ChemPhysChem

Research Article
doi.org/10.1002/cphc.202300027

Chemistry Europe
European Chemical Societies Publishing

www.chemphyschem.org

Accurate Interaction Energies of CO_2 with the 20 Naturally Occurring Amino Acids

Amarachi G. Sylvanus[a] and Konstantinos D. Vogiatzis*[a]

We have performed a series of highly accurate calculations between CO_2 and the 20 naturally occurring amino acids for the investigation of the attractive noncovalent interactions. Different nucleophilic groups present in the amino acid structures were considered (α-NH_2, COOH, side groups), and the stronger binding sites were identified. A database of accurate reference interactions energies was compiled as computed by explicitly-correlated coupled-cluster singles-and-doubles, together with perturbative triples extrapolated to the complete-basis-set limit. The CCSD(F12)(T)/CBS reference values were used for comparing a variety of popular density functionals with different basis sets. Our results show that most density functionals with the triple-zeta basis set def2-TZVPP align with the CCSD(F12)(T)/CBS reference values, but errors range from 0.1 kcal/mol up to 1.0 kcal/mol.

Introduction

Carbon dioxide (CO_2) is the prime greenhouse gas produced mainly from transportation, power plants, chemical plants, and other industrial processes, and has become a major concern across the globe as a prime cause of extreme climate changes.[1] CO_2 capture and separation comprise several technologies that capture, store and transport CO_2 at various stages from point sources.[2] On an industrial scale, amine-based solvents are largely used for CO_2 capture via chemisorption, which is an energy-intensive and prone to corrosion process.[3] This is an outcome of the high cost of solvent regeneration, the corrosive nature of the byproducts, while the high vapor pressure of the solvent leads to release of toxic amine gases into the atmosphere upon heating. In addition, the high energy required to remove the adsorbed CO_2 and break the carbamide bond formed between the CO_2 and the amine defeats the whole purpose of carbon reduction, towards mitigating climate change. This has propelled several researchers to seek alternatives to the capture of CO_2 via conventional chemisorption processes. These alternative methods range from the use of membrane,[4] porous materials as adsorbents such as zeolites and metal-organic frameworks (MOFs),[5] carbon molecular sieves,[6,7] cryogenic methods,[8] and ionic liquids.[9] These techniques are mostly used to capture CO_2 from post-combustion industrial sources. Although there are different methodologies for CO_2 capture, the challenge remains to find a balance between sustainable, cost-effective, and environmentally friendly technologies. Several studies have proposed the use of bio-inspired materials for the selective capture of CO_2, where amino acids (AAs), the building blocks of enzymes, can effectively bind CO_2 as suitable alternatives to conventional amine-based solvent capture.[10–15] Another experimental study explored the adsorption of CO_2 on solid AAs inside a thermal reactor.[16] In addition, AAs have low toxicity,[17] are non-volatile,[18] have high resistance to degradation,[19] and could be produced from various bio-sources.[20,21]

The 20 naturally occurring AAs are the building blocks of enzymes that catalyze biochemical reactions. They are composed of a carboxylic acid group and an amino group while they differ by their side groups (*R*) which can contain among others alkyl, aromatic, hydroxy, amine, or carboxylic acid groups. The presence of nucleophilic side chains makes AAs promising units for selective interactions with CO_2. The formation of weak, noncovalent interactions would eliminate the high regenerative cost of separating CO_2 from the conventional capture media that requires stronger carbamide bonds with CO_2. Also, the variety of nucleophilic groups on AAs offer synergistic binding with 2:1 ratio between AAs and one CO_2 molecule, which further increases the efficacy of the physisorption process. These AAs-based separation techniques can be used in post-combustion processes (high composition of CO_2), like amine solvents, or in direct air capture (DAC) processes by physisorption. The DAC can be used to capture CO_2 from smaller-scale and mobile sources like transportation media. DAC also bypasses the problem of storage and transportation of CO_2 and therefore can be incorporated into technologies that directly use CO_2 as feedstock.[22,23]

Although there are numerous reports on amino acid-functionalized MOFs,[24–31] amino acid salts and amino acid ionic liquids,[18,32] there are limited available theoretical data on the interaction strength of isolated AAs with CO_2. Hussain et al. reported the strength of the covalent and non-covalent interactions of the 20 naturally occurring AAs with CO_2 at various interaction sites.[33] This computational study utilized density functional theory (DFT) with the M05-2X exchange-correlational functional. A more recent work illustrated how the introduction of an extra carboxylate group on the AAs can increase their interaction with CO_2.[34] A combination of spectroscopic methods and quantum chemical calculations

[a] A. G. Sylvanus, Prof. Dr. K. D. Vogiatzis
Department of Chemistry, University of Tennessee, Knoxville, Tennessee 37996-1600, USA
E-mail: kvogiatz@utk.edu

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cphc.202300027

demonstrated how these groups could reduce the negative inductive effect of an amino group and accelerate the interaction of the molecule with CO_2. Another computational study focused on the chemisorption of CO_2 on amino acid ionic liquids using molecular dynamics (MD) simulations and DFT calculations performed with the M06-2X functional.[35]

In this work, the non-covalent interaction energies of the CO_2–AA molecular systems are explored with accurate quantum chemical methods. For this purpose, we have computed reference interaction energies using highly accurate explicitly correlated coupled-cluster singles-and-doubles with perturbative triples (abbreviated as CCSD(F12)(T)) at the complete basis set (CBS) limit.[36,37] We have explored a variety of possible CO_2 binding sites per AA and the generated reference data are used for benchmarking common density functionals. The theoretical methods used in this study are presented in Section "Computational Methods", while Section "Results and Discussion" introduces the CCSD(F12)(T)/CBS reference results. A detailed discussion for the CO_2–serine system is presented, followed by the analysis of the full CO_2–AA molecular dataset and the DFT benchmark study. Conclusions are provided in the last Section which will serve as the basis for future computational examination of biologically inspired macrostructures and materials for AA-based CO_2 separations. Our intention is to establish with this study an accurate computational procedure that can be applied on future studies between CO_2 and small oligopeptides for cooperative CO_2 binding.

