Article

A Minimum Quantum Chemistry CCSD(T)/CBS Data Set of Dimeric Interaction Energies for Small Organic Functional Groups: Heterodimers

Hsing-Hsiang Huang, Yi-Siang Wang, and Sheng D. Chao*

Read Online

ACCESS | 📊 Metrics & More | 📑 Article Recommendations | 🔲 Supporting Information

ABSTRACT: We extend our previous quantum chemistry calculations of interaction energies for 31 homodimers of small organic functional groups (the SOFG-31 data set) by including 239 heterodimers with monomers selected within the SOFG-31 data set, thus resulting in the SOFG-31+239 data set. The minimum-level theoretical scheme contains (1) the basis set superposition error corrected supermolecule (BSSE-SM) approach for intermolecular interactions; (2) the second-order Møller−Plesset perturbation theory (MP2) with the Dunning's aug-cc-pVXZ (X = D, T, Q) basis sets for the geometry optimization and correlation energy calculations; and (3) the single-point energy calculations with the coupled cluster with single, double, and perturbative triple excitations method at the complete basis set limit [CCSD(T)/CBS] using the well-tested extrapolation methods for the MP2 energy calibrations. In addition, we have performed a parallel series of energy decomposition calculations based on the symmetry adapted perturbation theory (SAPT) in order to gain chemical insights. That the above procedure cannot be further reduced has been proven to be very crucial for constructing reliable data sets of interaction energies. The calculated CCSD(T)/CBS interaction energy data can serve as a benchmark for testing or training less accurate but more efficient calculation methods, such as the electronic density functional theory. As an application, we employ a segmental SAPT model previously developed for the SOFG-31 data set to predict binding energies of large heterodimer complexes. These model energy "quanta" can be used in coarse-grained molecular dynamics simulations by avoiding large-scale calculations.
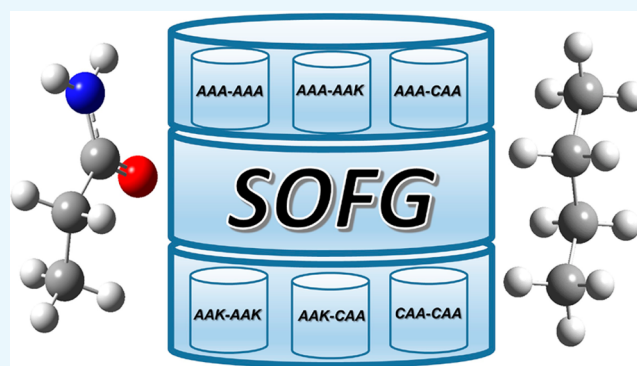
1. INTRODUCTION

Molecular modeling of complex materials has been a very useful tool of computational chemistry in gaining better understanding of intricate experimental observations. At the atomic level, the techniques are mainly concerned with developing classical force fields to model both chemical (covalent) bonding and intermolecular or noncovalent interactions. Traditional empirical force fields (EFFs)[1−5] have long been used in molecular mechanics, Monte Carlo simulations, and molecular dynamics simulations. Because many popular EFFs utilize extensive experimental data in their model constructions, the efficacy in reproducing experiments deteriorates very quickly once the models are used outside the original training sets. More fundamental chemical models (usually called ab initio force fields (AIFFs), to distinguish them from EFFs) are mainly based on quantum chemistry calculations with, hopefully, minimum inputs from experiments.[6−18] Most current generation force fields have employed various levels of potential energy data from electronic structure calculations, which are usually collected in the form of numerical data sets. These interaction energy data sets not only are useful for designing universal force fields but

also serve as a benchmark for testing and/or training lower-level but more computationally efficient calculation methods, such as the electronic density functional theory (DFT).[19−24] Therefore, it is a continuing effort to develop comprehensive data sets of accurate intermolecular interaction energies based on high-level quantum chemistry calculations.[25−32]

A reliable quantum chemistry calculation for interaction energies requires a size-consistent correlation method and a sizable basis set for error tolerance. An improper combination of method and basis set would render misleading, if not false, conclusions, making the human efforts wasteful and the calculated data futile. The issue of choosing a proper combination of a correlation method and a basis set has been carefully examined by previous database constructions, notably those by
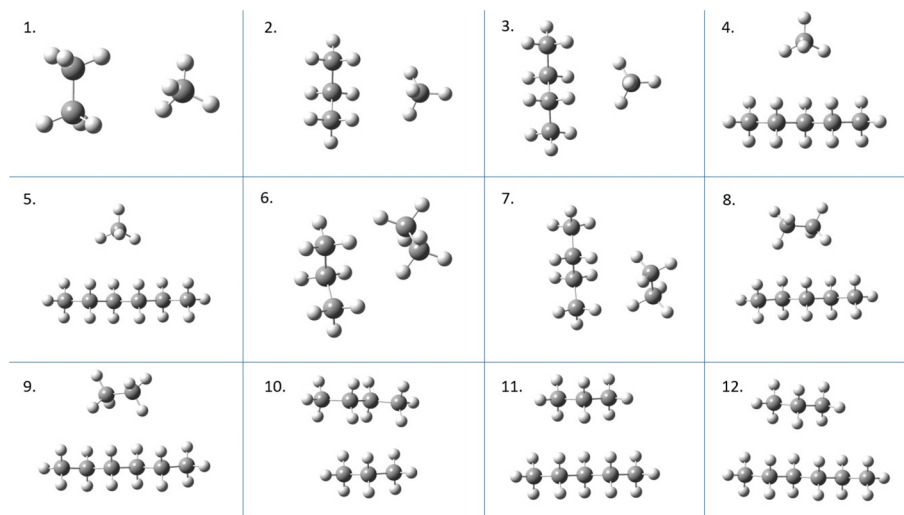
Figure 1. Optimized structures of the dimers in the Aa−Aa series.

the Hobza group, the Sherrill group, and the Grimme group, independently and respectively.[33−35] Thanks to these strenuous efforts, a consensus has been reached among active researchers in determining a minimum level theoretical scheme for the calculated interaction energy to bear a "sub-chemical accuracy" (ca. 0.1 kcal/mol).[36−38] It contains (1) the basis set super-position error[39,40] corrected supermolecule (BSSE-SM) approach[41−43] for intermolecular interactions; (2) the second-order Møller−Plesset perturbation theory (MP2)[44] with the Dunning's aug-cc-pVXZ (X = D, T, Q) basis sets[45] for the geometry optimization and correlation energy calculations; and (3) the single-point energy calculations with the coupled cluster with single, double, and perturbative triple excitations method at the complete basis set limit [CCSD(T)/CBS] using the well-tested extrapolation methods for the MP2 energy calibrations.[46,47] Complementary energy dissection methods, such as the symmetry adapted perturbation theory (SAPT), are often required in order to gain physical understanding of the calculated interaction energy.

One of the earlier efforts of collecting the benchmark interaction energy data into well-edited data sets was attributed to the Hobza group.[48] For example, the S22 data set[49] and its subsequent refinements[37,50] have served as a paradigm of first initiating a data set at a minimum theoretical level and subsequently extending the original scope. Indeed, because of their feeble magnitudes, the calculation of accurate interaction energies is a daunting task. For large noncovalent bounded systems, the above standard procedure is usually not feasible because of the enormous increase of the computational cost. Currently, for small complexes with less than 50 atoms, this line of practice is continued and being gradually revised.[51,52] For example, the Řezač group has recently launched the ATLAS project.[53] More comprehensive "super" data sets are also collected and maintained, notably by the Head-Gordon group,[22,54] the Grimme group,[55,56] the Shaw group,[57] and the QCArchive database.[58]

In a previous study, we constructed an interaction energy data set for the homodimers of 31 small organic functional groups (the SOFG-31 data set).[59] The SOFG-31 data set is a minimum CCSD(T)/CBS data set in the sense that these energies are calculated at the minimum-level theoretical scheme described above. In this paper, we extend the study to the heterodimers with the dimeric pair monomers selected from the SOFG-31

Table 1. Binding Energies of the Dimers in the Aa−Aa Series

	MP2 [CCSD(T)]			CCSD(T)
	aDZ	aTZ	aQZ	CBS
methane−ethane	−0.644	−0.746	−0.773	−0.825
	[−0.670]	[−0.791]	[−0.800]	(−0.827)[a]
methane−propane	−0.880	−1.002	−1.036	−1.092
	[−0.903]	[−1.033]		
methane−butane	−1.050	−1.196		−1.277
	[−1.067]	[−1.216]		
methane−pentane	−1.188	−1.333		−1.398
	[−1.192]			
methane−hexane	−1.260	−1.407		−1.466
	[−1.257]			
ethane−propane	−1.408	−1.616	−1.673	−1.716
	[−1.392]	[−1.617]		
ethane−butane	−1.598	−1.807		−1.881
	[−1.573]	[−1.793]		
ethane−pentane	−1.794	−2.020		−2.079
	[−1.758]			
ethane−hexane	−1.955	−2.187		−2.228
	[−1.898]			
propane−butane	−2.074	−2.324		−2.365
	[−2.010]			
propane−pentane	−2.337	−2.607		−2.651
	[−2.267]			
propane−hexane	−2.556	−2.839		−2.868
	[−2.466]			

[a]The A24 dataset[52].

data set. Because there are 239 (out of 465) heterodimers considered in this work, the resulting data set is called the SOFG-31+239 data set. The other part of this paper is organized into the following sections. In section 2 we briefly describe the theoretical considerations and computational details. Our main results and discussions are shown in section 3. We conclude this work in section 4, and numerical data of reference value are available in the Supporting Information.

2. QUANTUM CHEMISTRY CALCULATIONS

The theoretical scheme is similar to that used in the construction of the SOFG-31 data set.[59] Briefly, the basis set superposition
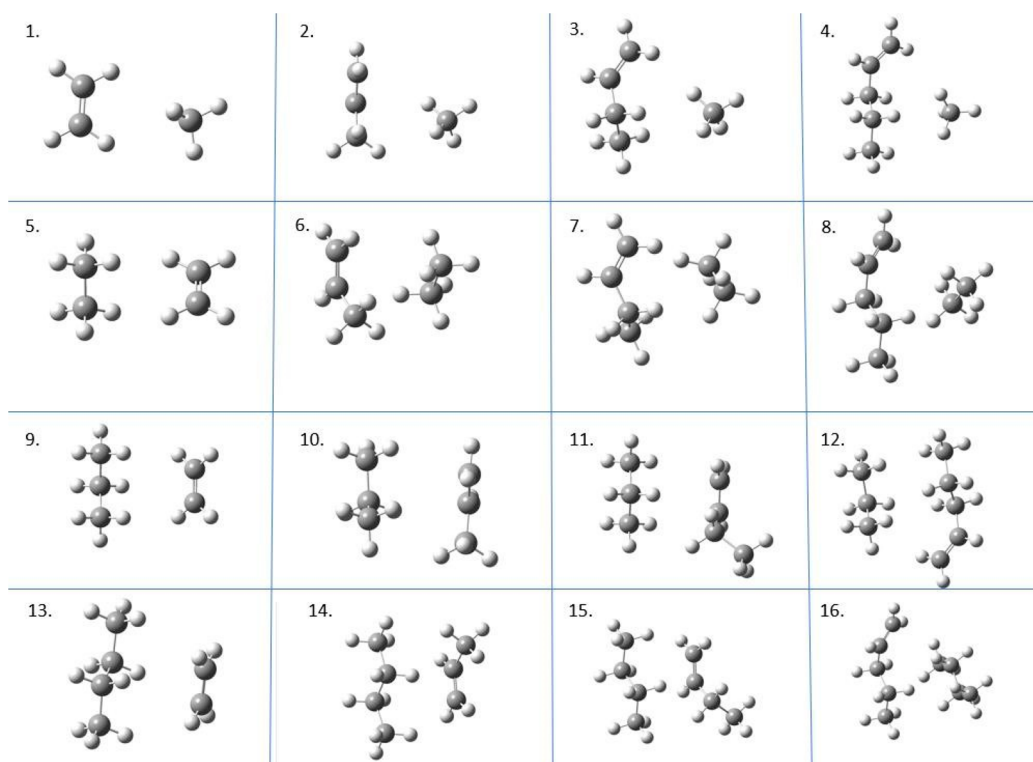
Figure 2. Optimized structures of the dimers in the Aa−Ae series.

error corrected supermolecule (BSSE-SM) approach was employed for calculating the interaction energies. The second-order Møller−Plesset perturbation theory (MP2) with the Dunning's aug-cc-pVXZ (X = D, T, Q) basis sets has been employed in the geometry optimization and energy calculations. The MP2 calculated energies have been calibrated by using the coupled cluster with single, double, and perturbative triple excitations method at the complete basis set limit [CCSD(T)/ CBS]. All the molecular orbital calculations and the Berny geometry optimization tasks were performed using the Gaussian 09 suite of programs.[60] No symmetry or rigid molecule constraints were imposed in the geometry optimization calculations. The normal-mode frequency analysis has been performed, and the found equilibrium complexes were carefully checked to ensure that all the obtained configurations are true energy minima on the respective potential energy surfaces.

