These were an overestimation of around 50%. Meanwhile, RPA(PBE) suggests that methane and fluorine do not interact with a computed binding energy of 0.183

kcal/mol. Inspection of the X40 test set revealed that all complexes were geometry optimized with the second-order Møller–Plesset perturbation theory (MP2) using cc-pVTZ basis sets (cc-pVTZ-PP for Br and I) and including counterpoise correction.^{6,8} Geometries were approximated by an interpolation of the minimum based on the CCSD(T)/CBS dissociation curve along the intermolecular distance. The CCSD(T) binding energies based on MP2 optimized structures may not be as accurate since MP2 has been shown to poorly predict NIs of polarizable molecules.⁶

RPA Convergence of NIs for Halogenated Complexes

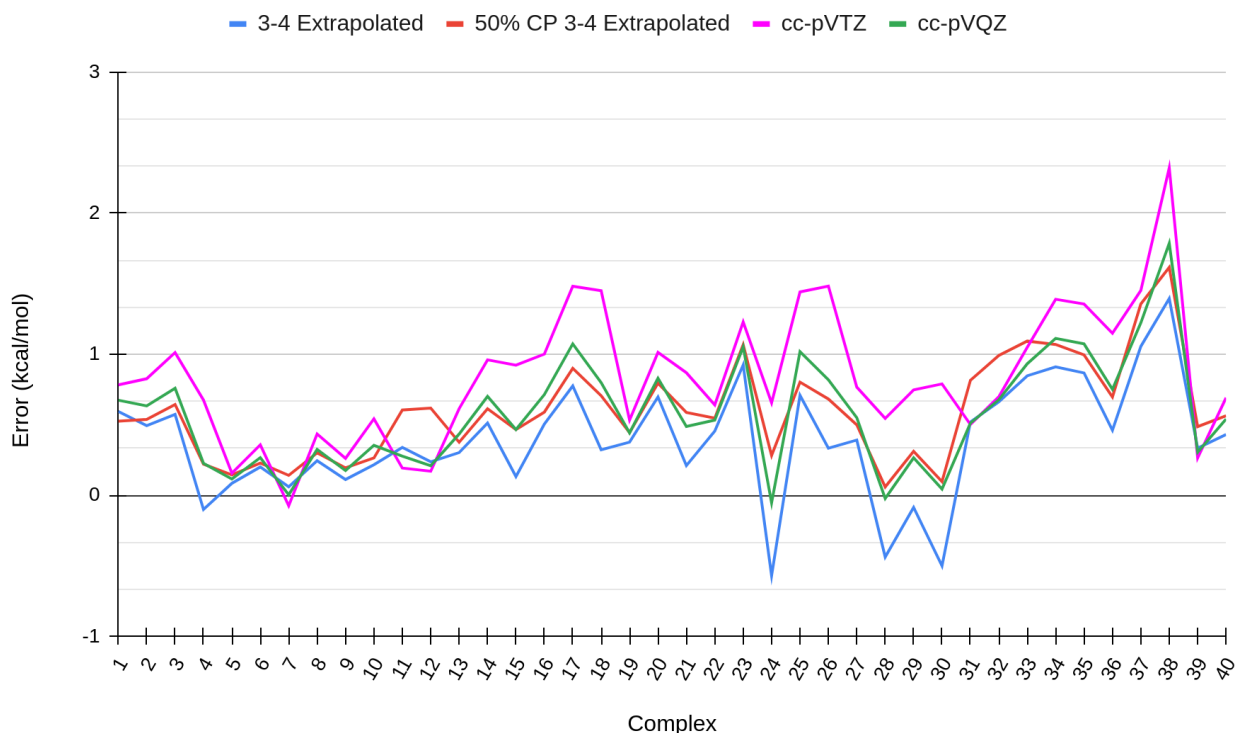


Figure 4. The RPA binding energy errors (kcal/mol) for the X40 test set⁶ computed using PBE reference orbitals with the Dunning’s basis sets,^{2,8} and 3-4 extrapolated energies with and without 50% CP correction. Positive sign indicates underbinding.