Computational Methods

Geometry Optimizations

Initial structures were generated with the OPLS-AA force field in a periodic box at 200 K in the NVT ensemble with a Nosé-Hoover thermostat using a 1.0 ps time constant and a 1.0 fs time step.[38] All the force-field inputs were generated using the LigParGen software.[39] The simulation was carried out using the Large Atomic/Molecular Massively Parallel Simulator (LAMMPS)[40] software package in 10,000 iterative steps. The 30 most stable conformations were selected and optimized using DFT. The PBE0 functional[41] with Grimme's dispersion correction (D3),[42] the Becke-Johnson (BJ) damping function, the resolution of identity (RI) approximation,[43] and the def2-TZVPP[44] basis set were used. All DFT calculations were performed with the TURBOMOLE 7.2.1[45] quantum chemical program package. This hybrid MD/DFT scheme allows the generation of molecular structures without user intervention and bias, it significantly enhances the probability to obtain the most stable conformer for a given supersystem, while it has been tested and successfully applied in previous studies on noncovalent interactions between CO_2 and a variety of organic molecules.[46,47]

To further assess the optimized DFT geometries, we selected four cases (arginine, asparagine, methionine, tyrosine) and we performed a scan along the AA–CO_2 coordinate. We selected the distance $R_{eq,DFT}$ between the alpha nitrogen atom (α-NH_2) of these four AAs and the carbon atom C(CO_2) of CO_2, and we reoptimized the full AA–CO_2 supersystem for $R_{eq,DFT} \pm 0.02$ Å, $R_{eq,DFT} \pm 0.04$ Å, and $R_{eq,DFT} + 0.06$ Å, by keeping the position of α-NH_2 and C(CO_2) atoms fixed (5 calculations for each AA–CO_2 case). These constrained geometry optimizations were performed at the PBE0-D3(BJ)/def2-TZVPP level, and verified the optimized unconstrained DFT geometries that were used for the reference CCSD(F12)(T) calculations (*vide infra*).

Explicitly Correlated Coupled-Cluster Calculations

The conventional CCSD(T) method has a strong basis set dependence, leading to slower convergence with increasing basis set size. Slow convergence can be addressed by using F12 methods, which include terms in the wavefunction that explicitly depend on the interelectronic distance r_{12}.[48–50] The CCSD(F12)(T) calculations were performed using the cc-pV*X*Z-F12[51,52] basis sets (X = D, T) and the corresponding complementary auxiliary basis sets (CABS). The cc-pV*X*Z-F12 auxiliary basis sets were used to fit the F12 and electron-repulsion integrals (CBAS) as well as the two-electron contributions to the Fock matrix (JKBAS). The 2B ansatz was used in all F12 calculations.[53] The aug-cc-pV*X*Z basis sets[54] (X = D, T) with the corresponding CABS, CBAS and JKBAS auxiliary basis sets were tested for one molecular system (serine + CO_2).

The perturbative triples (T) energy term was computed from both conventional CCSD(T) amplitudes, which is abbreviated as $(T)_{con}$ in our analysis, and explicitly-correlated CCSD(F12)(T) amplitudes, which we will refer as $(T)_{F12}$. For the estimation of the (T) at the CBS limit, we have applied the two-point formula of Helgaker and coworkers[55] (Eq. (1)):

$$\delta E_{(T)_i/CBS} = \frac{\delta E_{(T)_i/XZ} X^3 - \delta E_{(T)_i/YZ} Y^3}{X^3 - Y^3} \quad (1)$$

where $(T)_i$ is either $(T)_{con}$ or $(T)_{F12}$.

The reference electronic energies for all molecular structures were computed by summation of the terms shown on Eq. (2):

$$E_{CCSD(F12)(T)/CBS} = E_{HF/TZ} + \delta E_{CABS\ S/TZ} + \delta E_{CCSD(F12)/TZ} + \delta E_{(T)_{F12}/CBS} \quad (2)$$

where $E_{HF/TZ}$ is the Hartree-Fock (HF) energy while the δE terms represent electron correlation energies. The term "CABS S" represents the CABS singles correction to the HF energy.[56] The cc-pVDZ-F12 and cc-pVTZ-F12 basis sets are abbreviated by their cardinal numbers X = 2 and Y = 3, respectively. All CCSD(F12)(T) calculations were performed with the TURBOMOLE 7.2.1[45] software package.

DFT Benchmarking Calculations

DFT geometry optimizations for all 20 AA–CO_2 supersystems, AA monomers, and isolated CO_2 were performed using TURBOMOLE[45] 7.2.1 software package with the D3(BJ) dispersion correction, and the RI approximation. In this study, 13 density functionals were tested and the obtained interaction energies (E_{int}) were compared to the highly accurate reference CCSD(F12)(T)/CBS energies. The density functionals used in the study are: PBE,[57] BP86,[58] BLYP,[59,60] TPSS,[61] PW6B95, BHLYP,[62] PBE0,[41] TPSSH,[63] B3LYP,[64] B97D,[65] B973 C,[66] M06[67] and M06-2X.[67] The geometry optimizations were performed with each of these functionals, using the def2-TZVPP, def2-TZVP, and def2-SVP basis sets and the m4 grid.[44,68] The interaction energies were computed as the difference in the energy of the optimized supersystem and the energies of the optimized AA and CO_2 geometries.

For assessing the accuracy of the selected density functionals, we have computed the mean absolute error (MAE):

$$\text{Mean absolute error (MAE)} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \quad (3)$$

where y_i is the CCSD(F12)(T)/CBS reference energies, x_i is the DFT energy for each AA per density functional, and n is the number of AAs ($n = 20$). The root mean square error (RMSE) is defined as:

$$\text{Root mean square error (RMSE)} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - x_i)^2}{n}} \quad (4)$$

where y_i is the CCSD(F12)(T)/CBS reference energies, x_i is the DFT energy for each AA per density functional, and n is the number of AAs ($n = 20$). We are also reporting the maximum error (MAX) per functional.

Results and Discussion

Correlation Effects: Serine–CO_2 as a Test Case

Serine has been selected in this study as a representative example of the different electron correlation term contributions to the CO_2 interaction energies. In this analysis, we will abbreviate the two basis sets selected for these calculations (cc-pVDZ-F12 and cc-pVTZ-F12) as DZ and TZ, respectively. A short comparison with the aug-cc-pVDZ and aug-cc-pVXZ basis sets (abbreviated as aVDZ and aVTZ, respectively) is given at the end of this section since it is still unclear which of these two families of basis sets (cc-pVXZ-F12 and aug-cc-pVXZ) provide higher accuracy for noncovalent interactions.[69–71]

The slow convergence to the complete basis sets limit, an effect that originates from the poor description of the electron cusp when a truncated basis is used in post-HF calculations,[72] is addressed by the introduction of explicitly-correlated terms (F12). Since it is known that coupled-cluster methods require large basis sets to provide highly accurate results, we wanted to analyze the individual electron correlation terms on the interaction energy of a representative molecular system before we consider the full AA database. For that purpose, we have selected serine–CO_2, one of the simplest AA, and we have performed a detailed analysis of the individual electron correlation terms (Table 1). For this analysis, the most stable geometry of the serine–CO_2 supersystem was used, which involves a CO_2 weakly bound on the COOH site of the AA.