For benchmark data calibrations, the CCSD(T)/CBS energy is the well-recognized "gold standard". However, directly calculating the CCSD(T) energies at increasingly large basis sets is very computationally intense work. It is more feasible to first optimize the dimer structure using the MP2 method at a series of good-quality basis functions (such as Dunning's) and then use the well-tested extrapolation methods to obtain the CCSD(T)/CBS values. There are two standard ways for obtaining the complete basis set limit values. The first method of Helgaker et al.[47] is based on the theoretically justified power-law dependence of the energy on the aug-cc-pVXZ (X = D, T, Q, etc.) basis set. Using the calculated data at two basis functions of different X's, one can extrapolate the energy to the CBS value as X approaches infinity. On the other hand, the focal-point extrapolation method[61] is used to estimate the CBS value by considering the difference between the CCSD(T) and the MP2 interaction energies calculated at the same (smaller) basis set. It is assumed that although the absolute values of interaction energy converge very slowly, the difference between the values calculated by the two correlation methods is negligibly dependent on the basis set size, as long as a minimum basis function is used. This assumption has been thoroughly tested in the previous database constructions and is known to be reliable for a variety of noncovalently bonded complexes. In this work, the MP2/CBS binding energies were obtained from the extrapolation method of Helgaker et al.[47] with Dunning's correlation consistent basis set (aug-cc-pVXZ, X = D, T, and up to Q). The CCSD(T)/CBS binding energies were obtained using the focal point extrapolation method.[61]

The calculated interaction energies are further analyzed by the symmetry-adapted perturbation theory (SAPT0) with the jun-cc-pVXZ (X = D, T) basis set[62] as implemented in the PSI4 program.[63] The model segmental SAPT analysis was discussed in our previous paper and is illustrated and used here as an application of the SAPT data.[59]

3. RESULTS AND DISCUSSION

The SOFG-31 data set contains 31 homodimers with monomers distributed in three subsets. The alkane−alkene−alkyne (AAA) subset contains 6 alkanes (methane to hexane), 4 alkenes (ethene to 1-pentene), and 4 alkynes (ethyne to 1-pentyne). The alcohol−aldehyde−ketone (AAK) subset includes 4 alcohols (methanol to 1-butanol), 4 aldehydes (formaldehyde to butanal), and 3 ketones (acetone to 2-pentanone). The carboxylic acid−amide (CAA) subset consists of 3 carboxylic acids (formic acid to propanoic acid) and 3 amides (formamide to propanamide). With the intended heterodimers in mind, we consider 239 cross-group combinations with the pair monomers selected from respective subgroups. More specifically, we classify the binary complexes according to the following subsets. The AAA−AAA set contains 12 alkane−alkane (Aa−Aa), 16 alkane−alkene (Aa−Ae), and 6 alkene−alkene (Ae−Ae)

Table 2. Binding Energies of the Dimers in the Aa−Ae Series

	MP2 [CCSD(T)]			CCSD(T)
	aDZ	aTZ	aQZ	CBS
methane−ethylene	−0.702	−0.827	−0.863	−0.906
	[−0.706]	[−0.842]	[−0.880]	
methane−propylene	−0.920	−1.041	−1.081	−1.057
	[−0.870]	[−0.988]		
methane−butylene	−1.016	−1.168		−1.230
	[−1.012]	[−1.166]		
methane−pentylene	−1.253	−1.422		−1.489
	[−1.249]			
ethane−ethylene	−1.111	−1.302	−1.356	−1.362
	[−1.074]	[−1.280]	[−1.323]	
ethane−propylene	−1.543	−1.745	−1.808	−1.682
	[−1.380]	[−1.573]		
ethane−butylene	−1.652	−1.874		−1.900
	[−1.581]	[−1.806]		
ethane−pentylene	−1.886	−2.129		−2.139
	[−1.794]			
propane−ethylene	−1.362	−1.552	−1.610	−1.515
	[−1.237]	[−1.415]		
propane−propylene	−2.009	−2.252		−2.076
	[−1.760]	[−1.974]		
propane−butylene	−2.106	−2.353		−2.222
	[−1.871]			
propane−pentylene	−2.279	−2.549		−2.543
	[−2.159]			
butane−ethylene	−1.615	−1.806		−1.608
	[−1.378]			
butane−propylene	−2.355	−2.628		−2.435
	[−2.047]			
butane−butylene	−2.469	−2.737		−2.534
	[−2.153]			
butane−pentylene	−2.690	−3.009		−2.995
	[−2.542]			

Table 3. Binding Energies of the Dimers in the Ae−Ae Series

	MP2 [CCSD(T)]			CCSD(T)
	aDZ	aTZ	aQZ	CBS
ethylene−propylene	−1.687	−1.893	−1.962	−1.737
	[−1.438]	[−1.618]		
ethylene−butylene	−1.734	−1.937		−1.753
	[−1.490]	[−1.668]		
ethylene−pentylene	−2.008	−2.261		−2.150
	[−1.790]	[−2.030]		
propylene−butylene	−2.409	−2.665		−2.388
	[−2.024]			
propylene−pentylene	−2.810	−3.125		−2.863
	[−2.415]			

12 aldehyde−amide (Ad−Am) heterodimers (54 in total). Finally, we consider 15 binary complexes in the CAA−CAA set.

3.1. Data Set for the AAA−AAA Heterodimers.
3.1.1. Alkane−Alkane (Aa−Aa) Heterodimers. Figure 1 shows the optimized structures of the studied alkane−alkane heterodimers. Notice that in the data set only the *all-trans* *n*-alkanes are considered. We expect that the complexes are stabilized in a regular pattern due to their homology. Similar to their homodimer counterparts, larger heterodimers exhibit a binding pattern where the pair monomers are aligned in parallel with an inverse zigzag (staggered) contact geometry to avoid the stereorepulsion frustrations. This avoided stereorepulsion principle was first demonstrated clearly by Tsuzuki et al.[64,65] for the alkane homodimers and then verified by other groups, including ours.[66−69] Here we show that this principle also works for heterodimers.

In Table 1 we summarize the calculated MP2 and CCSD(T) energy data with the aug-cc-pVXZ (X = D, T, Q) basis sets (denoted as aDZ, aTZ, and aQZ, respectively) and the CBS extrapolation values for the alkane−alkane heterodimers. We see that the MP2 energy exhibits a systematic converging trend as the basis size increases. This indicates the good quality of Dunning's basis sets and the theoretically justified extrapolation rules. Our calculated energy data are consistent with previous benchmark CCSD(T)/CBS calculations for specific dimers. For example, the binding energy of the methane−ethane dimer is −0.825 (in kcal/mol), as compared to −0.827 (the A24 data set).[52]

3.1.2. Alkane−Alkene (Aa−Ae) Heterodimers. Figure 2 shows the optimized structures of the studied alkane−alkene heterodimers. Notice that for larger alkenes only the 1-alkenes are considered, so we will omit the numeral tag for brevity's sake. Overall, the complexes are stabilized in a regular pattern. For

heterodimers (34 in total). The AAA−AAK set contains 16 alkane−alcohol (Aa−Ac), 16 alkane−aldehyde (Aa−Ad), 12 alkane−ketone (Aa−K), 16 alkene−alcohol (Ae−Ac), 16 alkene−aldehyde (Ae−Ad), and 12 alkene−ketone (Ae−K) heterodimers (88 in total). The AAA−CAA contains 12 alkane−carboxylic acid (Aa−Ca), 12 alkane−amide (Aa−Am), 12 alkene−carboxylic acid (Ae−Ca), and 12 alkene−amide (Ae−Am) heterodimers (48 in total). The AAK−AAK set contains 6 alcohol−alcohol (Ac−Ac), and the AAK−CAA set contains 12 alcohol−carboxylic acid (Ac−Ca), 12 alcohol−amide (Ac−Am), 12 aldehyde−carboxylic acid (Ad−Ca), and
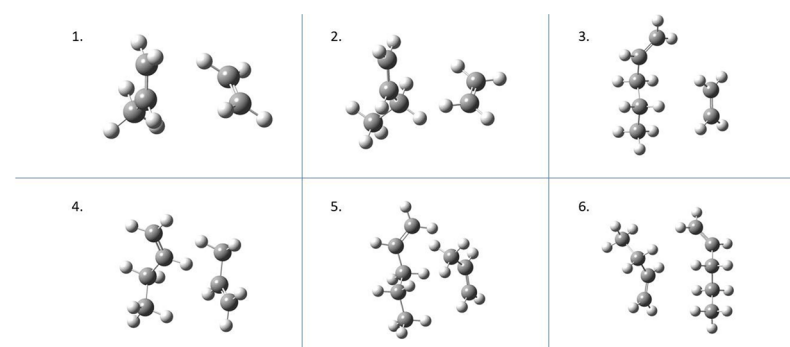

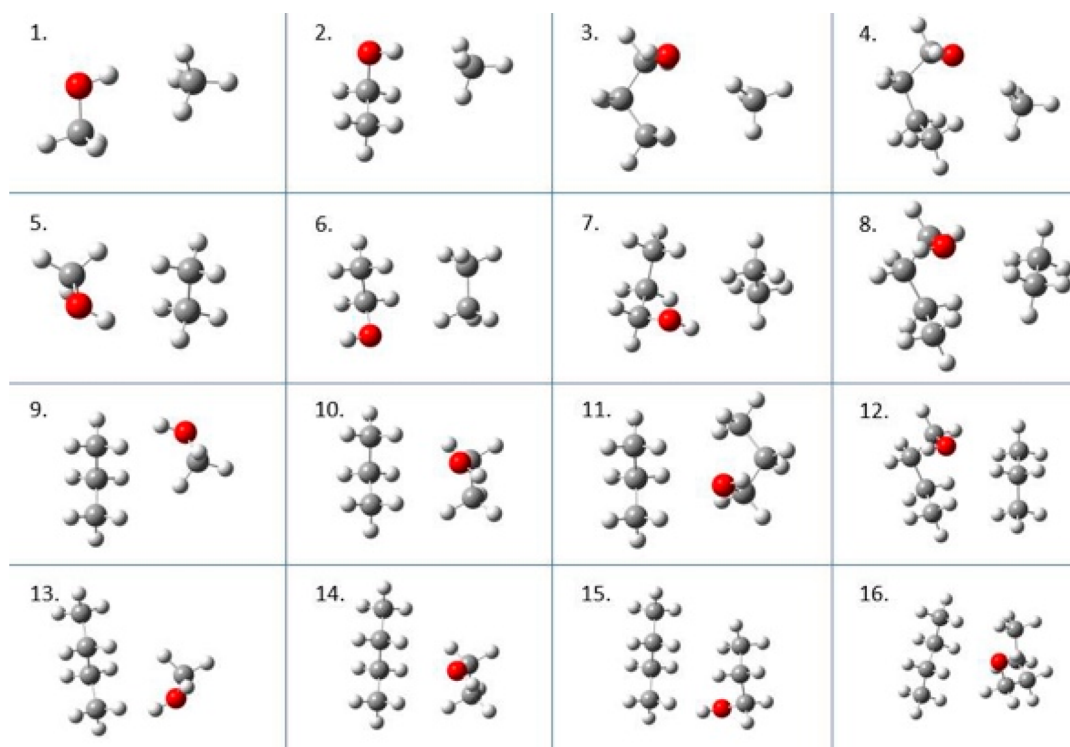**Figure 3.** Optimized structures of the dimers in the Ae−Ae series.

Figure 4. Optimized structures of the dimers in the Aa−Ac series.

both short-chain monomers, such as the methane−ethene and ethane−propene dimers, the alkane tends to incline toward the end double-bond of the paired alkene. As the chains become longer, they tend to align in parallel as in the alkane−alkane (Aa−Aa) heterodimers. In contrast to the latter, where the $\sigma-\sigma$ interaction (or dihydrogen bond) plays the role for stereo-repulsions, the short-chain alkane−alkene heterodimers employ the $\sigma-\pi$ interaction to avoid the orbital overlapping. For long-chain complexes, the alkyl tails tend to stabilize again with the $\sigma-\sigma$ interactions, thus yielding the binding patterns as shown in Figure 2.

Table 2 summarizes the MP2 and CCSD(T) energy data with different basis sets and their CBS extrapolation values for the alkane−alkene heterodimers. Looking at the specific values calculated from the aDZ to the aQZ basis sets, the binding energy follows a systematic converging trend as the basis size increases. This suggests the necessity of using at least the aTZ basis function for this series. The MP2/aQZ energy data are consistent with MP2/CBS calculations for specific dimers. For example, the binding energy of the methane−ethylene dimer is −0.863 kcal/mol, as compared to −0.889 kcal/mol.

3.1.3. Alkene−Alkene (Ae−Ae) Heterodimers. Figure 3 shows the optimized structures of the studied alkene−alkene heterodimers. For the alkene−alkene series, the functional active sites (heads) tend to form a T-shape cross pattern with respect to each other in order to keep the π bonds as far as possible and minimize the repulsion. This avoided stereorepulsion principle serves as a general stabilization mechanism for hydrocarbons. In this case, it is the $\pi-\pi$ interaction which plays the role of stereorepulsion.[70−72] For long-chain complexes, such as the butene−pentene heterodimer, the alkyl tails tend to stabilize using the $\sigma-\sigma$ interaction, thus competing with the functional heads.

Table 3 summarizes the MP2 and CCSD(T) energy data with different basis sets and their CBS extrapolation values for the

Table 4. Binding Energies of the Dimers in the Aa−Ac Series

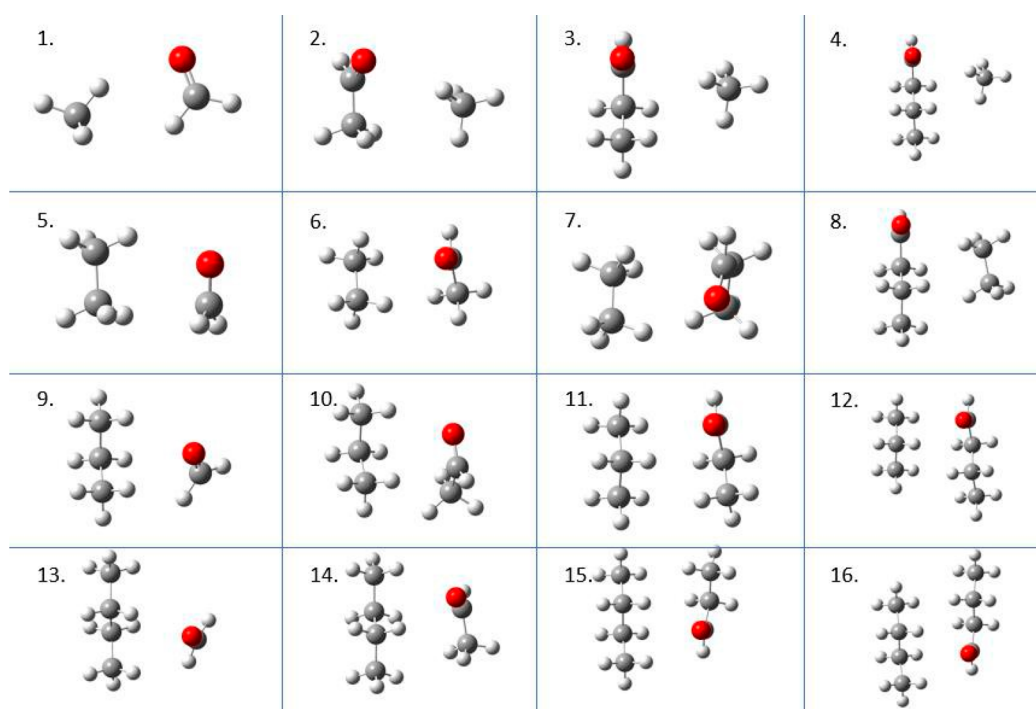|	MP2 [CCSD(T)]			CCSD(T)
	aDZ	aTZ	aQZ	CBS
methane−methanol	−0.957 [−1.006]	−1.185 [−1.259]	−1.236 [−1.315]	−1.352
methane−ethanol	−1.119 [−1.171]	−1.331 [−1.410]	−1.382	−1.498
methane−propanol	−1.242 [−1.265]	−1.382 [−1.426]	−1.433	−1.514
methane−butanol	−1.448 [−1.458]	−1.639 [−1.666]		−1.746
ethane−methanol	−1.396 [−1.443]	−1.697 [−1.783]	−1.773 [−1.849]	−1.904
ethane−ethanol	−1.481 [−1.477]	−1.687 [−1.710]	−1.751	−1.821
ethane−propanol	−1.693 [−1.727]	−1.951 [−2.021]	−2.022	−2.144
ethane−butanol	−1.882 [−1.861]	−2.127 [−2.125]		−2.228
propane−methanol	−1.592 [−1.646]	−1.912 [−1.992]	−1.993	−2.132
propane−ethanol	−2.059 [−2.018]	−2.333 [−2.317]	−2.410	−2.450
propane−propanol	−2.212 [−2.191]	−2.462 [−2.460]		−2.565
propane−butanol	−2.434 [−2.374]	−2.721 [−2.663]		−2.784
butane−methanol	−1.722 [−1.767]	−2.050 [−2.116]	−2.133	−2.260
butane−ethanol	−2.277 [−2.220]	−2.566 [−2.528]		−2.650
butane−propanol	−2.454 [−2.454]	−2.808 [−2.824]		−2.973
butane−butanol	−2.822 [−2.732]	−3.152		−3.201

Figure 5. Optimized structures of the dimers in the Aa—Ad series.