The unusually large percentage errors guided the study toward exploring the RPA(PBE) basis set convergence of NIs for halogenated complexes. From Fig. 4, the error for each complex ranges between -1 kcal/mol and 2.5 kcal/mol. With the addition of 50% counterpoise corrections, the error for most complexes is more positive than just RPA alone. This suggests significant BSSE where CP corrections may be up to 0.6 kcal/mol.

Basis Sets Convergence Plots

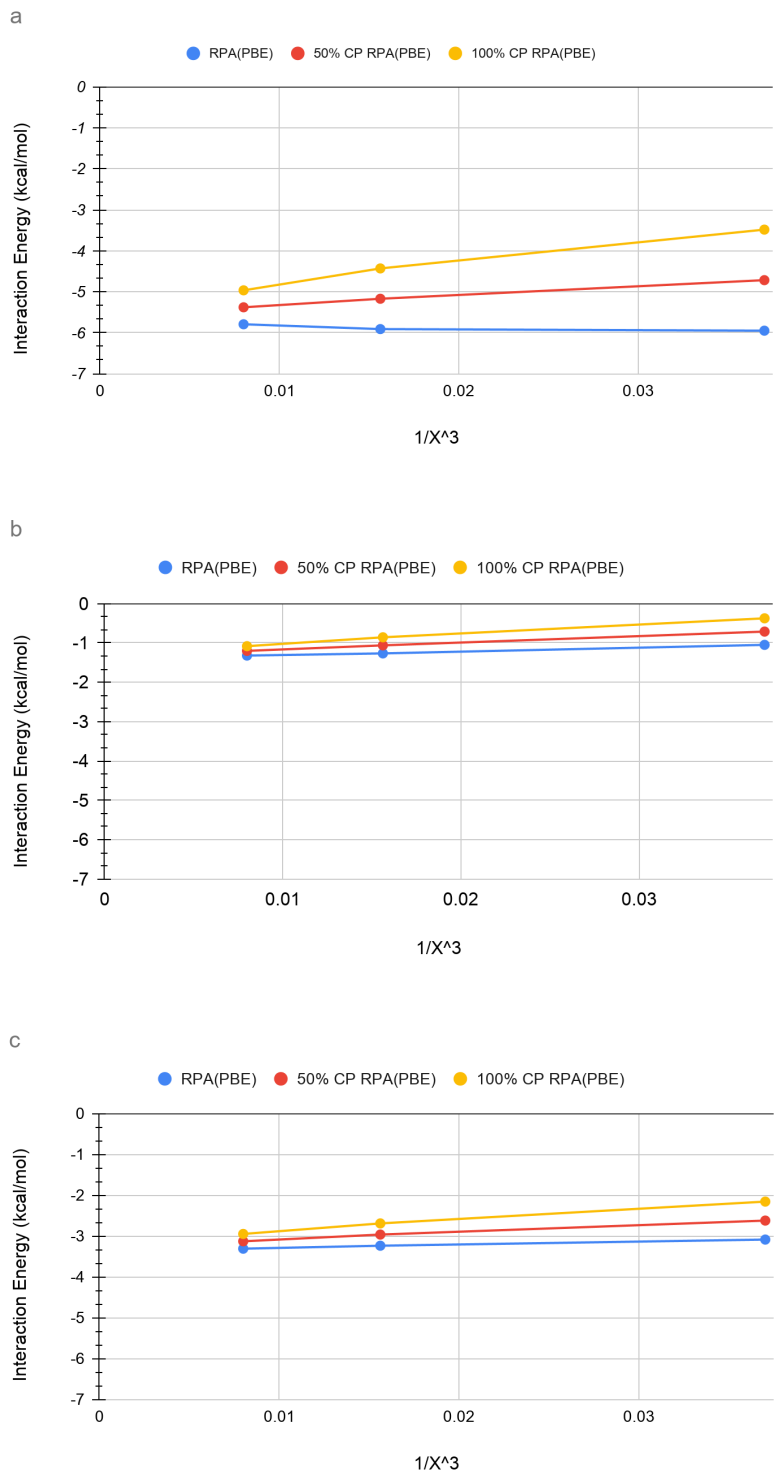


Figure 5. Complex basis set convergence plot computed using RPA(PBE) with 50% and 100% counterpoise corrections. Basis sets cc-pVTZ, cc-pVQZ, and cc-pV5Z were used for all atoms except iodine.⁸ The cc-pVTZ-PP, cc-pVQZ-PP, and cc-pV5Z-PP were used for iodine. $1/X^3$,

where X is the cardinal number, was used for extrapolation to form linear lines. **a.** Complex 12 contains stack interaction with fluorine. **b.** Complex 27 contains a halogen- π bond with bromine. **c.** Complex 40 contains hydrogen bond with chlorine.

Within the X40 test set, complexes 12, 27, and 40 were chosen to illustrate the good basis sets convergence for dispersion, halogen- π bond, and hydrogen bond, respectively. The convergence plot for Complex 12 shows that the 3-4 and 4-5 extrapolated converge to a similar interaction energy at CBS limit, see Fig. 5**a**. These extrapolated energies were within 0.40 kcal/mol. For complexes 27 and 40, the difference between the 3-4 and 4-5 extrapolated interaction energies was negligible (< 0.02 kcal/mol), see Fig. 5**b** and Fig. 5**c**. This suggests that the 3-4 extrapolation scheme with 50% CP correction is a good balance between efficiency and achieving the CBS limit for interaction energies.