The HF/TZ interaction energy is −2.10 kcal/mol, while the CABS singles correction to HF with the same basis is positive (0.02 kcal/mol). The CCSD correction to the HF interaction energy with the double- and triple-zeta basis sets is −2.22 and −2.34 kcal/mol, respectively, while the CCSD(F12)/TZ converges to −2.36 kcal/mol. Note that the CCSD(F12) energy term computed with the smaller basis set (DZ, −2.34 kcal/mol) is identical to the conventional CCSD with the larger triple-zeta basis. Thus, the gain from the explicit correlation becomes evident since the CCSD(F12)/DZ calculation requires less computational effort than a conventional CCSD calculation with a triple-zeta basis set. The perturbative triples (T) correction term contributed about 0.5 kcal/mol to the CCSD method, which corresponds to ~19% of the total correlation energy of the CO_2–serine interaction. Note that the (T) energy term computed from the CCSD(F12)(T) level of theory (shown as $(T)_{F12}$ in Table 1) differs from the conventional CCSD(T) (shown as $(T)_{con}$ in Table 1) since the computed t_i^a and t_{ij}^{ab} amplitudes have different values when the explicit correlation is included in the coupled-cluster projected equations. However, both extrapolated $(T)_{con}$/DT and $(T)_{F12}$/DT terms from Eq. 1 are identical (−0.56 kcal/mol), which means that both approaches converge to the same CBS limit. For that reason, we have used the $(T)_{F12}$/DT term for the computation of the reference values of the 20 AA–CO_2 supersystems (see below). Addition of the $(T)_{con}$/DT to the $\delta E_{CCSD(F12)/TZ}$ (−2.36 kcal/mol) provides the best estimate for

Table 1. Individual energy contributions to the interaction energy of CO_2 with serine. δE represents the interaction energy contribution of the different correlation terms (corrections to HF) and ΔE_{int} represents the total interaction energy (HF + corrections).

METHODS	δE [kcal/mol] cc-pVDZ-F12	cc-pVTZ-F12	ΔE_{int} [kcal/mol] cc-pVDZ-F12	cc-pVTZ-F12
HF	0.00	0.00	−2.06	−2.10
CABS S	0.00	0.02	−2.06	−2.08
CCSD	−2.22	−2.34	−4.29	−4.43
CCSD(F12)	−2.34	−2.36	−4.40	−4.44
$(T)_{con}$	−0.49	−0.54		
CCSD(T)	−2.72	−2.87	−4.78	−4.97
$(T)_{F12}$	−0.48	−0.53		
CCSD(F12)(T)	−2.82	−2.89	−4.88	−4.97
$(T)_{con}$/DT		−0.56		
$(T)_{F12}$/DT		−0.56		
CCSD(F12) + (T)/DT	−2.87	−2.91	−4.95	**−4.99**
METHODS	δE [kcal/mol] aug-cc-pVDZ	aug-cc-pVTZ	ΔE_{int} [kcal/mol] aug-cc-pVDZ	aug-cc-pVTZ
HF	0.00	0.00	−2.49	−2.24
CABS S	0.35	0.15	−2.14	−2.09
CCSD(F12)	−2.38	−2.43	−4.52	−4.52
$(T)_{F12}$/DT		−0.54		
CCSD(F12) + (T)/DT	−2.92	−2.96	−5.06	**−5.05**

the total correlation energy to the CBS (−2.91 kcal/mol). Interestingly, extrapolation of the total correlation energies computed at CCSD(F12)(T)/DZ and CCSD(F12)(T)/TZ levels by applying Eq. 1 provides a $\delta E_{CCSD(F12)(T)/CBS} = -2.92$ kcal/mol, which is almost identical (difference of 0.01 kcal/mol) with the energy term computed from the separate extrapolation of the CCSD via explicit correlation and (T) via the two-point Helgaker formula. Addition of the HF energy and the first-order correction from CABS Singles provides the best estimate (−4.99 kcal/mol) for the interaction energy between serine and CO_2.

The second half of Table 1 contains the contributions to the serine–CO_2 interaction energy from the aug-cc-pV*X*Z basis sets (*X* = D, T), abbreviated as a*VXZ*. Surprisingly, the conventional HF energies from aVDZ and aVTZ (−2.49 and −2.24 kcal/mol, respectively) significantly deviate from the equivalent values from the cc-pV*X*Z-F12 basis sets (−2.06 and −2.10 kcal/mol). On the contrary, both families of basis sets converge to the same HF limit upon addition of the first-order correction from CABS Singles (−2.08 and −2.09 kcal/mol for TZ and aVTZ, respectively). The $\delta E_{CCSD(F12)}$ correlation energy terms are in reasonable agreement (−2.36 and −2.43 kcal/mol for TZ and aVTZ, respectively), while the $(T)_{F12}$/DT extrapolated energies converge to almost the same value (−0.56 and −0.54 kcal/mol for DZ/TZ and aVDZ/aVTZ, respectively). Overall, the best estimates for the total interaction energies of serine–CO_2 system are in fair agreement (−4.99 and −5.05 kcal/mol).

In the next section, the sum of $E_{CCSD(F12)/TZ} + E_{(T)/DT}$ is used for the generation of a balanced set of accurate reference energies for the interaction of CO_2 with the 20 naturally occurring AAs. For sake of simplicity, we will refer to these energies as CCSD(F12)(T)/CBS.

An isolated AA has three potential sites that can form weak interactions with CO_2; the alpha amine (α-NH_2), the carboxylic group (COOH), and the side group *R* that is unique for each AA. For the purpose of this study, and for providing a complete dataset of highly accurate reference values, we have computed at the CCSD(F12)/TZ + (T)/DT level the CO_2 interaction energies for all 20 AAs and for all three different sites (α-NH_2, COOH, side group *R*). All reference data for the interaction energies and optimal atom distances between CO_2 and the AA atom closest to the carbon atom of CO_2 are included in Table 2, while all energies are shown graphically as a plot in Figure 1. Figure 2 shows the optimized geometries of the strongest interaction site for all the AA–CO_2 supersystems. For AAs with non-polar side groups, DFT geometry optimizations were not trapped on local minima, but they converged to molecular geometries where CO_2 resides closer to more polar sites. For these cases, the CO_2-side group interaction is indicated on Table 2 with a dash (–).

For almost all AAs, the interactions of CO_2 with the carboxylic acid (about −4.0 to −5.0 kcal/mol) are stronger than the interaction with the α-NH_2 group (about −2.5 to −3.5 kcal/mol). There is an exception to this with arginine, where the CO_2 interaction with the α-NH_2 group is further stabilized by interactions between O(CO_2) and H(side chain –NH_2), and C(CO_2) and N(side chain NH_2). Although CO_2 primarily interacts with the α-NH_2 group, this cooperative effect further enhances the CO_2 affinity of arginine.