Table 5. Binding Energies of the Dimers in the Aa—Ad Series

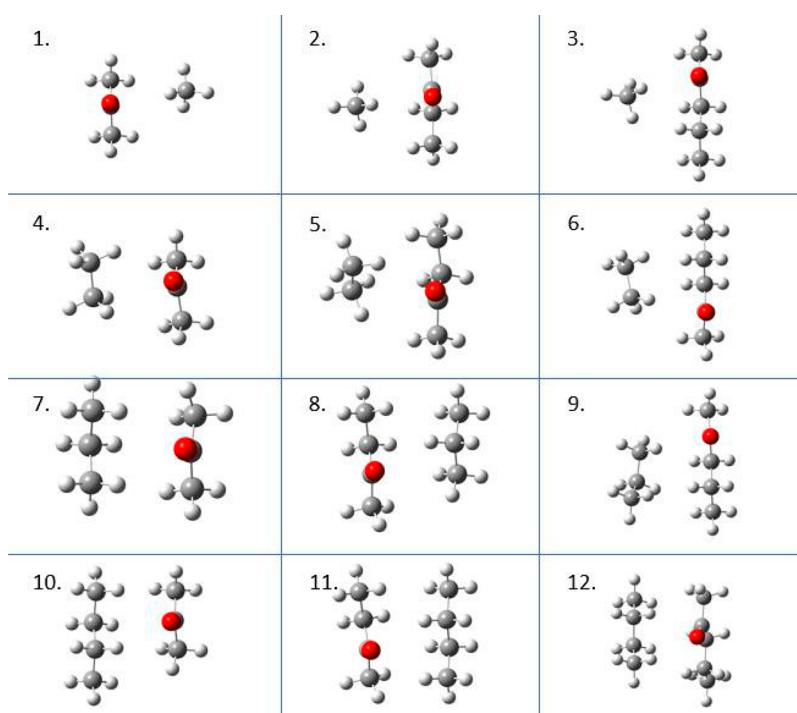|	MP2 [CCSD(T)]			CCSD(T)
	aDZ	aTZ	aQZ	CBS
methane−formaldehyde	−0.830	−0.981	−1.026	−1.125
	[−0.867]	[−1.042]	[−1.090]	
methane−acetaldehyde	−0.966	−1.111	−1.156	−1.182
	[−0.959]	[−1.106]	[−1.149]	
methane−propionaldehyde	−1.208	−1.385	−1.438	−1.430
	[−1.177]	[−1.355]		
methane−butyraldehyde	−1.336	−1.522		−1.560
	[−1.297]	[−1.482]		
ethane−formaldehyde	−1.074	−1.257	−1.319	−1.296
	[−1.011]	[−1.203]	[−1.251]	
ethane−acetaldehyde	−1.477	−1.712	−1.782	−1.751
	[−1.394]	[−1.630]		
ethane−propionaldehyde	−1.843	−2.090	−2.167	−2.127
	[−1.742]	[−1.994]		
ethane−butyraldehyde	−1.911	−2.174		−2.197
	[−1.824]	[−2.086]		
propane−formaldehyde	−1.456	−1.702	−1.776	−1.779
	[−1.402]	[−1.651]		
propane−acetaldehyde	−1.866	−2.141	−2.221	−2.145
	[−1.742]	[−2.007]		
propane−propionaldehyde	−2.121	−2.400		−2.387
	[−2.006]	[−2.270]		
propane−butyraldehyde	−2.425	−2.726		−2.712
	[−2.284]			
butane−formaldehyde	−1.582	−1.843	−1.923	−1.849
	[−1.464]	[−1.711]		
butane−acetaldehyde	−2.150	−2.432		−2.407
	[−2.018]	[−2.288]		
butane−propionaldehyde	−2.676	−3.021		−2.992
	[−2.502]			
butane−butyraldehyde	−2.990	−3.356		−3.305
	[−2.785]			

Figure 6. Optimized structures of the dimers in the Aa−K series.

Ae−Ae heterodimers. Similar to the Aa−Aa and Aa−Ae series, the aTZ basis is suggested for the binding energy calculations in this category. The MP2/aQZ energy data of ethylene−propylene is only 0.05 kcal/mol different from its MP2/CBS energy.

3.2. Data Set for the AAA−AAK Heterodimers. The molecules in the AAK groups contain an oxygen atom in the functional active site which introduces the possibility to form a hydrogen bond in a heterodimer. The strengths of such formed hydrogen bonds are expected to be weaker than the corresponding AAK−AAK homodimers. The details are discussed along with the following further specific complexes.

3.2.1. Alkane−Alcohol (Aa−Ac) Heterodimers. Figure 4 shows the optimized structures of the studied alkane−alcohol heterodimers. In this series, the functional hydroxyl end in the alcohol group tends to attract one carbon in the alkane group and the peripheral hydrogen atoms around the carbon are repelled from each other in order to minimize the repulsion. For long-chain complexes, the alkyl tails employ the same avoided stereorepulsion principle for hydrocarbons. We see the stabilized complexes are consistent with these two principles.

Table 4 summarizes the MP2 and CCSD(T) energy data with different basis sets and their CBS extrapolation values for the alkane−alcohol heterodimers. We see the energy follows a systematic converging trend as the basis size increases. This demonstrates the good quality of Dunning's basis set and the theoretically justified extrapolation rules, especially for larger alkyl groups. For example, The MP2/aQZ energy data of methane−methanol is only 0.037 kcal/mol different from its MP2/CBS energy.

3.2.2. Alkane−Aldehyde (Aa−Ad) Heterodimers. Figure 5 shows the optimized structures of the studied alkane−aldehyde heterodimers. Compared to alcohols, the carbonyl oxygen in an aldehyde tends to attract one hydrogen in the paired alkane group. Therefore, there is one hydrogen in the alkane pointing to the oxygen in the aldehyde. We see that the local −C=O−H

Table 6. Binding Energies of the Dimers in the Aa−K Series

	MP2 [CCSD(T)]			CCSD(T)
	aDZ	aTZ	aQZ	CBS
methane−acetone	−1.186	−1.346	−1.395	−1.402
	[−1.160]	[−1.317]		
methane−butanone	−1.464	−1.658		−1.698
	[−1.417]	[−1.608]		
methane−pentanone	−1.579	−1.781		−1.801
	[−1.523]	[−1.716]		
ethane−acetone	−1.857	−2.104	−2.177	−2.115
	[−1.746]	[−1.989]		
ethane−butanone	−2.259	−2.534		−2.507
	[−2.114]	[−2.391]		
ethane−pentanone	−2.231	−2.506		−2.506
	[−2.119]	[−2.390]		
propane−acetone	−2.529	−2.854		−2.781
	[−2.336]	[−2.644]		
propane−butanone	−2.528	−2.821		−2.782
	[−2.366]			
propane−pentanone	−2.672	−2.979		−2.954
	[−2.518]			
butane−acetone	−2.838	−3.175		−3.091
	[−2.632]	[−2.949]		
butane−butanone	−3.332	−3.717		−3.624
	[−3.077]			
butane−pentanone	−3.342	−3.733		−3.644
	[−3.088]			

structure appears in all the stabilized complexes. Again, for long-chain complexes, the alkyl tails tend to align in parallel. We see the stabilized complexes are consistent with these observations.

Table 5 summarizes the MP2 and CCSD(T) energy data with different basis sets and their CBS extrapolation values for the alkane−aldehyde heterodimers. The property of the systematic converging trend of basis size can also be seen here. Most
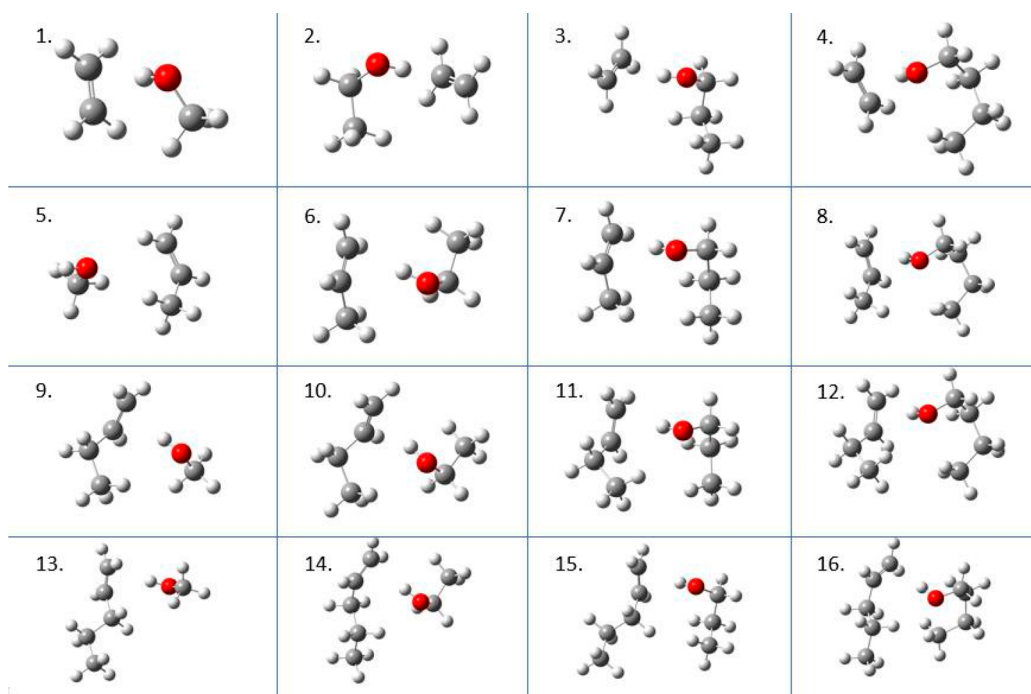
Figure 7. Optimized structures of the dimers in the Ae−Ac series.

MP2/aQZ data are presented here for the first time, and their CCSD(T)/CBS data can serve as benchmark values for comparison.

3.2.3. Alkane−Ketone (Aa−K) Heterodimers. Figure 6 shows the optimized structures of the studied alkane−ketone heterodimers. Similar to the aldehydes, the ketone functional oxygen also tends to attract one hydrogen in the alkane group. The side methyl group does not cause too much distortion of the local −C=O−H structures as this pattern is quite directional. As the alkyl chain gets long, the complexes tend to align in parallel.

Table 6 summarizes the MP2 and CCSD(T) energy data with different basis sets and their CBS extrapolation values for the alkane−ketone heterodimers. For the methane−acetone and the ethane−acetone dimers, where the MP2/aQZ optimization is converged, we see the energy follows a systematic converging trend as the basis size increases. Similar to the previous series, the improvement using the aTZ with respect to the aDZ basis sets is more significant than that using the aQZ with respect to the aTZ basis sets. For example, for the methane−acetone dimer, the energy difference MP2/aDZ-aTZ is 0.160 kcal/mol, while the MP2/aTZ-aQZ is only 0.049 kcal/mol. Thus, at least the aTZ basis function should be used for the geometry optimization.

3.2.4. Alkene−Alcohol (Ae−Ac) Heterodimers. Figure 7 shows the optimized structures of the studied alkene−alcohol heterodimers. For this series, the functional −OH end in the alcohol group tends to attract the nucleophilic region in the alkene group. Therefore, the local −OH−π pattern is found in all the stabilized complexes. This hydrogen-mediated bonding is of similar higher directionality to a hydrogen bond so when the chains get longer, the alkyl tails yield to this dominant structural pattern but do not always align in parallel. This subtle point can be seen very clearly in Figure 7.

Table 7 summarizes the MP2 and CCSD(T) energy data with different basis sets and their CBS extrapolation values for the alkene−alcohol heterodimers. The systematic converging trend

Table 7. Binding Energies of the Dimers in the Ae−Ac Series

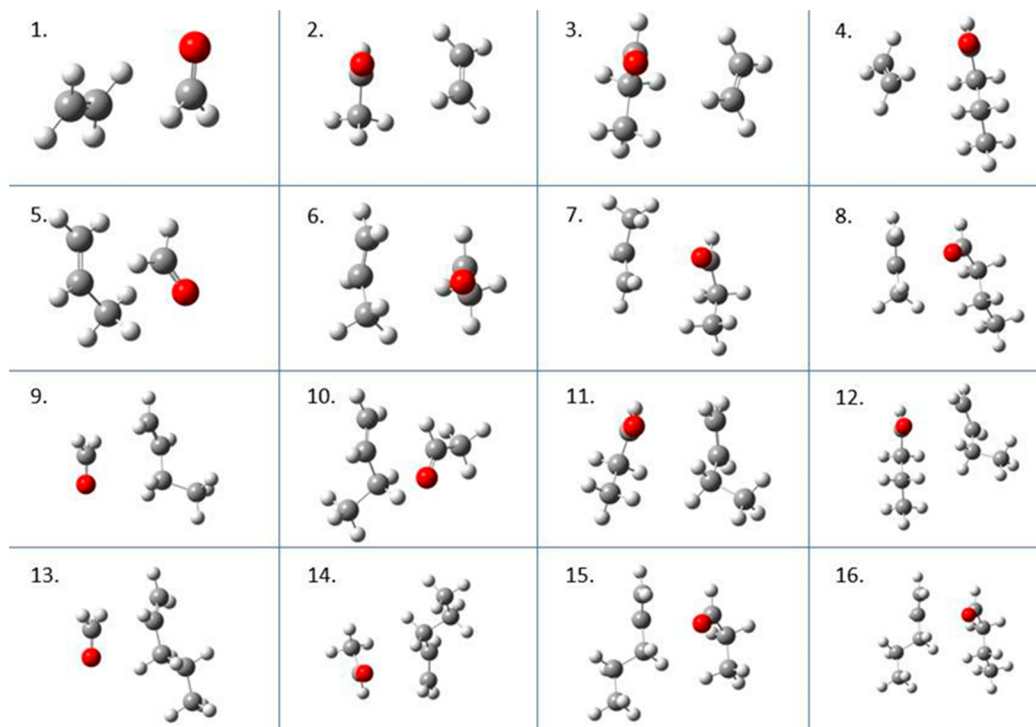|	MP2 [CCSD(T)]			CCSD(T)
	aDZ	aTZ	aQZ	CBS
ethylene−methanol	−2.614	−2.934	−3.018	−2.835
	[−2.394]	[−2.711]	[−2.784]	
ethylene−ethanol	−2.786	−3.129	−3.217	−3.026
	[−2.537]	[−2.874]		
ethylene−propanol	−2.917	−3.262		−3.131
	[−2.651]	[−2.986]		
ethylene−butanol	−3.043	−3.392		−3.266
	[−2.776]	[−3.119]		
propylene−methanol	−2.228	−2.471	−2.556	−2.526
	[−2.114]	[−2.379]		
propylene−ethanol	−3.645	−4.053		−3.934
	[−3.356]	[−3.762]		
propylene−propanol	−3.878	−4.306		−4.142
	[−3.545]	[−3.962]		
propylene−butanol	−4.104	−4.542		−4.359
	[−3.737]			
butylene−methanol	−3.602	−4.045	−4.167	−4.046
	[−3.384]	[−3.835]		
butylene−ethanol	−3.864	−4.321		−4.278
	[−3.627]	[−4.086]		
butylene−propanol	−4.049	−4.511		−4.428
	[−3.771]			
butylene−butanol	−4.307	−4.790		−4.701
	[−4.015]			
pentylene−methanol	−3.531	−3.914		−3.834
	[−3.288]	[−3.673]		
pentylene−ethanol	−3.828	−4.309		−4.303
	[−3.619]			
pentylene−propanol	−4.047	−4.471		−4.298
	[−3.695]			
pentylene−butanol	−4.592	−5.121		−5.076
	[−4.324]			

Figure 8. Optimized structures of the dimers in the Ae−Ad series.