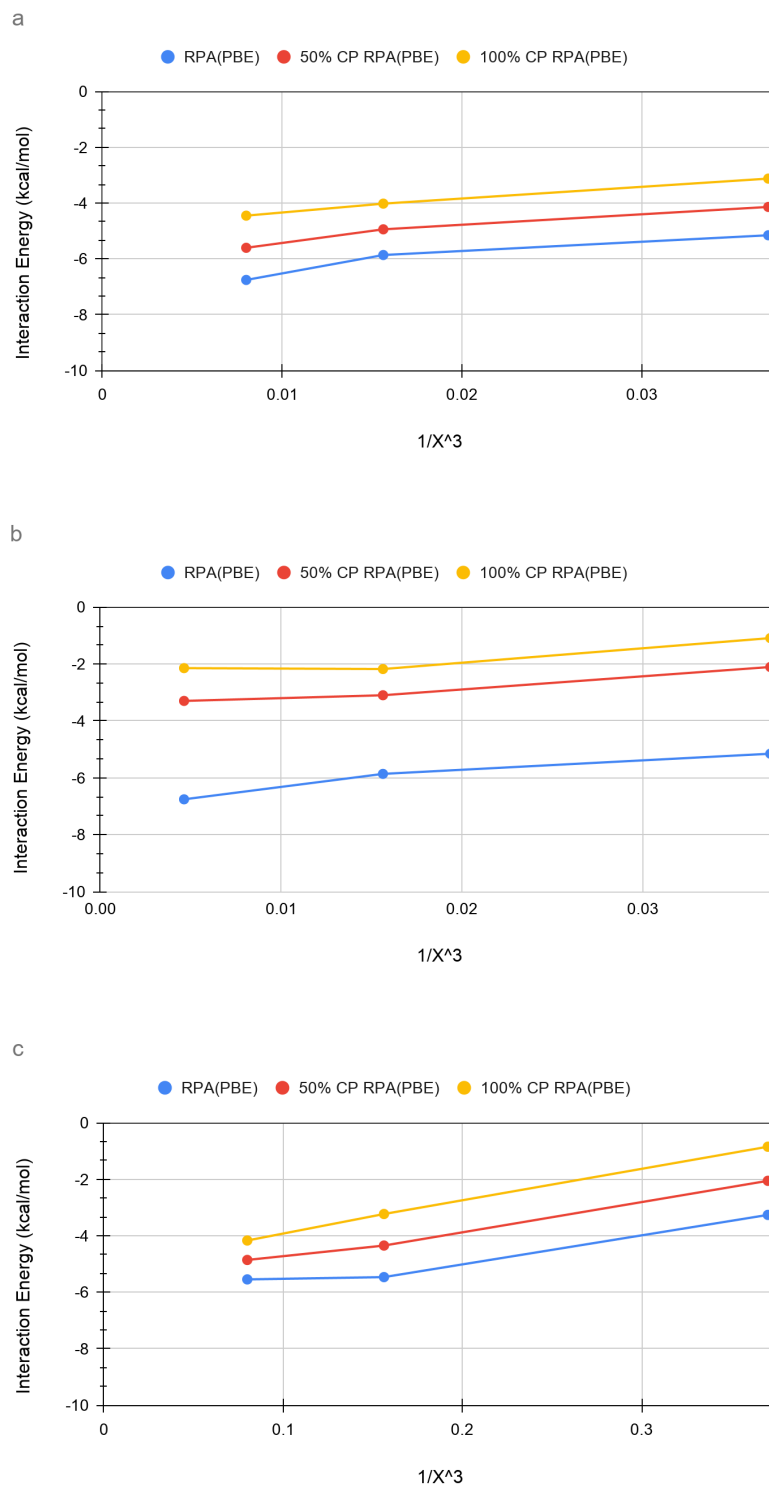


Figure 6. Complex 24 basis set convergence plot computed using RPA(PBE) with 50% and 100% counterpoise corrections. $1/X^3$, where X is the cardinal number, was used for extrapolation to form linear lines. This complex contains halogen bonds with iodine. **a.** For all atoms except for iodine, basis sets cc-pVTZ, cc-pVQZ, and cc-pV5Z were used. Basis sets cc-pVTZ-PP,

cc-pVQZ-PP, and cc-pV5Z-PP were used for iodine. **b.** The same basis sets were used as in part **a.** except the subvalence basis sets cc-pV(Q+2f)Z-PP for iodine is introduced.⁸ **c.** Basis sets def2-SVP, def2-TZVP, and def2-QZVP were used for all atoms.

Meanwhile, heavier and polarizable halogen atoms such as iodine present a tougher basis set convergence for interaction energies. Complex 24 from Fig. 6a and 6b appears to show diverging lines. The extrapolation line from cc-pVTZ to cc-pVQZ for RPA, 50% counterpoise corrections (CP), and 100% counterpoise corrections appear to be parallel. From cc-pVQZ to cc-pV5Z, the lines are no longer parallel and instead extend into different directions. For basis sets cc-pVTZ-PP, cc-pVQZ-PP, and cc-pV5Z-PP, the extrapolated lines also start off parallel then diverge. The convergence plot for basis sets def2-SVP, def2-TZVP, and def2-QZVP did not diverge. The extrapolation from def2-SVP to def2-TZVP does appear parallel, however, it can be noticed that the extrapolations converge from def2-TZVP to def2-QZVP.