Leucine (Leu–CO_2) exhibits the strongest interaction energy (−6.17 kcal/mol) in the entire series (Table 2), where CO_2 preferably interacts with the COOH group. This can be attributed to the dipole-induced dipole interactions between the C(CO_2) and the O(COOH) as well as additional stabilizing hydrogen bonding interactions that are absent in other AAs.[73,74] These arise from the interactions of H atoms located on the side chain methyl groups and the oxygen atoms of CO_2. This interaction is also present in isoleucine-structural isomer of leucine, but in a form of head-on interaction compared to the parallel position of CO_2 next to leucine. Arginine (Arg–CO_2) has the second strongest interaction energy (−6.12 kcal/mol) with CO_2 interacting with the side chain that has three amine groups. This rich nucleophilic center results in the observed interaction

Table 2. CCSD(F12)(T)/CBS reference interaction energies (E_{Int}, in kcal/mol) and interatomic distances (in Å) of the 20 naturally occurring AAs and CO_2 (in alphabetical order). Bold font indicates the most preferable interaction site. A dash (–) indicates that no favorable interaction was found between CO_2 and the side groups (e.g., the methyl group of alanine).

AA	α-NH_2		COOH		Side Group	
	E_{Int}	$R_{N...C}$	E_{Int}	$R_{O...C}$	E_{Int}	$R_{X...C}$
Ala	−2.55	3.058	**−3.89**	2.834	–	–
Arg	−5.24	3.553	−4.46	2.838	**−6.12**	2.804
Asn	−2.45	3.054	−3.37	2.951	**−6.08**	2.791
Asp	−2.36	3.065	−4.05	3.492	**−5.53**	2.881
Cys	−2.59	3.024	**−3.93**	2.866	−3.62	3.540
Glu	−2.58	3.018	**−5.19**	2.911	−3.89	2.809
Gln	−2.67	3.003	−4.46	2.884	**−4.85**	2.796
Gly	−2.36	3.040	**−3.84**	2.830	–	–
His	−2.58	3.045	−4.55	2.784	**−4.75**	2.836
Ile	−2.42	3.079	**−4.20**	2.845	–	–
Leu	−2.89	3.112	**−6.17**	2.859	–	–
Lys	−2.43	3.060	−4.05	3.307	**−4.14**	3.025
Met	−2.44	3.046	**−4.16**	2.815	−2.78	3.296
Phe	−3.08	3.120	**−4.60**	2.835	–	–
Pro	−4.19	2.878	**−4.43**	3.096	–	–
Ser	−2.65	3.014	**−4.99**	2.810	−4.00	2.805
Thr	−2.61	3.020	**−4.61**	2.845	−4.37	2.785
Trp	−3.39	3.120	−3.92	2.993	**−4.26**	3.175
Tyr	−2.54	3.044	**−5.28**	2.835	−4.66	3.751
Val	−2.43	3.066	**−4.15**	2.847	–	–

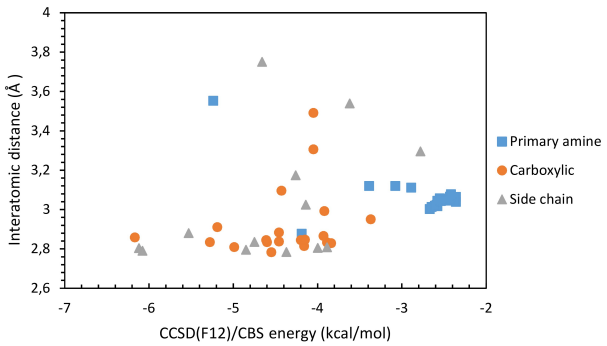


Figure 1. CCSD(F12)(T)/CBS interaction energies and the respective interatomic distances of the 20 naturally occurring AAs and CO_2 at different interaction points

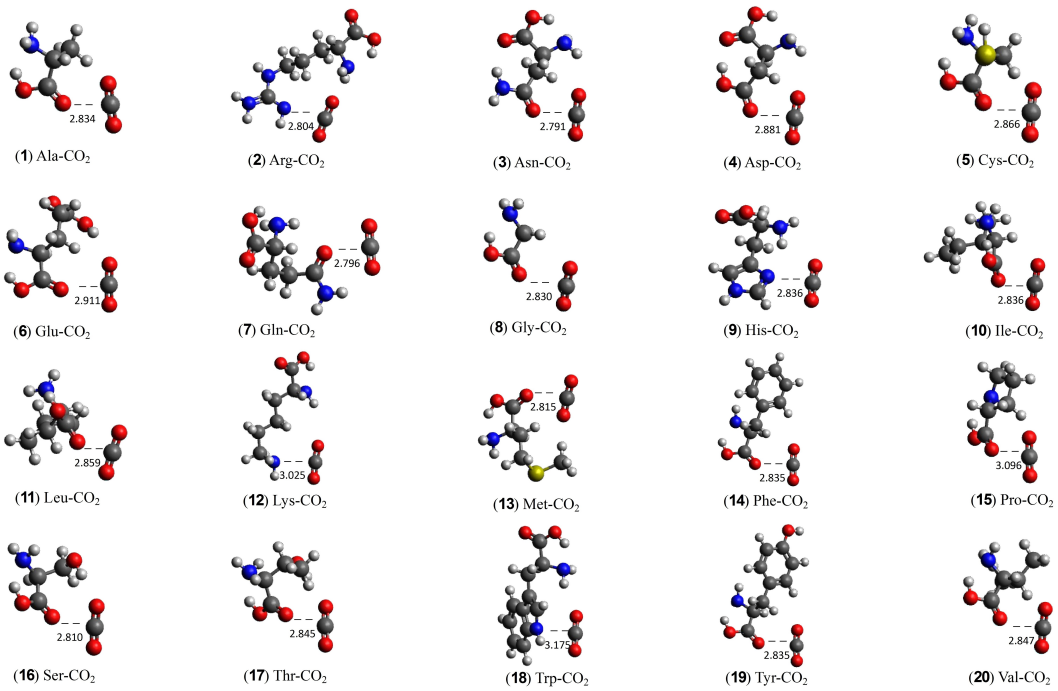


Figure 2. Optimized geometries of the 20 naturally occurring AAs and CO_2 (in alphabetical order). The most preferable interaction site with CO_2 is shown.