Table 8. Binding Energies of the Dimers in the Ae−Ad Series

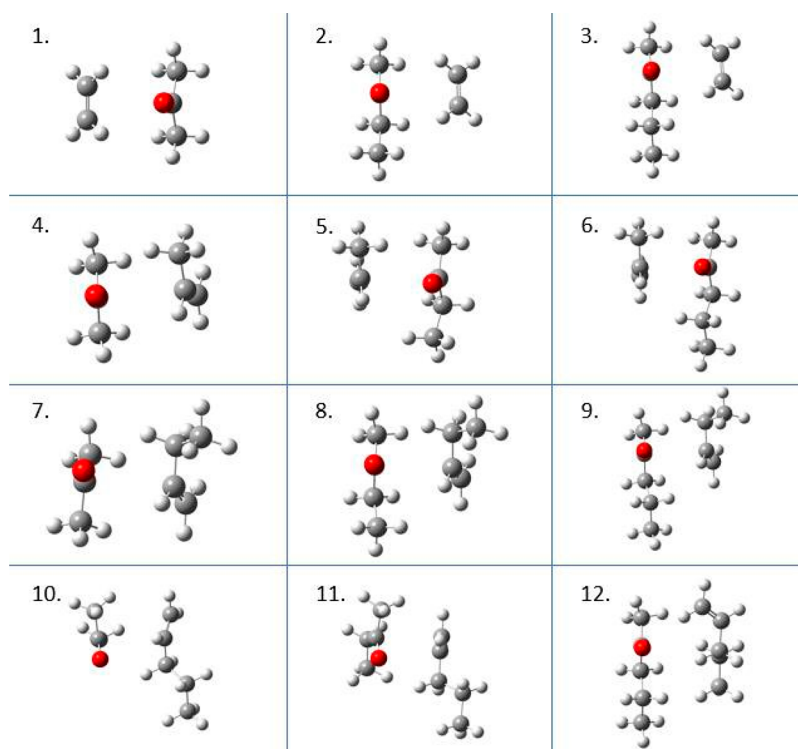|	MP2 [CCSD(T)]			CCSD(T)
	aDZ	aTZ	aQZ	CBS
ethylene−formaldehyde	−1.629	−1.845	−1.924	−1.781
	[−1.468]	[−1.661]	[−1.725]	
ethylene−acetaldehyde	−2.038	−2.287	−2.370	−2.226
	[−1.851]	[−2.082]	[−2.165]	
ethylene−propanal	−2.302	−2.576		−2.413
	[−2.049]	[−2.298]		
ethylene−butanal	−2.437	−2.719		−2.517
	[−2.144]	[−2.398]		
propylene−formaldehyde	−2.536	−2.845	−2.964	−2.668
	[−2.204]	[−2.462]		
propylene−acetaldehyde	−2.767	−3.091		−2.884
	[−2.466]	[−2.748]		
propylene−propanal	−2.842	−3.158		−2.962
	[−2.544]	[−2.829]		
propylene−butanal	−3.027	−3.360		−3.208
	[−2.735]			
butylene−formaldehyde	−2.515	−2.816	−2.930	−2.662
	[−2.213]	[−2.465]		
butylene−acetaldehyde	−3.034	−3.350		−3.105
	[−2.701]	[−2.974]		
butylene−propanal	−2.990	−3.300		−3.158
	[−2.717]			
butylene−butanal	−3.217	−3.535		−3.359
	[−2.907]			
pentylene−formaldehyde	−2.555	−2.851		−2.620
	[−2.248]	[−2.495]		
pentylene−acetaldehyde	−2.924	−3.228		−3.011
	[−2.623]	[−2.883]		
pentylene−propanal	−3.165	−3.479		−3.313
	[−2.867]			
pentylene−butanal	−3.441	−3.767		−3.574
	[−3.111]			

Figure 9. Optimized structures of the dimers in the Ae−K series.

with increasing basis size provides confidence of the CCSD(T)/ CBS calculations. In general, the larger the alkyl group is, the higher the binding energy.

3.2.5. Alkene−Aldehyde (Ae−Ad) Heterodimers. Figure 8 shows the optimized structures of the studied alkene−aldehyde heterodimers. The aldehyde functional oxygen tends to attract one hydrogen in the alkene group. However, the attraction is largely reduced by the confronting $\pi-\pi$ repulsion, which is similar to the alkene−alkene heterodimers. The two double bonds are thus tending to avoid each other. Therefore, we see the local T-shape structure appear in all the stabilized complexes. This bonding pattern is of high directionality, so when the chains get longer, the alkyl tails yield to this dominant structural pattern but not always align in parallel. This subtle point can be seen very clearly in Figure 8.

Table 8 summarizes the MP2 and CCSD(T) energy data with different basis sets and their CBS extrapolation values for the alkane−aldehyde heterodimers. Both the energy converging trends with respect to the basis size and the alkyl group size can be seen in this series. In comparison to the corresponding Ae−Ac serious, the Ae−Ad series has lower binding energy. This might be due to the confronting $\pi-\pi$ repulsion as described in above.

3.2.6. Alkene−Ketone (Ae−K) Heterodimers. Figure 9 shows the optimized structures of the studied alkene−ketone hetero-dimers. Similar to aldehydes, the ketone functional oxygen also tends to attract one hydrogen in the alkene group but is hindered by the confronting $\pi-\pi$ repulsion. This is further complicated by the side methyl group which tends to tilt the local perpendicular structures. Similar to the alkene−aldehyde heterodimers, we observe the local T-shape structure in all the stabilized complexes, and for longer chains, the alkyl tails do not always align in parallel. We see in this case there are several competing mechanisms for stabilizing the overall conformations.

Table 9 summarizes the MP2 and CCSD(T) energy data with different basis sets and their CBS extrapolation values for the

Table 9. Binding Energies of the Dimers in the Ae−K Series

	MP2 [CCSD(T)]			CCSD(T)
	aDZ	aTZ	aQZ	CBS
ethylene−acetone	−2.618 [−2.334]	−2.914 [−2.603]	−3.011	−2.771
ethylene−butanone	−2.773 [−2.457]	−3.086 [−2.738]		−2.870
ethylene−pentanone	−2.880 [−2.542]	−3.194 [−2.820]		−2.952
propylene−acetone	−3.516 [−3.094]	−3.863 [−3.397]		−3.543
propylene−butanone	−3.896 [−3.394]	−4.280		−3.949
propylene−pentanone	−4.050 [−3.518]	−4.432		−4.061
butylene−acetone	−3.623 [−3.193]	−3.965		−3.679
butylene−butanone	−4.010 [−3.505]	−4.387		−4.041
butylene−pentanone	−4.171 [−3.638]	−4.546		−4.171
pentylene−acetone	−3.731 [−3.290]	−4.073		−3.776
pentylene−butanone	−4.048 [−3.554]	−4.420		−4.083
pentylene−pentanone	−4.027 [−3.747]	−4.462		−4.365

alkene−ketone heterodimers. The computational cost when enlarging the alkyl groups on both the alkene and ketone sites is significant for this series, mainly because there is no regular expected configurations to initiate the optimization. Therefore, a proper choice of a smaller basis size for balancing the com-putational cost is necessary. For example, the MP2 energy of the
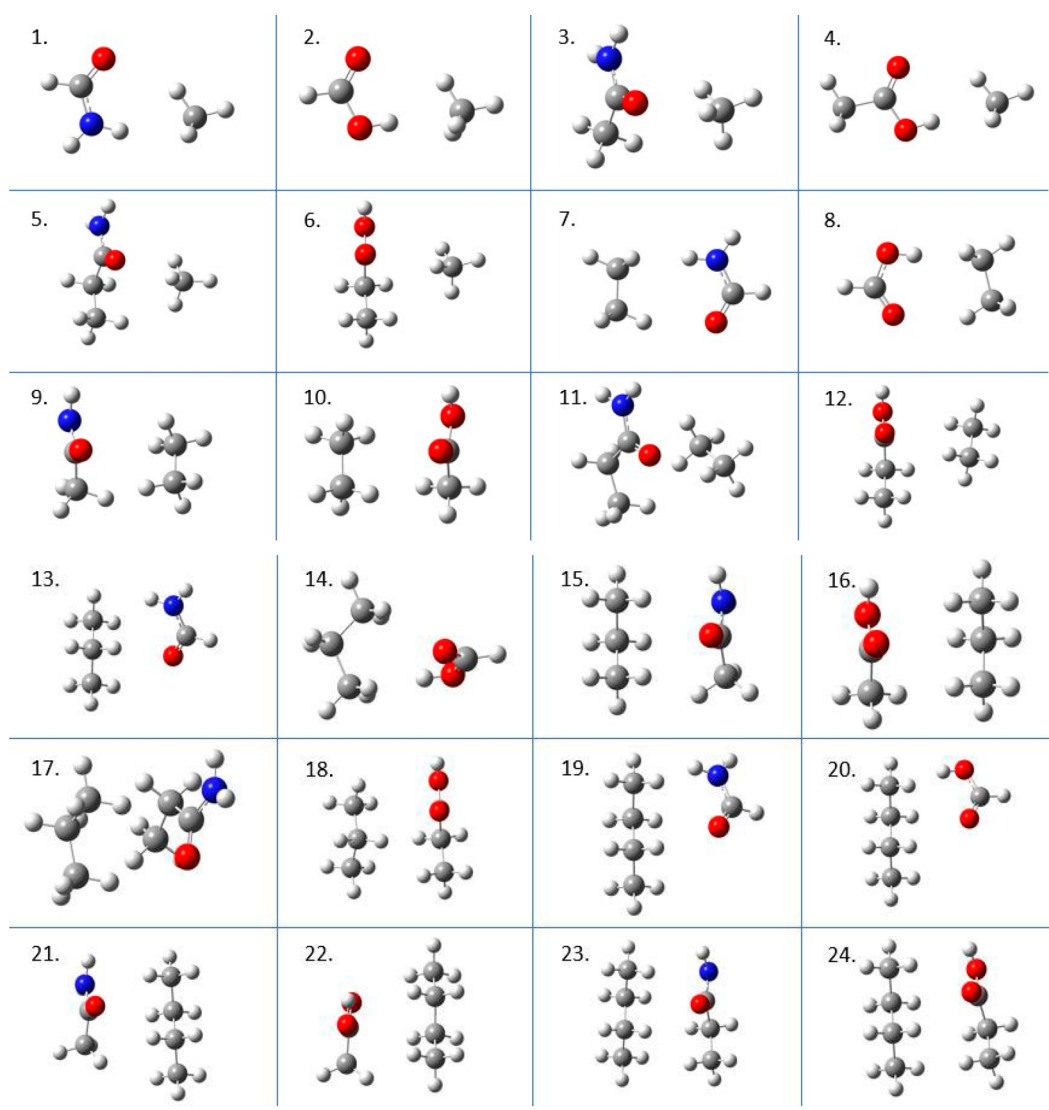
Figure 10. Optimized structures of the dimers in the Aa−Am and Aa−Ca series.

ethylene−acetone dimer follows a systematic converging trend as the basis size increases, and the aTZ basis set is suggested.

3.3. Data Set for the AAA−CAA Heterodimers. The molecules in the CAA groups all contain two functional active sites for the hydrogen bond donor and acceptor, respectively. However, the paired monomers are hydrocarbons, so the possibility of forming the double hydrogen bonding pattern decreases as the chains get longer. A more generally expected pattern would be the formation of a weak hydrogen bond, similar to the AAA−AAK heterodimers discussed in section 3.2. The strengths of interaction are expected to be stronger than the corresponding AAA−AAK heterodimers but may compete with those of the AAK−AAK homodimers. The details are discussed along with the following further specific complexes.

3.3.1. Alkane−Amide (Aa−Am) and Alkane−Carboxylic Acid (Aa−Ca) Heterodimers. Figure 10 shows the optimized structures of the studied Aa−Am and Aa−Ca heterodimers in the alkane−CAA series. For this series, the functional hydrogen donor (acceptor) site tends to attract one carbon (hydrogen) in the alkane group. Overall the pattern is dominated by a major single hydrogen bond with a compromise in balancing a weaker electrostatic interaction and a van der Waals bond. The

peripheral hydrogen atoms are repelled from each other so as to minimize the repulsion. For long-chain complexes, the alkyl tails employ the same avoided stereorepulsion principle for hydrocarbons. We see the stabilized complexes are consistent with these principles.

Tables 10 and 11 summarize the MP2 and CCSD(T) energy data with different basis sets and their CBS extrapolation values for the Aa−Am and Aa−Ca heterodimers, respectively. The basis set effect can clearly be seen in these two categories. The energy follows a systematic converging trend as the basis size increases, especially for the energy difference calculated between the aDZ and the aTZ basis sets. This demonstrates the good quality of Dunning's basis sets and the theoretically justified extrapolation rules. Because of the competition mechanism, the hydrogen bonding pattern is not significant in this series, so the binding energy is lower than the usual strength of a hydrogen bond. This implies that the electrostatic interaction is not the dominate attraction term. The MP2/aQZ energy data of the methane−formic acid dimer is only 0.047 kcal/mol different from its MP2/CBS energy.