Subvalence Dunning's Basis Sets

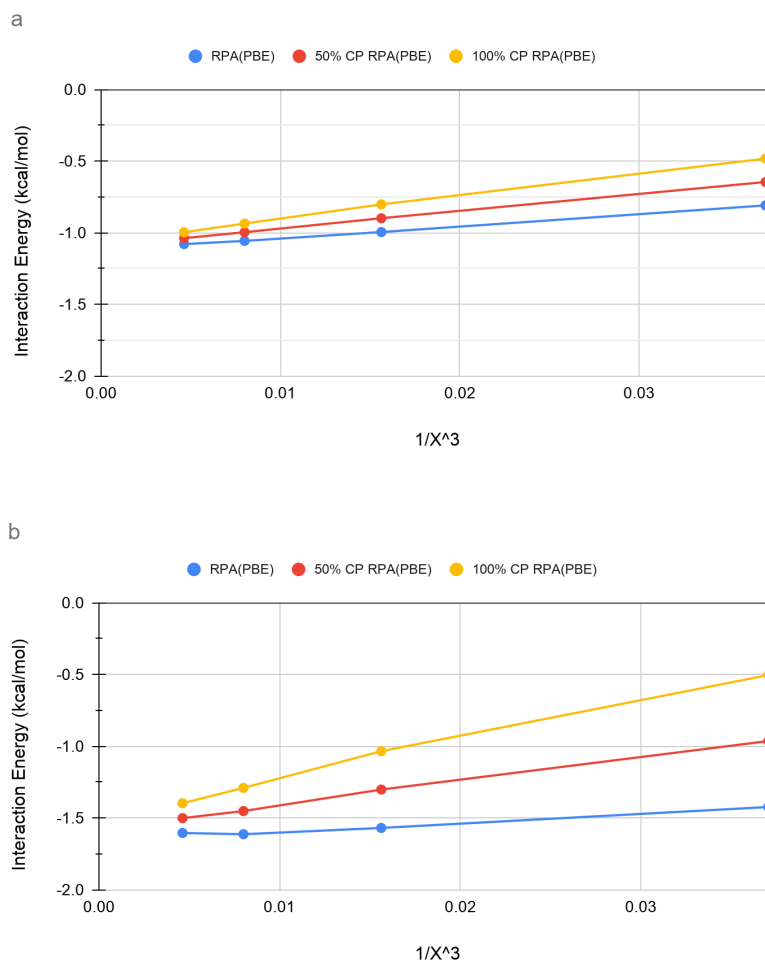


Figure 7. Complex basis set convergence plot computed using RPA(PBE) with 50% and 100% counterpoise corrections. Basis sets cc-pV(T+d)Z, cc-pV(Q+d)Z, cc-pV(5+d)Z, cc-pV(6+d)Z were used.^{8,14} $1/X^3$, where X is the cardinal number, was used for extrapolation to form linear lines. **a.** Complex 10 contains dipole-dipole interactions with chlorine. **b.** Complex 40 contains hydrogen bonding with chlorine.

The convergence plots with subvalence basis sets for complex 10 from Fig. 7a shows that 3-4, 4-5, and 5-6 extrapolations with and without CP all converge to a similar interaction energy at CBS limit. This interaction energy converged around -1.1 kcal/mol. The lines for the convergence plot for complex 40 in Fig. 7b do not appear parallel with each other and converges as the lines approach 0. The 3-4, 4-5, and 5-6 extrapolations for complex 40 also yield similar interaction energies converging to -1.6 kcal/mol.

Conclusions

The CBS convergence plots computed using RPA(PBE) reveal that the most adequate method of calculating binding energy for the different interactions and halogens represented in the X40 Test Set is the 3-4 extrapolation scheme with 50% counterpoise corrections. Comparing the 3-4, 4-5, and 5-6 extrapolations for RPA, the binding energy values obtained do differ by around 0.1 kcal/mol. It would be cost efficient to use the 3-4 extrapolations as it yields nearly the same results as 4-5 and 5-6 extrapolations. It would be beneficial to use RPA(PBE) because this method produces results with little error and without needing corrections. PBE methods with dispersion corrections produce better results since the mean absolute error for PBE-D3 is 0.431 kcal/mol and PBE-D4 is 0.332 kcal/mol however, they do not systematically improve calculations like RPA(PBE). For some complexes, PBE with no corrections performs better than PBE-D3 and PBE-D4, which isn't consistent. RPA(PBE) performs well as the average error shown in Table 1. was 0.598 kcal/mol and the average percent error is 22.45%. RPA(PBE) is parameters free making it a good choice to study halogen bonds.