energy with CO_2. In addition, there is a stabilizing hydrogen bonding between the O(CO_2) and the H(primary NH_2). Asparagine (Asn–CO_2) has the third strongest interaction energy (−6.08 kcal/mol) where CO_2 is in close proximity to the side chain amide group. This dipole-induced dipole interaction is stabilized by hydrogen bonding. Asparagine has also one of the shortest interatomic distances with CO_2 with a value of 2.791 Å. For all other AAs, the reported interactions energies are between −2.36 kcal/mol to −5.53 kcal/mol (Table 2). Most of the AAs interacted preferentially with CO_2 through the primary carboxyl group, with an average interaction energy of −4.43 kcal/mol and an average interatomic distance of 2.92 Å. On the other hand, interactions at the α-amine groups displayed the weakest energies (average $E_{INT} = -2.73$ kcal/mol) and longest interatomic distances (average $R_{N\ldots C(CO2)} = 3.09$ Å). in a few cases, we found an intramolecular interaction between the H(COOH) and N(α-NH_2) on the AA backbone, which competes with the CO_2–N(α-NH_2) and eventually reduces the strength of the CO_2 interaction.

Discussion on Leucine, Isoleucine and Valine Interacting with CO_2

In this section, the differences and similarities of the interaction energies between CO_2 and leucine, isoleucine and valine are discussed. These three AAs have similar, non-polar side groups, but their computed interaction energies with CO_2 show significant deviations (Figure 3(a)). While leucine has one of the strongest CO_2 interaction energies (−6.17 kcal/mol at the CCSD(F12)(T)/CBS level), isoleucine and valine found to have a significantly weaker interaction with CO_2 (−4.20 and −4.15 kcal/mol at the CCSD(F12)(T)/CBS level, respectively).

All three systems are stabilized with dipole-induced dipole interactions between C(CO_2) and O(COOH) (~2.86 Å for all three cases), while weak "hydrogen bonding interactions" between H(methyl) and O(CO_2) provide additional stability. For all three cases, the closest O(CO_2)…H(methyl) distance is ~2.67 Å. However, the orientation and angular placement of CO_2 in the leucine supersystem differ from the isoleucine and valine cases. In particular, the absence of a methyl group in beta position from the COOH group in leucine creates less steric repulsion and allows CO_2 to be in close proximity to the AA. This is shown in Figure 3(b), and it is also validated by the measured distance between the second O(CO_2) that its distance from the AA is ~3.0 Å. On the contrary, the presence of the beta methyl group in isoleucine and valine (shown in dashed red circle in Figure 3(b)) introduces steric repulsion to CO_2 (Figure 3(d)), and the distance of the second O(CO_2) exceeds the 3.5 Å from the AA (both isoleucine and valine).

Screening of Different Density Functionals

The stable conformers from the classical mechanics simulations were further optimized with different density functionals (PBE, BP86, BLYP, TPSS, PW6B95, BHLYP, PBE0, TPSSH, B3LYP, B97D, B973C, M06, and M06-2X) and basis sets (def2-SVP, def2-TZVP, and def2-TZVPP). The performance of each of these density functionals was evaluated with respect to the CCSD(F12)(T)/CBS reference energies. For this analysis, we considered only the supersystem geometries with the strongest interaction energies

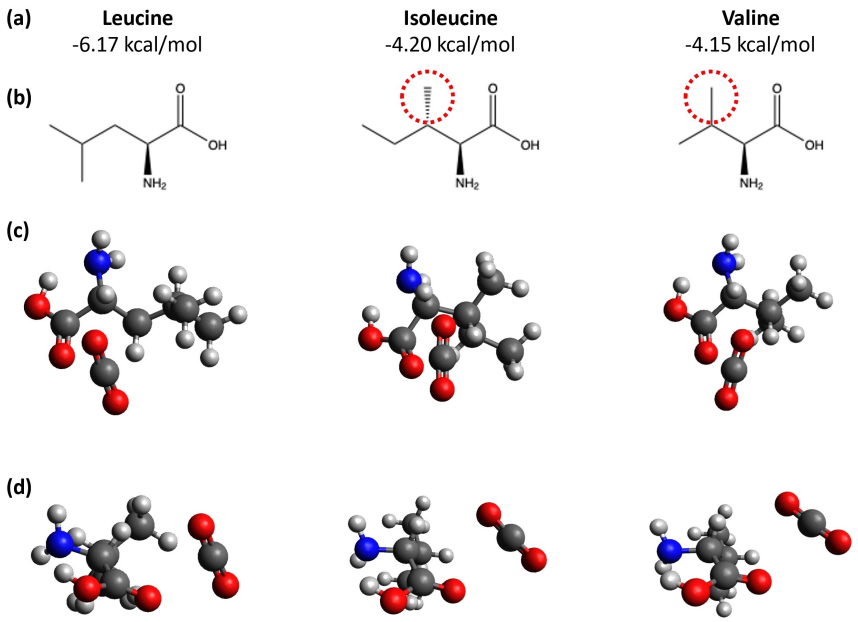


Figure 3. (a) Interaction energies of leucine, isoleucine, and valine with CO_2 calculated at the CCSD(F12)(T)/CBS limit. (b) Molecular structures of the three AAs. (c) Optimized geometries at the PBE0-D3(BJ)/def2-TZVPP level (top view) and (d) rotated by 90° (side view).

of Table 2. This analysis would identify which density functionals quntitatively describe to a considerable degree the noncovalent interactions of CO_2 and AAs. The MAE, RMSE, and MAX error values of the different density functionals with def2-TZVPP basis sets are shown in Table 3. Detailed tables containing results with all the density functionals and basis sets combinations used in this study are included in the Supporting Information, together with the corresponding statistical analysis of their errors (Sections S1–S3, Tables S1–S7).

The MAE for the interaction energy (kcal/mol) and interatomic distances (Å) of the different functionals and def2-TZVPP and def2-TZVP basis sets are presented in Figure 4. Of all the functionals used in this study, the GGA density functionals B97-3c-D3(BJ) and BP86-D3(BJ) provided the largest errors, with MAE of 0.92 kcal/mol and 0.87 kcal/mol, respectively (def2-TZVPP basis set). The hybrid density functionals PBE0-D3(BJ) and B3LYP-D3(BJ) provided the highest accuracy and lowest MAE (0.13 kcal/mol and 0.17 kcal/mol, respectively, with the def2-TZVPP basis set). A similar behavior was observed for the def2-TZVP basis sets, while results with the smaller def2-SVP basis sets and for all density functionals significantly deviated from the reference values (Supporting Information, Figure S1). For example, the density functionals BHLYP-D3(BJ) had the highest MAE of 3.06 kcal/mol. The RMSE value computed with BP86-D3(BJ)/def2-TZVPP was the highest of all methods used in this study (1.04 kcal/mol, see Supporting Information, Figure S2), while PBE0-D3(BJ)/def2-TZVPP had the lowest RMSE (0.16 kcal/mol). It is noteworthy that B3LYP-D3/BJ had also a low RMSE of 0.21 kcal/mol. Overall, all density functionals together with the triple-zeta quality basis sets were in good agreement with respect to the reference values, with RMSEs not exceeding the 1.04 kcal/mol.