3.3.2. Alkene−Amide (Ae−Am) and Alkene−Carboxylic Acid (Ae−Ca) Heterodimers. The optimized heterodimers

Table 10. Binding Energies of the Dimers in the Aa−Am Series

	MP2 [CCSD(T)]			CCSD(T)
	aDZ	aTZ	aQZ	CBS
methane−formamide	−1.170 [−1.220]	−1.398 [−1.483]	−1.453 [−1.542]	−1.585
methane−acetamide	−1.263 [−1.282]	−1.405 [−1.425]	−1.462 [−1.473]	−1.515
methane−propanamide	−2.432 [−2.479]	−2.534 [−2.572]	−2.587	−2.664
ethane−formamide	−1.563 [−1.630]	−1.792 [−1.903]	−1.857	−2.015
ethane−acetamide	−1.849 [−1.816]	−2.073 [−2.044]	−2.146	−2.174
ethane−propionamide	−3.125 [−3.126]	−3.292 [−3.291]		−3.361
propane−formamide	−2.131 [−2.113]	−2.451 [−2.462]	−2.539	−2.614
propane−acetamide	−2.647 [−2.567]	−2.987 [−2.883]		−3.026
propane−propanamide	−3.604 [−3.561]	−3.808		−3.831
butane−formamide	−2.376 [−2.326]	−2.709 [−2.683]		−2.823
butane−acetamide	−2.869 [−2.761]	−3.179		−3.202
butane−propanamide	−4.305 [−4.225]	−4.593		−4.634

Table 11. Binding Energies of the Dimers in the Aa−Ca Series

	MP2 [CCSD(T)]			CCSD(T)
	aDZ	aTZ	aQZ	CBS
methane−formic acid	−1.238 [−1.274]	−1.564 [−1.644]	−1.628	−1.759
methane−acetic acid	−1.210 [−1.269]	−1.517 [−1.620]	−1.580	−1.733
methane−propanoic acid	−1.390 [−1.368]	−1.593 [−1.574]	−1.651	−1.674
ethane−formic acid	−1.598 [−1.661]	−1.904 [−2.018]	−1.979	−2.148
ethane−acetic acid	−1.607 [−1.560]	−1.850 [−1.808]	−1.922	−1.933
ethane−propanoic acid	−1.859 [−1.808]	−2.114 [−2.069]		−2.176
propane−formic acid	−1.950 [−2.005]	−2.287 [−2.382]	−2.376	−2.536
propane−acetic acid	−2.194 [−2.111]	−2.510 [−2.424]		−2.557
propane−propanoic acid	−2.237 [−2.156]	−2.525		−2.565
butane−formic acid	−2.130 [−2.117]	−2.551 [−2.571]		−2.748
butane−acetic acid	−2.397 [−2.289]	−2.732 [−2.617]		−2.758
butane−propanoic acid	−2.927 [−2.804]	−3.308		−3.345

paired by the alkene−amide (Ae−Am) and the alkene−carboxylic (Ae−Ca) groups are shown in Figure 11. The dimers are bonded together by an −H−π interaction with a −C=O−H side interaction, where the former comes from the −NH in the Ae−Am and the −OH in the Ae−Ca groups, respectively. The functional hydrogen bond donor site (−OH for the carboxylic acid and − NH for the amide) tends to attract the electrophilic region in the alkene group. Therefore, the local −OH−π or −NH−π pattern is found in all the stabilized complexes. This hydrogen bonding is of high directionality, so when the chains
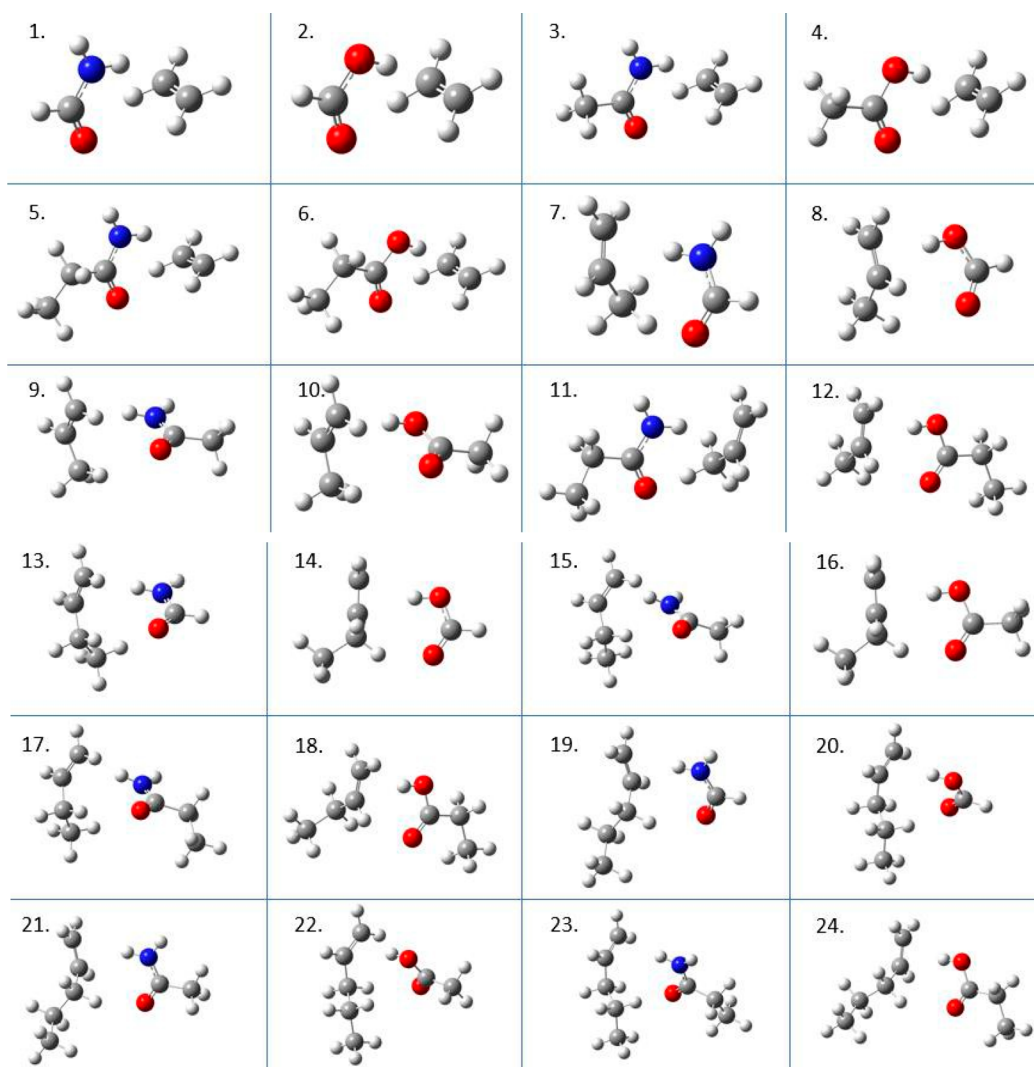
Figure 11. Optimized structures of the dimers in the Ae−Am and Ae−Ca series.

get longer, the alkyl tails yield to this dominate structural pattern but do not always align in parallel.

Tables 12 and 13 summarize the MP2 and CCSD(T) energy data with different basis sets and their CBS extrapolation values for the Ae−Am and the Ae−Ca heterodimers, respectively. The improvement using the aTZ with respect to the aDZ basis sets is more significant than that of using the aQZ with respect to the aTZ basis sets. Generally, the binding energy of an Ae−Ca dimer is slightly larger than that of the corresponding Ae−Am dimer.

3.4. Database for the AAK−AAK Heterodimers. Figure 12 shows the optimized structures of the stabilized dimers in the AAK groups. As expected, all bonding patterns show the single −O−H−O hydrogen-bonded configurations. When the alkyl group gets longer, there is a competition between the other attractive components and the hydrogen bonding. The alcohol series is clearly dominated by the electrostatic energy, while the dispersion energy due to the alkyl group adds up to modify the configuration. As can be seen in Figure 12, the electrostatic energy and the dispersion energy compete in these dimers.

Table 14 summarizes the MP2 and CCSD(T) energy data with different basis sets and their CBS extrapolation values for the AAK groups. The data exhibit a systematic convergence trend, which again shows the calculations are of high-level quality. Overall, our MP2/aQZ calculated energy data are within

Table 12. Binding Energies of the Dimers in the Ae−Am Series

	MP2 [CCSD(T)]			CCSD(T)
	aDZ	aTZ	aQZ	CBS
ethylene−formamide	−3.266 [−3.081]	−3.630 [−3.470]	−3.736	−3.653
ethylene−acetamide	−3.327 [−3.169]	−3.673 [−3.538]	−3.779	−3.721
ethylene−propionamide	−4.261 [−4.149]	−4.502 [−4.411]		−4.512
propylene−formamide	−4.134 [−3.864]	−4.506 [−4.260]	−4.624	−4.464
propylene−acetamide	−4.092 [−3.868]	−4.499 [−4.293]		−4.474
propylene−propionamide	−5.043 [−4.842]	−5.279 [−5.090]		−5.189
butylene−formamide	−4.212 [−3.971]	−4.657 [−4.436]	−4.790	−4.632
butylene−acetamide	−4.311 [−4.095]	−4.733		−4.692
butylene−propionamide	−5.288 [−5.109]	−5.603		−5.557
pentylene−formamide	−4.300	−4.664		−4.559

Table 12. continued

	MP2 [CCSD(T)]			CCSD(T)
	aDZ	aTZ	aQZ	CBS
pentylene—acetamide	[−4.012] −4.323 [−4.065]	[−4.397] −4.671		−4.560
pentylene— propionamide	−5.788 [−5.559]	−6.143		−6.063

Table 13. Binding Energies of the Dimers in the Ae—Ca Series

	MP2 [CCSD(T)]			CCSD(T)
	aDZ	aTZ	aQZ	CBS
ethylene—formic acid	−3.875 [−3.581]	−4.392 [−4.134]	−4.516	−4.348
ethylene—acetic acid	−3.715 [−3.475]	−4.222 [−4.016]	−4.346	−4.230
ethylene—propanoic acid	−3.700 [−3.473]	−4.205 [−4.010]		−4.223
propylene—formic acid	−4.921 [−4.530]	−5.466 [−5.107]	−5.606	−5.349
propylene—acetic acid	−4.564 [−4.229]	−5.131 [−4.825]		−5.064
propylene—propanoic acid	−4.687 [−4.368]	−5.205 [−4.912]		−5.130
butylene—formic acid	−4.946 [−4.564]	−5.481 [−5.126]	−5.619	−5.365
butylene—acetic acid	−4.726 [−4.404]	−5.237		−5.130
butylene—propanoic acid	−4.720 [−4.396]	−5.227		−5.116
pentylene—formic acid	−5.212 [−4.834]	−5.867		−5.765
pentylene—acetic acid	−5.118 [−4.783]	−5.754		−5.687
pentylene—propanoic acid	−4.800 [−4.486]	−5.305		−5.204

0.1 kcal/mol difference from the MP2/CBS calculations for specific dimers (e.g., the methanol—ethanol and methanol—propanol dimers).

3.5. Database for the AAK—CAA Heterodimers.
3.5.1. Alcohol—Amide (Ac—Am) and Alcohol—Carboxylic Acid (Ac—Ca) Heterodimers. Figure 13 shows the optimized structures of the studied Ac—Am and Ac—Ca heterodimers.

In these two series, there are one hydrogen bond donor and one hydrogen bond acceptor on each monomer site, which offers the opportunity of forming double hydrogen bonds within the pairs. For the Ac—Am dimers, the two hydrogen bonds stem from an −OH on the alcohol with an oxygen on the amide, and an −NH on the amide with an oxygen on the alcohol. On the other hand, for the Ac—Ca dimers, the two hydrogen bonds stem from an −OH on the alcohol with an oxygen on the carboxylic acid and an −OH on the carboxylic acid with an oxygen on the alcohol. If the alcohol is methanol, the double hydrogen bond forms a planar ring. The short alkyl tail in methanol does not alter the double hydrogen bond. However, as the tail of alcohol gets longer, the alkyl group comes into play and leads to more complicated structures.

From Tables 15 and 16, where we summarize the calculated MP2 and CCSD(T) energy data for the Ac—Am and the Ac—Ca heterodimers, respectively, we can see that the hydrogen bond dominates the binding energy of these heterodimers (ca. 10−11 kcal/mol). However, because of the competition mechanism, larger alkyl groups on the alcohols do not always render larger binding energies. For example, the binding energy of propanol—formic acid is actually larger than that of butanol—formic acid. Similarly, longer alkyl groups on the carboxylic acids do not guarantee larger binding energies either. This is because the partial charges on the oxygen atom, both in the alcohols and the carboxylic acids, are modified by the longer alkyl groups.

3.5.2. Aldehyde—Amide (Ad—Am) and Aldehyde—Carboxylic Acid (Ad—Ca) Heterodimers. Figure 14 shows the optimized structures of the studied Ad—Am and Ad—Ac heterodimers. In the two series, an Ad—Ac or an Ad—Am dimer tends to form a planar double hydrogen bonded ring with the corresponding carbonyl functional groups. However, there is a subtle difference between these two series. The Ad—Am dimer indeed remains the planar pattern for shorter chains, but the carbonyl oxygen interacts with the other hydrogens on the longer alkyl groups (e.g., contrast panels 15, 17, and 19 and panels 16, 18, and 20 in Figure 14).

In Tables 17 and 18 we summarize the calculated MP2 and CCSD(T) energy data for the Ad—Am and the Ad—Ca heterodimers, respectively. We can see that the hydrogen bond dominates the binding energy of these heterodimers (ca. 8−9 kcal/mol). However, because of the competition mechanism, larger alkyl groups on both chains do not always render larger binding energies. This is also because the partial charges on the involved atoms are modified by the longer alkyl groups. The binding energies in the series are generally less than those of the corresponding Ac—Am and Ac—Ca series.
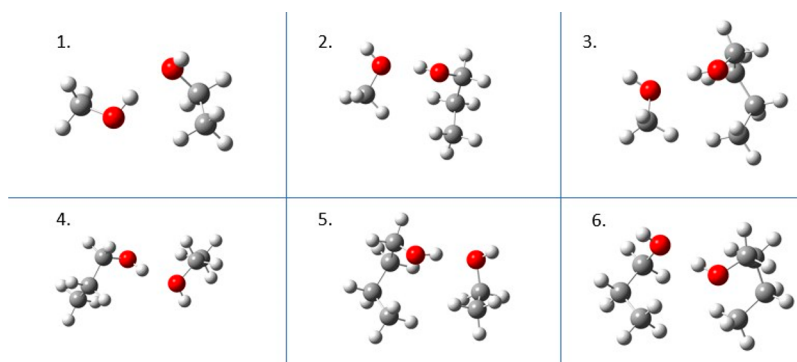

Figure 12. Optimized structures of the dimers in the Ac—Ac series.