Complexes containing dispersion, stacking, and hydrogen bonding were found to have converged better than other interactions. Fluorine, chlorine, and bromine were also found to converge very well compared to iodine in the test set. Slightly better convergences were observed for cc-pV(T+d)Z, cc-pV(Q+d)Z, cc-pV(5+d)Z, and cc-pV(6+d)Z basis sets. The complexes with iodine, however, did not converge when the basis set cc-pVTZ-PP, cc-pVQZ-PP, and cc-pV5Z-PP were used in the calculations. Better results can be obtained by using the basis sets def2-SVP, def2-TZVP, and def2-QZVP for complexes containing iodine. The recommendation is to use subvalence basis sets to study the halogen bonds and further investigations are needed for the basis sets errors with iodine.

Previously studied for RPA: X40 and XB51 test sets

<https://digitalcommons.montclair.edu/sigma-xi/2019/poster-1/57/>

Bibliography

- 1) Könst, Z. A.; Szklarski, A. R.; Pellegrino, S.; Michalak, S. E.; Meyer, M.; Zanette, C.; Cencic, R.; Nam, S.; Voora, V. K.; Horne, D. A.; Pelletier, J.; Mobley, D. L.; Yusupova, G.; Yusupov, M.; Vanderwal, C. D. Synthesis Facilitates an Understanding of the Structural Basis for Translation Inhibition by the Lissoclimides. *Nat. Chem.* **2017**, 9 (11), 1140–1149.
- 2) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, 77 (18), 3865–3868.

- 3) Grimme, S. Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory. *Eur. J. Chem.* **2012**, *18* (32), 9955–9964.
- 4) Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. A Generally Applicable Atomic-Charge Dependent London Dispersion Correction. *J. Chem. Phys.* **2019**, *150* (15), 154122.
- 5) Nguyen, B. D.; Chen, G. P.; Agee, M. M.; Burow, A. M.; Tang, M. P.; Furche, F. Divergence of Many-Body Perturbation Theory for Noncovalent Interactions of Large Molecules. *J. Chem. Theory Comput.* **2020**, *16* (4), 2258–2273.
- 6) Řezáč, J.; Riley, K. E.; Hobza, P. Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules. *J. Chem. Theory Comput.* **2012**, *8* (11), 4285–4292.
- 7) 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* (18), 3297.
- 8) Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90* (2), 1007–1023.
- 9) Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98* (2), 1358–1371 DOI: 10.1063/1.464303.
- 10) Peterson, K. A.; Figgen, D.; Goll, E.; Stoll, H.; Dolg, M. Systematically Convergent Basis Sets with Relativistic Pseudopotentials. II. Small-Core Pseudopotentials and Correlation Consistent Basis Sets for the Post-d Group 16–18 Elements. *J. Chem. Phys.* **2003**, *119* (21), 11113–11123.
- 11) Peterson, K. A.; Shepler, B. C.; Figgen, D.; Stoll, H. On the Spectroscopic and Thermochemical Properties of ClO, BrO, IO, and Their Anions. *J. Chem. Phys. A* **2006**, *110* (51), 13877–13883.
- 12) Eshuis, H.; Furche, F. Basis Set Convergence of Molecular Correlation Energy Differences within the Random Phase Approximation. *J. Chem. Phys.* **2012**, *136* (8), 084105.
- 13) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N₂, and H₂O. *Chem. Phys. Lett.* **1998**, *286* (3-4), 243–252 DOI: 10.1016/s0009-2614(98)00111-0.
- 14) Dunning, T. H.; Peterson, K. A.; Wilson, A. K. Gaussian Basis Sets for Use in Correlated Molecular Calculations. X. The Atoms Aluminum through Argon Revisited. *J. Chem. Phys.* **2001**, *114* (21), 9244–9253.
- 15) Kozuch, S.; Martin, J. M. L. Halogen Bonds: Benchmarks and Theoretical Analysis. *J. Chem. Theory Comput.* **2013**, *9* (4), 1918–1931.