Table 3. MAE, RMSE, and MAX Error of the interaction energy (in kcal/mol) of the different density functionals and def2-TZVPP basis sets. Results are listed in increasing MAE order.

Density Functional	MAE	RMSE	MAX
PBE0-D3(BJ)	0.13	0.16	0.36
B3LYP-D3(BJ)	0.17	0.21	0.41
PW6B95-D3(BJ)	0.19	0.26	0.73
M06-2X	0.23	0.29	0.61
BLYP-D3(BJ)	0.24	0.27	0.51
PBE-D3(BJ)	0.31	0.53	2.07
TPSSh-D3(BJ)	0.59	0.77	2.57
B97d-D3(BJ)	0.60	0.65	1.08
BHLYP-D3(BJ)	0.62	0.64	1.01
TPSS-D3(BJ)	0.70	0.87	2.79
M06	0.81	0.85	1.41
BP86-D3(BJ)	0.87	1.04	3.10
B97-3c-D3(BJ)	0.92	0.95	1.43

The source of the large deviations from the smaller def2-SVP basis set is due to the basis set incompleteness which leads to the basis set superposition error (BSSE). In order to further evaluate this effect, we applied the counterpoise (CP) correction proposed by Boys and Bernardi[75] in the serine–CO_2 system and for the PBE0-D3(BJ) density functional. The uncorrected PBE0-D3(BJ)/def2-SVP interaction energy is −7.66 kcal/mol, which differs by more than 2.5 kcal/mol from the CCSD(F12)(T) reference (−4.99 kcal/mol). On the contrary, the CP-corrected PBE0-D3(BJ)/def2-SVP interaction energy is −5.14 kcal/mol, which is in better agreement with the reference, as well as with results obtained from the larger triple-zeta quality basis sets (−5.05 and −5.04 kcal/mol for def2-TZVP and def2-TZVPP, respectively).

For the interatomic distances, the GGA density functionals B97-3c-D3(BJ) and B97D-D3(BJ) provided the largest errors with

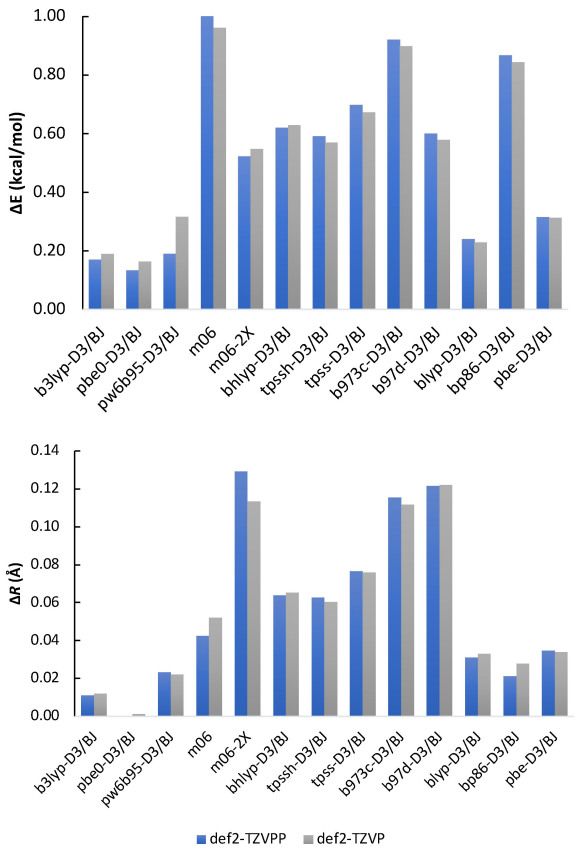


Figure 4. Top: CO_2 interaction energy mean absolute error (MAE) of different density functionals with respect to the CCSD(F12)(T)/CBS reference (in kcal/mol). Bottom: Interatomic distance MAE of different functionals with respect to the PBE0-D3(BJ)/def2-TZVPP geometries (in Å).

a MAE of 0.12 Å with def2-TZVPP basis sets. On the contrary, B3LYP-D3(BJ) was in excellent agreement with the PBE0-D3(BJ)/def2-TZVPP geometry that was used as reference in this study (MAE of 0.01 Å, def2-TZVPP). The def2-TZVP data followed a similar behavior, with density functionals PBE0-D3(BJ) and B3LYP-D3(BJ) providing the lowest MAE. As expected, large deviations from the reference interactions energies were observed from calculations with the smaller def2-SVP basis set.

Conclusions and Outlook

Inspired by recent work on CO_2 capture and separation through bio-inspired materials, we have performed a quantum chemical study on the noncovalent interactions between the 20 natural AAs and CO_2. The AAs contain various nucleophilic groups, like amine, carboxyl, hydroxyl, and thiol groups, that can interact favorably with CO_2, exceeding the −6.0 kcal/mol (e.g. arginine–CO_2 supersystem). We started by performing explicitly-correlated CCSD calculations together with perturbative triples at the CBS limit for the computation of accurate interaction energies of the systems of interest. The CCSD(F12)(T)/CBS energies for every possible interaction site of the 20 AAs with CO_2 served as reference data for testing different density functional and basis sets. We concluded that the hybrid density functionals PBE0-D3(BJ) and B3LYP-D3(BJ) provide the highest accuracy, with MAEs of 0.13 kcal/mol and 0.18 kcal/mol, respectively (def2-TZVPP basis set). Our results showed that polar functional groups enhance the CO_2 interaction strength, as expected. A remarkably strong interaction was found between the carboxylic acid of leucine and CO_2, which is attributed to less steric repulsion from the non-polar side group. Comparison between leucine, isoleucine and valine further verified our computational outcome, while it revealed an interplay between multiple attraction sites that enhance CO_2 interactions. As we shift from single AAs to larger oligopeptides that can be incorporated on surfaces or inside porous materials, such cooperative interactions become more complex. In the future, we are planning to utilize conclusions extracted from this work in hybrid quantum chemical/machine learning studies for the elucidation of noncovalent interactions between oligopeptides and CO_2. The presence of a large variety of nucleophilic side chains on the oligopeptides would be a key feature in enhancing the optimal physisorption of CO_2.