Table 14. Binding Energies of the Dimers in the Ac−Ac Series

	MP2 [CCSD(T)]			CCSD(T)
	aDZ	aTZ	aQZ	CBS
methanol−ethanol	−5.852	−6.226	−6.402	−6.512
	[−5.759]	[−6.208]		
methanol−propanol	−5.629	−6.008	−6.186	−6.292
	[−5.555]	[−5.984]		
methanol−butanol	−5.721	−6.121		−6.266
	[−5.646]	[−6.098]		
ethanol−propanol	−5.997	−6.382		−6.518
	[−5.905]	[−6.356]		
ethanol−butanol	−6.528	−6.968		−7.123
	[−6.436]	[−6.938]		
propanol−butanol	−6.591	−7.070		−7.232
	[−6.530]	[−7.030]		

3.6. Database for the CAA−CAA Heterodimers.
Figure 15 shows the optimized structures of the studied Am−Am, Am−Ca, and Ca−Ca heterodimers. For these dimers, there are clearly one hydrogen bond donor and one hydrogen bond acceptor on the paired monomers, respectively. It is expected to form a double hydrogen bond pattern. Three types of double hydrogen bonds are shown in Figure 15, namely, two N−H−O hydrogen bonds in the Am−Am dimers, two O−H−O

hydrogen bonds in the Ca−Ca dimers, and one N−H−O hydrogen bond and one O−H−O hydrogen bond in the Am−Ca dimers. The double hydrogen bonded functional groups invariantly form a planar ring structure which represents a significant feature for such complexes.[73]

From Table 19, where we summarize the calculated MP2 and CCSD(T) energy data for the CAA−CAA heterodimers, we can see that the double hydrogen bond dominates the binding energy of each heterodimer. Although a longer alkyl tail does not imply a larger binding energy, the contribution from the alkyl tails is less significant than the other series discussed in the above.

3.7. Segmental SAPT Energy Decomposition Analysis.
In order to gain chemical insights of the calculated (total) interaction energies, we perform an energy decomposition analysis based on the symmetry-adapted perturbation theory (SAPT0/jun-cc-pVXZ, X = D and T).[74] Here, the full interaction energy is decomposed into four components: electrostatic, induction, dispersion, and exchange. Table S1 (see the Supporting Information) lists the four components of the SAPT binding energies for all the studied dimers. The attractive energy is composed of the electrostatic energy, the induction energy, and the dispersion energy, whereas the repulsive energy stems from the exchange term. To see the interesting interplay of the
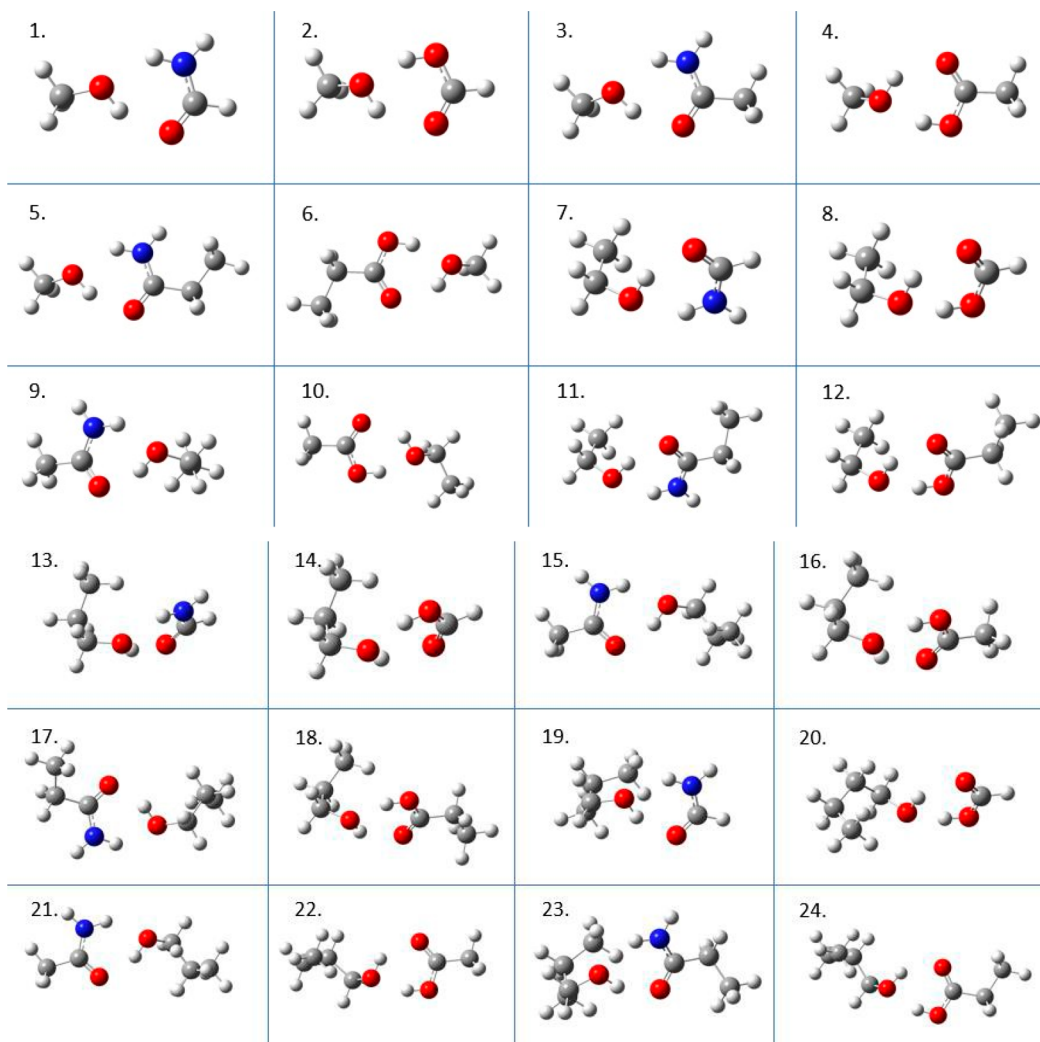

Figure 13. Optimized structures of the dimers in the Ac−Am and Ac−Ca series.

Table 15. Binding Energies of the Dimers in the Ac−Am Series

	MP2 [CCSD(T)]			CCSD(T)
	aDZ	aTZ	aQZ	CBS
methanol−formamide	−9.091	−9.714	−10.003	−10.277
	[−9.052]	[−9.765]	[−10.066]	
methanol−acetamide	−9.443	−10.064	−10.360	−10.651
	[−9.429]	[−10.139]		
methanol−propionamide	−9.587	−10.225		−10.578
	[−9.581]	[−10.309]		
ethanol−formamide	−9.236	−9.863	−10.154	−10.412
	[−9.186]	[−9.909]		
ethanol−acetamide	−9.612	−10.255	−10.554	−10.837
	[−9.585]	[−10.320]		
ethanol−propionamide	−10.501	−11.021		−11.321
	[−10.518]	[−11.102]		
propanol−formamide	−9.468	−10.158	−10.451	−10.683
	[−9.404]	[−10.176]	[−10.469]	
propanol−acetamide	−9.715	−10.330		−10.649
	[−9.690]	[−10.390]		
propanol−propionamide	−10.593	−11.075		−11.365
	[−10.614]	[−11.162]		
butanol−formamide	−9.349	−10.208		−10.569
	[−9.292]	[−10.207]		
butanol−acetamide	−9.749	−10.374		−10.601
	[−9.716]			
butanol−propionamide	−10.779	−11.421		−11.671
	[−10.762]			

Table 16. Binding Energies of the Dimers in the Ac−Ca Series

	MP2 [CCSD(T)]			CCSD(T)
	aDZ	aTZ	aQZ	CBS
methanol−formic acid	−9.828	−10.59	−10.917	−11.147
	[−9.681]	[−10.559]	[−10.908]	
methanol−acetic acid	−9.743	−10.514	−10.851	−11.134
	[−9.673]	[−10.551]		
methanol−propanoic acid	−9.69	−10.455		−10.836
	[−9.643]	[−10.514]		
ethanol−formic acid	−10.09	−10.883	−11.212	−11.407
	[−9.923]	[−10.838]		
ethanol−acetic acid	−10.094	−10.822	−11.158	−11.372
	[−9.949]	[−10.791]		
ethanol−propanoic acid	−9.954	−10.759		−11.143
	[−9.890]	[−10.805]		
propanol−formic acid	−10.09	−10.883	−11.365	−11.831
	[−10.021]	[−10.997]		
propanol−acetic acid	−10.094	−10.822		−11.285
	[−10.000]	[−10.978]		
propanol−propanoic acid	−9.954	−10.759		−11.302
	[−9.991]	[−10.963]		
butanol−formic acid	−10.163	−10.914		−11.178
	[−9.996]	[−10.862]		
butanol−acetic acid	−10.079	−10.845		−11.081
	[−9.993]			
butanol−propanoic acid	−10.035	−10.795		−11.052
	[−9.972]			

attractive energy components, we present the relative percentage contribution of each component shown in the parentheses. We see that there is a crossing of the relative electrostatic and dispersion components around the AAK−AAK groups. The main stabilization attractive energy contributions shift from the AAA−AAA groups (mainly dispersion bound) to the CAA−CAA groups (mainly hydrogen bonded).

As an application of the calculated SAPT energy data, we have proposed a segmental model where we further dissect a functional group molecule into chemically identified segments,
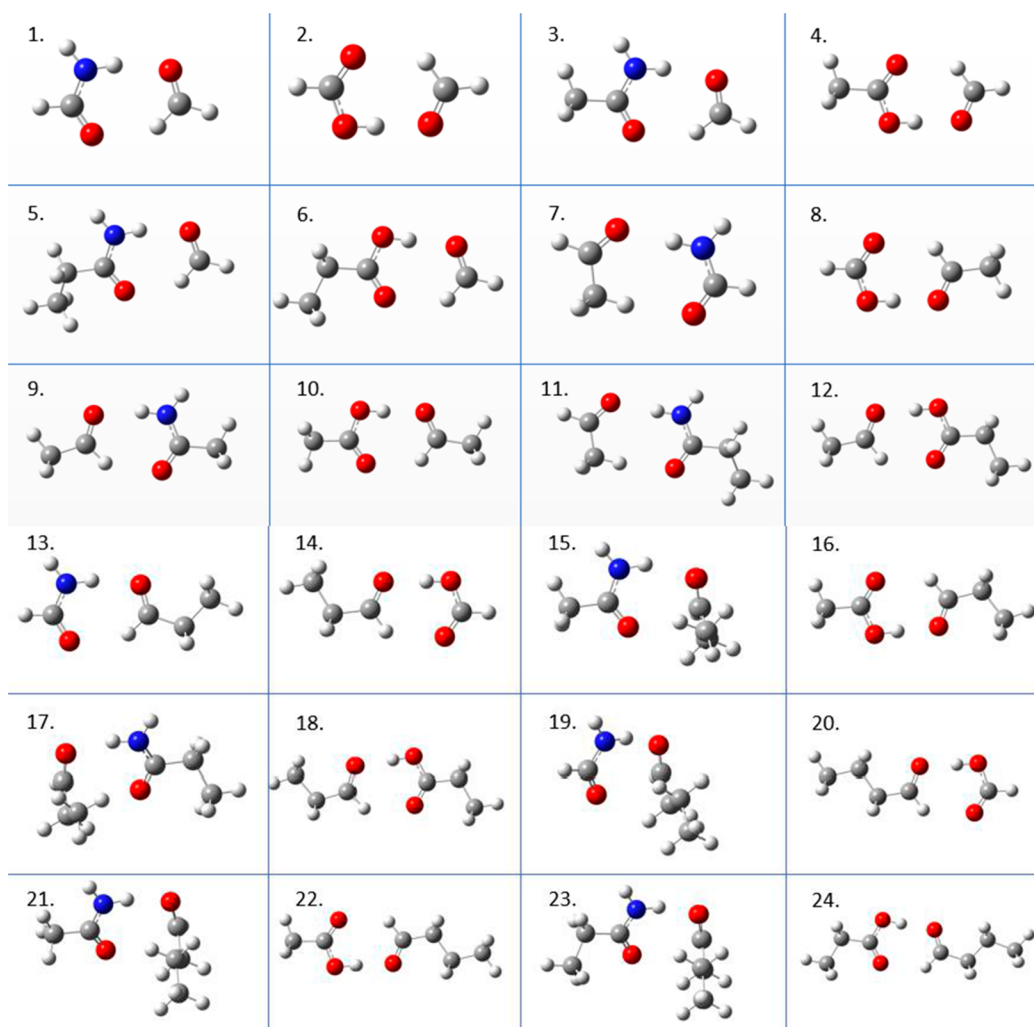
Figure 14. Optimized structures of the dimers in the Ad−Am and Ad−Ca series.

each as an effective united atom. To each segment we attribute electric features such as effective charges and geometrical features such as molecular volumes, so that the pair summed intersegment interactions can reproduce the SAPT component energies for the dimers. For repulsion and electrostatic interactions, formally the charge pair $(+, -)$ is counted as electrostatic, while the charge pairs $(+, +)$ and $(-, -)$ are counted as exchange. The induction energy is modeled as a charge−dipole interaction; that is, the charge at one segment interacts with the closest dipole at the other segment. The dispersion interaction is modeled by a power law with respect to the molecular volume[75,76] (see also the Supporting Information). In this way our model is similar to the usual fragment-based energy partition schemes, such as the recent A-SAPT and F-SAPT methods,[77−80] where the goal is to construct an effective two-body partition model of the SAPT energy components to localized chemically recognizable segments.