Acknowledgements

This material is based on work supported by the National Science Foundation under Grant CHE-2143354 (NSF CAREER Award). The authors acknowledge the Infrastructure for Scientific Applications and Advanced Computing (ISAAC) of the University of Tennessee for computational resources. We would like to thank Rebecca Ryan for fruitful discussions during the early stages of this project.

Conflict of Interests

The authors declare no conflict of interest.

Data Availability Statement

Cartesian coordinates of all PBE0-D3(BJ)/def2-TZVPP geometries, CO_2 interaction energies from different density functionals and for different nucleophilic centers, together with statistical analysis of interaction energies and interatomic distances relative to CCSD(F12)(T)/CBS reference are available as supplementary material of this article.

Keywords: quantum chemistry · noncovalent interactions · coupled-cluster · amino acids · CO_2 capture

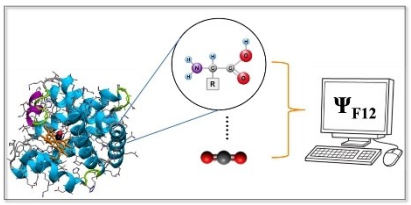
[1] D. C. R. H.-O. Pörtner, M. Tignor, E. S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.) *IPCC, 2022: Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; 2022.
[2] M. E. Boot-Handford, J. C. Abanades, E. J. Anthony, M. J. Blunt, S. Brandani, N. Mac Dowell, J. R. Fernández, M.-C. Ferrari, R. Gross, J. P. Hallett, *Energy Environ. Sci.* **2014**, *7*, 130–189.

[3] E. E. Ünveren, B. Ö. Monkul, Ş. Sarıoğlan, N. Karademir, E. Alper, *Petroleum* **2017**, *3*, 37–50.
[4] E. S. Kentish, A. C. Scholes, W. G. Stevens, *Recent Pat. Chem. Eng.* **2008**, *1*, 52–66.
[5] A. L. Chaffee, G. P. Knowles, Z. Liang, J. Zhang, P. Xiao, P. A. Webley, *Int. J. Greenhouse Gas Control* **2007**, *1*, 11–18.
[6] M. Cao, L. Zhao, D. Xu, R. Ciora, P. K. Liu, V. I. Manousiouthakis, T. T. Tsotsis, *J. Membr. Sci.* **2020**, *605*, 118028.
[7] J. D. Carruthers, M. A. Petruska, E. A. Sturm, S. M. Wilson, *Microporous Mesoporous Mater.* **2012**, *154*, 62–67.
[8] S. Burt, A. Baxter, L. Baxter, **2009**, https://scholar.google.com/citations?view op=view citation&hl=en&user=pBJFGdMAAAAJ&cstart=100&pagesize=100&sortby=pubdate&citation for view=pBJFGdMAAAAJ:YFjsv pBGBYC, https://www.researchgate.net/profile/Larry-Baxter-4/publication/264875049 Cryogenic CO2 Capture as a Cost-Effective CO2 Capture Process/links/547c7dfd0cf2cfe203bfb637/Cryogenic-CO2-Capture-as-a-Cost-Effective-CO2-Capture-Process.pdf.
[9] E. Torralba-Calleja, J. Skinner, D. Gutiérrez-Tauste, *J. Chem.* **2013**, *2013*, 473584.
[10] R. Custelcean, K. A. Garrabrant, P. Agullo, N. J. Williams, *Cell Rep. Phys. Sci.* **2021**, *2*, 100385.
[11] D. Stamberga, N. A. Thiele, R. Custelcean, *MRS Adv.* **2022**, *7*, 399–403.
[12] R. Custelcean, N. J. Williams, K. A. Garrabrant, P. Agullo, F. M. Brethomé, H. J. Martin, M. K. Kidder, *Ind. Eng. Chem. Res.* **2019**, *58*, 23338–23346.
[13] R. Ramezani, S. Mazinani, R. Di Felice, *Rev. Chem. Eng.* **2022**, *38*, 273–299.
[14] C. B. Bavoh, B. Lal, H. Osei, K. M. Sabil, H. Mukhtar, *J. Nat. Gas Sci. Eng.* **2019**, *64*, 52–71.
[15] V. Sang Sefidi, P. Luis, *Ind. Eng. Chem. Res.* **2019**, *58*, 20181–20194.
[16] S. Chatterjee, S. Rayalu, S. D. Kolev, R. J. Krupadam, *J. Environ. Chem. Eng.* **2016**, *4*, 3170–3176.
[17] I. Eide-Haugmo, O. G. Brakstad, K. A. Hoff, K. R. Sørheim, E. F. da Silva, H. F. Svendsen, *Energy Procedia* **2009**, *1*, 1297–1304.
[18] Z. Zhang, Y. Li, W. Zhang, J. Wang, M. R. Soltanian, A. G. Olabi, *Renewable Sustainable Energy Rev.* **2018**, *98*, 179–188.
[19] G. Hu, K. H. Smith, Y. Wu, K. A. Mumford, S. E. Kentish, G. W. Stevens, *Chin. J. Chem. Eng.* **2018**, *26*, 2229–2237.
[20] T. Hirasawa, H. Shimizu, *Curr. Opin. Biotechnol.* **2016**, *42*, 133–146.
[21] M. J. Allison, *J. Anim. Sci.* **1969**, *29*, 797–807.
[22] K. S. Lackner, *Energy* **2013**, *50*, 38–46.
[23] A. Kumar, D. G. Madden, M. Lusi, K. J. Chen, E. A. Daniels, T. Curtin, J. J. Perry IV, M. J. Zaworotko, *Angew. Chem. Int. Ed.* **2015**, *54*, 14372–14377; *Angew. Chem.* **2015**, *127*, 14580–14585.
[24] R. Vaidhyanathan, D. Bradshaw, J. N. Rebilly, J. P. Barrio, J. A. Gould, N. G. Berry, M. J. Rosseinsky, *Angew. Chem. Int. Ed.* **2006**, *45*, 6495–6499; *Angew. Chem.* **2006**, *118*, 6645–6649.
[25] Y. Xie, Z. Yu, X. Huang, Z. Wang, L. Niu, M. Teng, J. Li, *Chem. Eur. J.* **2007**, *13*, 9399–9405.
[26] C. H. Görbitz, *Chem. Eur. J.* **2007**, *13*, 1022–1031.
[27] J. Perez Barrio, J. N. Rebilly, B. Carter, D. Bradshaw, J. Bacsa, A. Y. Ganin, H. Park, A. Trewin, R. Vaidhyanathan, A. I. Cooper, *Chem. Eur. J.* **2008**, *14*, 4521–4532.
[28] D. V. Soldatov, I. L. Moudrakovski, J. A. Ripmeester, *Angew. Chem. Int. Ed.* **2004**, *116*, 6468–6471.
[29] E. N. Koukaras, A. D. Zdetsis, G. E. Froudakis, *J. Phys. Chem. Lett.* **2011**, *2*, 272–275.
[30] J. Rabone, Y.-F. Yue, S. Chong, K. Stylianou, J. Bacsa, D. Bradshaw, G. Darling, N. Berry, Y. Khimyak, A. Ganin, *Science* **2010**, *329*, 1053–1057.
[31] R. Vaidhyanathan, S. S. Iremonger, G. K. Shimizu, P. G. Boyd, S. Alavi, T. K. Woo, *Science* **2010**, *330*, 650–653.
[32] X. Wang, N. G. Akhmedov, Y. Duan, D. Luebke, D. Hopkinson, B. Li, *ACS Appl. Mater. Interfaces* **2013**, *5*, 8670–8677.
[33] M. A. Hussain, Y. Soujanya, G. N. Sastry, *Environ. Sci. Technol.* **2011**, *45*, 8582–8588.
[34] F. F. Chen, K. Huang, Y. Zhou, Z. Q. Tian, X. Zhu, D. J. Tao, D. e. Jiang, S. Dai, *Angew. Chem. Int. Ed.* **2016**, *128*, 7282–7286.
[35] A. R. Shaikh, H. Karkhanechi, E. Kamio, T. Yoshioka, H. Matsuyama, *J. Phys. Chem. C* **2016**, *120*, 27734–27745.
[36] P. M. Gill, M. Head-Gordon, J. A. Pople, *Int. J. Quantum Chem.* **1989**, *36*, 269–280.
[37] T. Helgaker, W. Klopper, D. P. Tew, *Mol. Phys.* **2008**, *106*, 2107–2143.
[38] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
[39] L. S. Dodda, I. Cabeza de Vaca, J. Tirado-Rives, W. L. Jorgensen, *Nucleic Acids Res.* **2017**, *45*, W331–W336.
[40] S. Plimpton, *J. Comput. Phys.* **1995**, *117*, 1–19.
[41] J. P. Perdew, M. Ernzerhof, K. Burke, *J. Chem. Phys.* **1996**, *105*, 9982–9985.
[42] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456–1465.
[43] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs, *Chem. Phys. Lett.* **1995**, *240*, 283–290.
[44] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
[45] R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, *Chem. Phys. Lett.* **1989**, *162*, 165–169.
[46] C. R. Maroon, J. Townsend, K. R. Gmernicki, D. J. Harrigan, B. J. Sundell, J. A. Lawrence, S. M. Mahurin, K. D. Vogiatzis, B. K. Long, *Macromolecules* **2019**, *52*, 1589–1600.
[47] J. Townsend, C. P. Micucci, J. H. Hymel, V. Maroulas, K. D. Vogiatzis, *Nat. Commun.* **2020**, *11*, 3230.
[48] W. Kutzelnigg, *Theor. Chim. Acta* **1985**, *68*, 445–469.
[49] W. Klopper, W. Kutzelnigg, *Chem. Phys. Lett.* **1987**, *134*, 17–22.
[50] W. Kutzelnigg, W. Klopper, *J. Chem. Phys.* **1991**, *94*, 1985–2001.
[51] D. P. Tew, W. Klopper, C. Neiss, C. Hättig, *Phys. Chem. Chem. Phys.* **2007**, *9*, 1921–1930.
[52] C. Hättig, D. P. Tew, A. Köhn, *J. Chem. Phys.* **2010**, *132*, 231102.
[53] W. Klopper, C. C. Samson, *J. Chem. Phys.* **2002**, *116*, 6397–6410.
[54] R. A. Kendall, T. H. Dunning, R. J. Harrison, *J. Chem. Phys.* **1992**, *96*, 6796–6806.
[55] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, A. K. Wilson, *Chem. Phys. Lett.* **1998**, *286*, 243–252.
[56] A. Köhn, D. P. Tew, *J. Chem. Phys.* **2010**, *132*, 024101.
[57] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865.
[58] J. P. Perdew, *Phys. Rev. B* **1986**, *33*, 8822.
[59] A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098.
[60] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.
[61] J. Tao, J. P. Perdew, V. N. Staroverov, G. E. Scuseria, *Phys. Rev. Lett.* **2003**, *91*, 146401.
[62] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 1372–1377.
[63] V. N. Staroverov, G. E. Scuseria, J. Tao, J. P. Perdew, *J. Chem. Phys.* **2003**, *119*, 12129–12137.
[64] A. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.
[65] S. Grimme, *J. Comput. Chem.* **2006**, *27*, 1787–1799.
[66] J. G. Brandenburg, C. Bannwarth, A. Hansen, S. Grimme, *J. Chem. Phys.* **2018**, *148*, 064104.
[67] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215–241.
[68] A. Hellweg, C. Hättig, S. Höfener, W. Klopper, *Theor. Chem. Acc.* **2007**, *117*, 587–597.
[69] H. Kruse, R. Szabla, J. Sponer, *J. Chem. Phys.* **2020**, *152*, 214104.
[70] D. A. Sirianni, L. A. Burns, C. D. Sherrill, *J. Chem. Theory Comput.* **2017**, *13*, 86–99.
[71] N. N. Dutta, K. Patkowski, *J. Chem. Theory Comput.* **2018**, *14*, 3053–3070.
[72] J. Townsend, J. K. Kirkland, K. D. Vogiatzis, Post-Hartree-Fock methods: configuration interaction, many-body perturbation theory, coupled-cluster theory. In *Mathematical Physics in Theoretical Chemistry*, Blinder, S. M.; House, J. E., Eds. Elsevier: 2018.
[73] K. D. Vogiatzis, A. Mavrandonakis, W. Klopper, G. E. Froudakis, *ChemPhysChem* **2009**, *10*, 374–383.
[74] K. D. Vogiatzis, W. Klopper, J. Friedrich, *J. Chem. Theory Comput.* **2015**, *11*, 1574–1584.
[75] S. F. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553–566.

Manuscript received: January 11, 2023
Revised manuscript received: April 6, 2023
Accepted manuscript online: April 8, 2023
Version of record online: ■■, ■■

RESEARCH ARTICLE

CO_2-philicity inspired by Nature: Highly accurate gas-phase quantum chemical calculations on the noncovalent interactions between amino acids and CO_2. Examination of multiple sites that enhance CO_2 binding.



*A. G. Sylvanus, Prof. Dr. K. D. Vogiatzis**

1 – 9

Accurate Interaction Energies of CO_2 with the 20 Naturally Occurring Amino Acids