Let us illustrate the assignment of segments using a butane molecule (Figure S1). We dissect a butane molecule into four segments of two types, that is, A is the methyl radical (CH_3-) and B is the methylene radical $(-CH_2-)$, so a butane molecule is represented by $A^+B^-B^+A^-$, where we have assigned symbolically the alternating (positive−negative) charges on each segment. Next for each energy component, we simply count the suitable paired segments and list the interactions. Let us first

consider the ethane−propane dimer. Here we can count the suitable pairs for the electrostatic, induction, and exchange energies (in kcal/mol) as follows:

$$E_{elec} = 2E^{elec}_{AA} + E^{elec}_{AB} = -0.793$$

$$E_{ind} = 2E^{ind}_{A,A-B} + E^{ind}_{A,A-A} + E^{ind}_{B,A-A}$$

$$= -0.164 + E^{ind}_{B,A-A} = -0.196$$

$$E_{exch} = 2E^{exch}_{AA} + E^{exch}_{AB} = 2.261$$

Here we have used the previously determined segmental component energies from the SOFG-31 data set.[59] Only one unknown variable is to be determined, so we used the SAPT energy to obtain $E^{ind}_{B,A-A} = -0.032$. Continuing this procedure we can list similar energy equations with unknown intersegment energies to be determined sequentially. Please refer to the Supporting Information for the full list of supplementary figures, tables, and equations. It is found that for the alkane heterodimer series using the energy data up to the ethane−pentane dimer (the training set) is sufficient to sort out all the intersegment interactions for each energy component. The detailed analysis and calculations for the other molecules are shown in the Supporting Information. In Table S6 we summarize the resulting segmental SAPT energies. We can see that the model works

Table 17. Binding Energies of the Dimers in the Ad−Am Series

	MP2 [CCSD(T)]			CCSD(T)
	aDZ	aTZ	aQZ	CBS
formaldehyde−formamide	−7.215	−7.587	−7.814	−8.179
	[−7.324]	[−7.785]	[−8.013]	
formaldehyde−acetamide	−7.140	−7.504	−7.728	−8.126
	[−7.288]	[−7.739]		
formaldehyde−propionamide	−7.997	−8.233		−8.597
	[−8.193]	[−8.498]		
acetaldehyde−formamide	−7.173	−7.513	−7.724	−8.014
	[−7.226]	[−7.649]		
acetaldehyde−acetamide	−7.452	−7.825	−8.056	−8.270
	[−7.650]	[−8.109]		
acetaldehyde−propionamide	−7.889	−8.088		−8.380
	[−8.034]	[−8.296]		
propionaldehyde−formamide	−7.566	−7.946	−8.176	−8.606
	[−7.736]	[−8.205]		
propionaldehyde−acetamide	−7.181	−7.636		−7.893
	[−7.216]	[−7.701]		
propaldehyde−propionamide	−8.304	−8.682		−8.868
	[−8.331]			
butyraldehyde−formamide	−7.510	−8.023		−8.197
	[−7.426]	[−7.981]		
butyraldehyde−acetamide	−7.568	−8.056		−8.214
	[−7.521]			
butyaldehyde−propionamide	−8.508	−8.884		−9.028
	[−8.494]			

Table 18. Binding Energies of the Dimers in the Ad−Ca Series

	MP2 [CCSD(T)]			CCSD(T)
	aDZ	aTZ	aQZ	CBS
formaldehyde−formic acid	−8.442	−9.054	−9.329	−9.653
	[−8.460]	[−9.166]	[−9.452]	
formaldehyde−acetic acid	−8.064	−8.654	−8.927	−9.332
	[−8.160]	[−8.842]	[−9.125]	
formaldehyde−propion acid	−7.994	−8.576		−9.028
	[−8.109]	[−8.783]		
acetaldehyde−formic acid	−9.220	−9.871	−10.156	−10.499
	[−9.257]	[−10.006]		
acetaldehyde−acetic acid	−8.717	−9.341	−9.623	−10.057
	[−8.851]	[−9.569]		
acetaldehyde−propion acid	−8.641	−9.254		−9.761
	[−8.796]	[−9.503]		
propanal−formic acid	−9.257	−9.905	−10.187	−10.551
	[−9.314]	[−10.063]		
propanal−acetic acid	−8.737	−9.358		−9.871
	[−8.894]	[−9.610]		
propanal−propion acid	−8.661	−9.273		−9.710
	[−8.840]			
butanal−formic acid	−9.257	−9.940		−10.395
	[−9.359]	[−10.107]		
butanal−acetic acid	−8.766	−9.385		−9.813
	[−8.933]			
butanal−propion acid	−8.690	−9.301		−9.749
	[−8.881]			

surprisingly well. For most cases, we can reproduce the corresponding SAPT energies to an accuracy of about 10% errors. As a further test of the validity of this model, let us predict the binding energies for larger heterodimers. For example, consider the undecane−dodecane heterodimer. We see that there are an additional 3 pairs of (A, B) and 17 pairs of (B, B) electrostatic interactions, 4 pairs of (A, A-B)/(B, A-B) and 33 pairs of (B, B−B) charge−dipole interactions, and 1 pair of (A, B) and 9 pairs of (B, B) exchange interactions. By simply counting the pairs we can list the energy equations as follows:
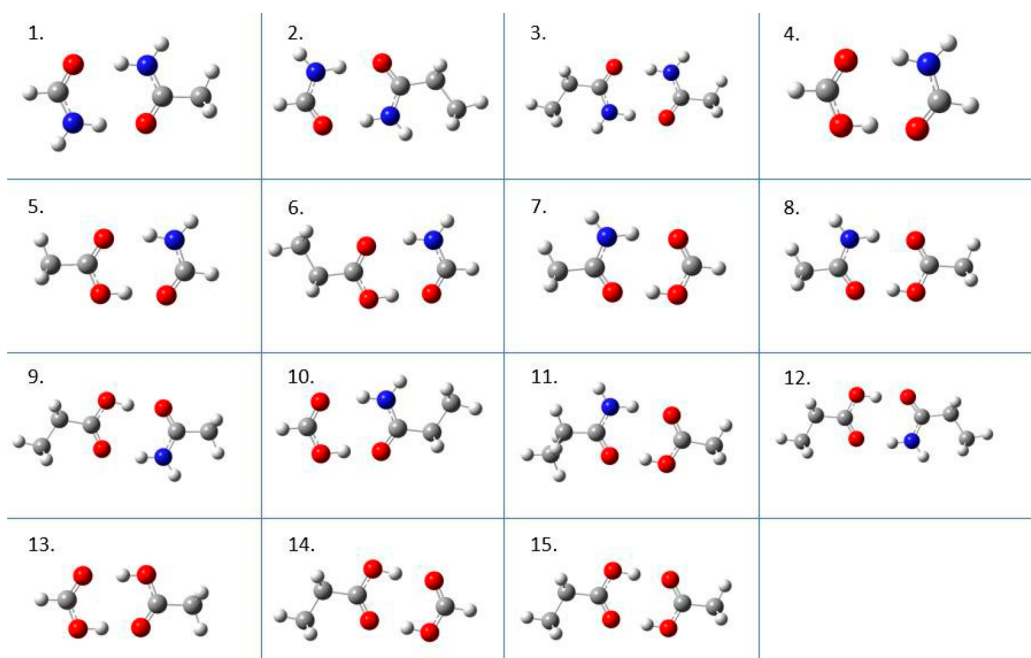
Figure 15. Optimized structures of the dimers in the Am−Am, Ca−Ca, and Am−Ca series

Table 19. Binding Energies of the Dimers in the Am−Am, Ca−Ca, and Am−Ca Series

	MP2 [CCSD(T)]			CCSD(T)
	aDZ	aTZ	aQZ	CBS
formamide−acetamide	−13.692	−14.318	−14.677	−15.139
	[−13.761]	[−14.518]		
formamide−propionamide	−13.847	−14.487	−14.845	−15.315
	[−13.923]	[−14.696]		
acetamide−propionamide	−13.958	−14.590		−15.110
	[−14.081]	[−14.844]		
formamide−formic acid	−14.314	−15.278	−15.688	−16.206
	[−14.269]	[−15.377]	[−15.817]	
formamide−acetic acid	−14.015	−14.952	−15.364	−15.854
	[−14.065]	[−15.141]		
formamide−propanoic acid	−13.919	−14.846		15.450
	[−13.994]	[−15.060]		
acetamide−formic acid	−14.997	−15.986	−16.399	−16.807
	[−14.958]	[−16.093]		
acetamide−acetic acid	−14.598	−15.557		−16.168
	[−14.669]	[−15.764]		
acetamide−propanoic acid	−14.496	−15.444		−16.081
	[−14.594]	[−15.682]		
propanoamide−formic acid	−15.270	−16.283		−16.822
	[−15.232]	[−16.395]		
propanoamide−acetic acid	−15.431	−16.273		−16.847
	[−15.546]	[−16.492]		
propanoamide−propic acid	−14.742	−15.711		−15.924
	[−14.848]			
formic acid−acetic acid	−14.334	−15.581	−16.041	−16.508
	[−14.324]	[−15.712]		
formic acid−propanoic acid	−14.296	−15.533		−16.208
	[−14.308]	[−15.687]		
acetic acid−propanoic acid	−14.449	−15.683		−16.437
	[−14.545]	[−15.917]		

Table 20. Comparison of the Reference MP2/CBS Energy Data[81] and the Model Predicted Energies (kcal/mol) Using the Segment SAPT Analysis

complex	MP2/CBS	model	error %
heptane−octane	−5.44	−5.77	−6.1%
octane−nonane	−5.81	−6.13	−5.5%
nonane−decane	−6.66	−6.43	+3.1%
decane−undecane	−6.98	−6.74	+3.4%
undecane−dodecane	−8.33	−8.62	−3.4%

$$E_{elec} = 3E_{AB} + E_{AA} + 17E_{BB} = -4.191$$

$$E_{ind} = 4E_{A,A-B} + 4E_{B,A-B} + 33E_{B,B-B} = -1.229$$

$$E_{exch} = E_{AA} + 9E_{BB} + E_{AB} = 12.105$$

$$E_{disp} = -15.305$$

Our predicted SAPT energy is −8.62, which can be compared with the MP2/CBS value of −8.33 calculated by the Hobza group.[81] As can be seen, in this case we accidentally obtain a very accurate energy with only a 3% error off the reference value. The application of this model to other heterodimers is shown and compared to their MP2/CBS values in Table 20. We see that the overall performance is very good. Therefore, it is promising to utilize this model in coarse-grained molecular modeling for larger molecules.

4. CONCLUSION AND OUTLOOK

We have constructed a minimum-level CCSD(T)/CBS-calculated interaction energy data set with the MP2/aug-cc-pVXZ (X = D, T, and up to Q) optimized geometries for 239 heterodimers of small organic functional groups. The monomers are selected from the SOFG-31 data set, including the alkane, alkene, alkyne, alcohol, aldehyde, ketone, carboxylic acid, and amide groups. Together with the SOFG-31 set, this extended set is called the SOFG-31+239 (SOFG-270) data set. The MP2/aug-cc-pVTZ level of theory is reliable for the geometry optimization, and the CCSD(T)/CBS binding energies can serve as benchmark reference data which supplements and/or complements existing data sets. Overall, a chemical accuracy (∼0.1 kcal/mol), consistent for each individual noncovalent complex, can be assigned with this data set. A comprehensive SAPT analysis is also performed in order to gain more chemical insights into the calculated full interaction energies. A further segment modeling provides finer details of the segmental contributions for each molecule. These segmental energy "quanta" can then be used to predict intermolecular interaction energies for large molecules and in the construction of coarse-grained force fields for molecular simulations.

This minimum set of energy data can be enlarged along with available computer resources. However, the scope is limited with the most stable conformations of each pair of monomers. To reach our goal of constructing a universal force field without empirical inputs, we need the full potential energy surfaces. One standard way is to sample a set of relative orientations of the paired monomers and scan the corresponding potential energy curves at a sequence of distance points along the dissociation coordinates. The computational cost is roughly proportional to the sample number of orientations and the number of scanning points, respectively. Within current computer capacity attained similar in this work, this task can be routinely studied. A further complication for larger organic functional groups is the issue of

isomers, both for monomers and dimers. It is well-known that the most stable dimer is not necessarily formed by the most stable monomers. Therefore, there may exist several local stable complexes which are related through isomerization pathways. The computational costs are expected to be quite intense because the number of isomers for a specific pair of monomers increases combinatorially fast. Apparently we are just toeing the (starting) line. A considerable amount of computer resources and human collaboration is required in this fundamental and important subfield of computational chemistry.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c01888.

Detailed SAPT energies for the 239 dimers; complete segment SAPT analysis including model construction, algebraic equations, and results (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Sheng D. Chao − *Institute of Applied Mechanics, National Taiwan University, Taipei 10617 Taiwan, R.O.C.;* ⓞ orcid.org/0000-0002-5896-6700; Email: sdchao@spring.iam.ntu.edu.tw

Authors

Hsing-Hsiang Huang − *Institute of Applied Mechanics, National Taiwan University, Taipei 10617 Taiwan, R.O.C.*

Yi-Siang Wang − *School of Chemistry & Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c01888

Notes

The authors declare no competing financial interest.
The data that supports the findings of this study are available within the article and its Supporting Information.

■ REFERENCES

(1) Gieseler, J.; Heber, B.; Herbst, K. An Empirical Modification of the Force Field Approach to Describe the Modulation of Galactic Cosmic Rays Close to Earth in a Broad Range of Rigidities. *J. Geophys. Res. Space Phys.* **2017**, *122* (11), 10964−10979.

(2) Ma, R.; Huang, D.; Zhang, T.; Luo, T. Determining influential descriptors for polymer chain conformation based on empirical force-fields and molecular dynamics simulations. *Chem. Phys. Lett.* **2018**, *704*, 49−54.

(3) Santra, S.; Kundu, T.; Jana, M. Microscopic investigation on empirical force-field model dependent structure and dynamical properties of amino acids in aqueous medium. *J. Indian Chem. Soc.* **2018**, *95* (12), 1617−1630.

(4) Stroet, M.; Koziara, K. B.; Malde, A. K.; Mark, A. E. Optimization of empirical force fields by parameter space mapping: A single-step perturbation approach. *J. Chem. Theory Comput.* **2017**, *13* (12), 6201−6212.

(5) Van der Spoel, D.; Ghahremanpour, M. M.; Lemkul, J. A. Small molecule thermochemistry: a tool for empirical force field development. *J. Phys. Chem. A* **2018**, *122* (45), 8982−8988.

(6) Philips, A.; Marchenko, A.; Ducati, L. C.; Autschbach, J. Quadrupolar 14N NMR Relaxation from Force-Field and Ab Initio Molecular Dynamics in Different Solvents. *J. Chem. Theory Comput.* **2019**, *15* (1), 509−519.

(7) Xu, P.; Guidez, E. B.; Bertoni, C.; Gordon, M. S. Perspective: Ab initio force field methods derived from quantum mechanics. *J. Chem. Phys.* **2018**, *148* (9), 090901.

(8) Chung, Y. H.; Li, A. H. T.; Chao, S. D. Computer simulation of trifluoromethane properties with ab initio force field. *J. Comput. Chem.* **2011**, *32* (11), 2414−2421.

(9) Higgoda, U. A.; Hellmann, R.; Koller, T. M.; Fröba, A. P. Self-diffusion coefficient and viscosity of methane and carbon dioxide via molecular dynamics simulations based on new ab initio-derived force fields. *Fluid Ph. Equilibria* **2019**, *481*, 15−27.

(10) Li, A. H. T.; Chao, S. D.; Chang, C. C. Determination of a silane intermolecular force field potential model from an ab initio calculation. *Phys. Rew. A* **2010**, *82* (6), 062520.

(11) Li, A. H. T.; Huang, S. C.; Chao, S. D. Molecular dynamics simulation of liquid carbon tetrachloride using ab initio force field. *J. Chem. Phys.* **2010**, *132* (2), 024506.

(12) Pereyaslavets, L.; Kurnikov, I.; Kamath, G.; Butin, O.; Illarionov, A.; Leontyev, I.; Olevanov, M.; Levitt, M.; Kornberg, R.; Fain, B. On the importance of accounting for nuclear quantum effects in ab initio calibrated force fields in biological simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (36), 8878−8882.

(13) Pérez-Conesa, S.; Torrico, F.; Martínez, J. M.; Pappalardo, R. R.; Marcos, E. S. A general study of actinyl hydration by molecular dynamics simulations using ab initio force fields. *J. Chem. Phys.* **2019**, *150* (10), 104504.

(14) Wang, S. B.; Li, A. H. T.; Chao, S. D. Liquid properties of dimethyl ether from molecular dynamics simulations using ab initio force fields. *J. Comput. Chem.* **2012**, *33* (9), 998−1003.

(15) Rezac, J.; Huang, Y.; Hobza, P.; Beran, G. J. Benchmark calculations of three-body intermolecular interactions and the performance of low-cost electronic structure methods. *J. Chem. Theory Comput.* **2015**, *11* (7), 3065−3079.

(16) McDaniel, J. G.; Choi, E.; Son, C. Y.; Schmidt, J. R. Ab Initio Force Fields for Imidazolium-Based Ionic Liquids. *J. Phys. Chem. B* **2016**, *120*, 7024−7036.

(17) Rackers, J. A.; Liu, C.; Ren, P.; Ponder, J. W. A physically grounded damped dispersion model with particle mesh Ewald summation. *J. Chem. Phys.* **2018**, *149*, 084115.

(18) Gökcan, H.; Kratz, E.; Darden, T. A.; Piquemal, J.-P.; Cisneros, G. A. QM/MM Simulations with the Gaussian Electrostatic Model: A Density-based Polarizable Potential. *J. Phys. Chem. Lett.* **2018**, *9*, 3062−3067.

(19) Antony, J.; Grimme, S. Density functional theory including dispersion corrections for intermolecular interactions in a large benchmark set of biologically relevant molecules. *Phys. Chem. Chem. Phys.* **2006**, *8* (45), 5287−5293.

(20) Grimme, S.; Waletzke, M. A combination of Kohn−Sham density functional theory and multi-reference configuration interaction methods. *J. Chem. Phys.* **1999**, *111* (13), 5645−5655.

(21) Grimme, S. Supramolecular binding thermodynamics by dispersion-corrected density functional theory. *Chem.—Eur. J.* **2012**, *18* (32), 9955−9964.

(22) Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.* **2017**, *115* (19), 2315−2372.

(23) Hait, D.; Head-Gordon, M. How accurate is density functional theory at predicting dipole moments? An assessment using a new database of 200 benchmark values. *J. Chem. Theory Comput.* **2018**, *14* (4), 1969−1981.

(24) Witte, J.; Neaton, J. B.; Head-Gordon, M. Push it to the limit: Characterizing the convergence of common sequences of basis sets for intermolecular interactions as described by density functional theory. *J. Chem. Phys.* **2016**, *144* (19), 194306.

(25) Kodrycka, M.; Patkowski, K. Platinum, gold, and silver standards of intermolecular interaction energy calculations. *J. Chem. Phys.* **2019**, *151*, 070901.

(26) Schneebeli, S. T.; Bochevarov, A. D.; Friesner, R. A. Parameterization of a B3LYP Specific Correction for Noncovalent Interactions and Basis Set Superposition Error on a Gigantic Data Set of CCSD(T) Quality Noncovalent Interaction Energies. *J. Chem. Theory Comput.* **2011**, *7*, 658−668.

(27) Mardirossian, N.; Head-Gordon, M. ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **2016**, *144*, 214110.

(28) Smith, D. G. A.; Burns, L. A.; Patkowski, K.; Sherrill, C. D. Revised Damping Parameters for the D3 Dispersion Correction to Density Functional Theory. *J. Phys. Chem. Lett.* **2016**, *7*, 2197−2203.

(29) Burns, L. A.; Faver, J. C.; Zheng, Z.; Marshall, M. S.; Smith, D. G. A.; Vanommeslaeghe, K.; MacKerell, A. D., Jr.; Merz, K. M., Jr.; Sherrill, C. D. The BioFragment Database (BFDb): An open-data platform for computational chemistry analysis of noncovalent interactions. *J. Chem. Phys.* **2017**, *147*, 161727.

(30) McGibbon, R. T.; Taube, A. G.; Donchev, A. G.; Siva, K.; Hernández, F.; Hargus, C.; Law, K.-H.; Klepeis, J. L.; Shaw, D. E. Improving the accuracy of Møller-Plesset perturbation theory with neural networks. *J. Chem. Phys.* **2017**, *147*, 161725.

(31) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184−32215.

(32) Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets: Hydrogen Bonding. *J. Chem. Theory Comput.* **2020**, *16*, 2355−2368.

(33) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7* (8), 2427−2438.

(34) Hohenstein, E. G.; Sherrill, C. D. Wavefunction methods for noncovalent interactions. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2* (2), 304−326.

(35) Goerigk, L.; Grimme, S. A General Database for Main Group Thermochemistry, Kinetics, and Noncovalent Interactions - Assessment of Common and Reparameterized (meta-)GGA Density Functionals. *J. Chem. Theory Comput.* **2010**, *6* (1), 107−126.

(36) Burns, L. A.; Marshall, M. S.; Sherrill, C. D. Comparing Counterpoise-Corrected, Uncorrected, and Averaged Binding Energies for Benchmarking Noncovalent Interactions. *J. Chem. Theory Comput.* **2014**, *10*, 49−57.

(37) Marshall, M. S.; Burns, L. A.; Sherrill, C. D. Basis set convergence of the coupled-cluster correction, δ-CCSD(T)-MP2: Best practices for benchmarking non-covalent interactions and the attendant revision of the S22, NBC10, HBC6, and HSG databases. *J. Chem. Phys.* **2011**, *135*, 194102.

(38) Granatier, J.; Pitoňák, M.; Hobza, P. Accuracy of Several Wave Function and Density Functional Theory Methods for Description of Noncovalent Interaction of Saturated and Unsaturated Hydrocarbon Dimers. *J. Chem. Theory Comput.* **2012**, *8*, 2282−2292.

(39) Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19* (4), 553−566.

(40) Van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G.; van Lenthe, J. H. State of the Art in Counterpoise Theory. *Chem. Rev.* **1994**, *94* (7), 1873−1885.

(41) Gresh, N. Energetics of Zn²⁺ binding to a series of biologically relevant ligands − A molecular mechanics investigation grounded on ab-initio self-consistent-field supermolecular computations. *J. Comput. Chem.* **1995**, *16* (7), 856−882.

(42) Gresh, N.; Leboeuf, M.; Salahub, D. Energetics and structure in model neutral, anionic and cationic hydrogen-bonded complexes − combined ab-initio SCF MP2 supermolecular, density-functional, and

molecular mechanics. In *Modeling the Hydrogen Bond*; Smith, D. A., Ed.; 1994; pp 82−112.

(43) Kóňa, J.; Zahradník, P.; Kozmon, S.; Fabian, W. M. Role of solvent effects on nucleophilic substitution of 4H-pyran-4-one and its 2, 6-dimethyl derivative with hydroxide ion in aqueous solution: ab initio and density functional theory studies on a supermolecular reaction model. *J. Mol. Struc.:THEOCHEM* **2005**, *728* (1−3), 117−122.

(44) Marchetti, O.; Werner, H.-J. Accurate Calculations of Intermolecular Interaction Energies Using Explicitly Correlated Coupled Cluster Wave Functions and a Dispersion-Weighted MP2Method. *J. Phys. Chem. A* **2009**, *113* (43), 11580−11585.

(45) 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.

(46) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. Quadratic configuration interaction. A general technique for determining electron correlation energies. *J. Chem. Phys.* **1987**, *87* (10), 5968−5975.

(47) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-set convergence of correlated calculations on water. *J. Chem. Phys.* **1997**, *106* (23), 9639−9646.

(48) Benchmark Energy & Geometry Database (BEGDB). http://www.begdb.org/index.php.

(49) Gráfová, L.; Pitonak, M.; Rezac, J.; Hobza, P. Comparative study of selected wave function and density functional methods for noncovalent interaction energy calculations using the extended S22 data set. *J. Chem. Theory Comput.* **2010**, *6* (8), 2365−2376.

(50) Takatani, T.; Hohenstein, E. G.; Malagoli, M.; Marshall, M. S.; Sherrill, C. D. Basis set consistent revision of the S22 test set of noncovalent interaction energies. *J. Chem. Phys.* **2010**, *132* (14), 144104.

(51) Hobza, P. The calculation of intermolecular interaction energies. *Annu. Rep. Prog. Chem., Sect. C* **2011**, *107*, 148−168.

(52) Řezáč, J.; Hobza, P. Describing Noncovalent Interactions beyond the Common Approximations: How Accurate Is the "Gold Standard," CCSD(T) at the Complete Basis Set Limit? *J. Chem. Theory Comput.* **2013**, *9*, 2151−2155.

(53) Non-Covalent Interactions Atlas. http://www.nciatlas.org/.

(54) Khaliullin, R. Z.; Cobar, E. A.; Lochan, R. C.; Bell, A. T.; Head-Gordon, M. Unravelling the origin of intermolecular interactions using absolutely localized molecular orbitals. *J. Phys. Chem. A* **2007**, *111* (36), 8753−8765.

(55) Pitoňák, M.; Neogrády, P.; Černý, J.; Grimme, S.; Hobza, P. Scaled MP3 non-covalent interaction energies agree closely with accurate CCSD(T) benchmark data. *ChemPhysChem.* **2009**, *10* (1), 282−289.

(56) Hyla-Kryspin, I.; Haufe, G.; Grimme, S. Weak Hydrogen Bridges: A Systematic Theoretical Study on the Nature and Strength of C-H··· F-C Interactions. *Chem.—Eur. J.* **2004**, *10* (14), 3411−3422.

(57) Donchev, A. G.; Taube, A. G.; Decolvenaere, E.; Hargus, C.; McGibbon, R. T.; Law, K. H.; Gregersen, B. A.; Li, J.-L.; Palmo, K.; Shaw, D. E.; et al. Quantum chemical benchmark databases of gold-standard dimer interaction energies. *Scientific data* **2021**, *8* (1), 55.

(58) Smith, D. G.; Altarawy, D.; Burns, L. A.; Welborn, M.; Naden, L. N.; Ward, L.; Ellis, S.; Pritchard, B. P.; Crawford, T. D. The MolSSI QCArchive project: An open-source platform to compute, organize, and share quantum chemistry data. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2021**, *11* (2), No. e1491.

(59) Chang, Y. M.; Wang, Y. S.; Chao, S. D. A minimum quantum chemistry CCSD (T)/CBS dataset of dimeric interaction energies for small organic functional groups. *J. Chem. Phys.* **2020**, *153* (15), 154301.

(60) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Fox, D. J. et al., *Gaussian 09*, rev. D. 01.; Gaussian, Inc.: Wallingford, CT, 2009.

(61) East, A. L.; Allen, W. D. The heat of formation of NCO. *J. Chem. Phys.* **1993**, *99* (6), 4638−4650.

(62) Jeziorski, B.; Moszynski, R.; Szalewicz, K. J. C. R. Perturbation theory approach tointermolecular potential energy surfaces of van der Waals complexes. *Chem. Rev.* **1994**, *94* (7), 1887−1930.

(63) Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Crawford, T. D.; et al. Psi4: an open-source ab initio electronic structure program. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2* (4), 556−565.

(64) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; et al. Estimated MP2 and CCSD (T) interaction energies of n-alkane dimers at the basis set limit: Comparison of the methods of Helgaker and Feller. *J. Chem. Phys.* **2006**, *124* (11), 114304.

(65) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M. Magnitude of interaction between n-alkane chains and its anisotropy: High-level ab initio calculations of n-butane, n-petane, and n-hexane dimers. *J. Phys. Chem. A* **2004**, *108* (46), 10311−10316.

(66) Li, A. H.-T.; Chao, S. D. A Refined Intermolecular Interaction Potential for Methane: Spectral Analysis and Molecular Dynamics Simulations. *J. Chin. Chem. Soc.* **2016**, *63*, 282−289.

(67) Chao, S. W.; Li, A. H.-T.; Chao, S. D. Molecular dynamics simulations of fluid methane properties using ab initio intermolecular interaction potentials. *J. Comput. Chem.* **2009**, *30*, 1839−1849.

(68) Li, A. H.-T.; Chao, S. D. Interaction energies of dispersion-bound methane dimer from coupled cluster method at complete basis set limit. *J. Mol. Struc.:THEOCHEM* **2009**, *897*, 90−94.

(69) Li, A. H.-T.; Chao, S. D. Intermolecular potentials of the methane dimer calculated with M?ller-Plesset perturbation theory and density functional theory. *J. Chem. Phys.* **2006**, *125*, 094312.

(70) Echeverria, J.; Aullon, G.; Danovich, D.; Shaik, S.; Alvarez, S. Dihydrogen contacts in alkanes are subtle but not faint. *Nat. Chem.* **2011**, *3*, 323−330.

(71) Fokin, A. A.; Gerbig, D.; Schreiner, P. R. σ/σ- And π/π-interactions are equally important: multilayered graphanes. *J. Am. Chem. Soc.* **2011**, *133*, 20036−20039.

(72) Danovich, D.; Shaik, S.; Neese, F.; Echeverria, J.; Aullon, G.; Alvarez, S. Understanding the Nature of the CH···HC Interactions in Alkanes. *J. Chem. Theory Comput.* **2013**, *9*, 1977−1991.

(73) Zeng, Z. Y.; Wang, Y. S.; Chao, S. D. Hydrogen bonded dimers of small alkyl substituted amides: Structures, energetics, and spectral analyses based on density functional theory calculations. *Comput. Theor. Chem.* **2017**, *1113*, 1−7.

(74) Parker, T. M.; Burns, L. A.; Parrish, R. M.; Ryno, A. G.; Sherrill, C. D. Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies. *J. Chem. Phys.* **2014**, *140*, 094106.

(75) Stone, A. J. *The theory of intermolecular forces*; Oxford University Press: Oxford, 2013.

(76) Sperling, L. H. *Introduction to Physical Polymer Sciences*; Wiley: NewYork, 2001.

(77) Parrish, R. M.; Parker, T. M.; Sherrill, C. D. Chemical Assignment of Symmetry-Adapted Perturbation Theory Interaction Energy Components: The Functional-Group SAPT Partition. *J. Chem. Theory Comput.* **2014**, *10*, 4417.

(78) Parrish, R. M.; Sherrill, C. D. Spatial assignment of symmetry adapted perturbation theory interaction energy components: The atomic SAPT partition. *J. Chem. Phys.* **2014**, *141*, 044115.

(79) Parrish, R. M.; Gonthier, J. F.; Corminboeuf, C.; Sherrill, C. D. Communication: Practical intramolecular symmetry adapted perturbation theory via Hartree-Fock embedding. *J. Chem. Phys.* **2015**, *143*, 051103.

(80) Parrish, R. M.; Sherrill, C. D. Quantum-Mechanical Evaluation of $\pi−\pi$ versus Substituent-π Interactions in π Stacking: Direct Evidence for the Wheeler−Houk Picture. *J. Am. Chem. Soc.* **2014**, *136*, 17386.

(81) Sedlak, R.; Janowski, T.; Pitoňák, M.; Řezáč, J.; Pulay, P.; Hobza, P. Accuracy of quantum chemical methods for large noncovalent complexes. *J. Chem. Theory Comput.* **2013**, *9* (8), 3364−3374.