Electronic structure benchmark calculations of
inorganic and biochemical carboxylation reactions

Oscar A. Douglas-Gallardo, David Adrián Sáez, Stefan Vogt-Geisse, and
Esteban Vöhringer-Martinez*

Departamento de Físico-Química, Facultad de Ciencias Químicas, Universidad de
Concepción, Chile.

E-mail: evohringer@udec.cl
Phone: +56 41 220-4986
Abstract

Carboxylation reactions represent a very special class of chemical reactions that is characterized by the presence of a carbon dioxide (CO$_2$) molecule as reactive species within its global chemical equation. These reactions work as fundamental gear to accomplish CO$_2$ fixation and thus to build up more complex molecules through different technological and biochemical processes. In this context, a correct description of the CO$_2$ electronic structure turns out to be crucial to study the chemical and electronic properties associated with this kind of reactions. Here, a systematic study of CO$_2$ electronic structure and its contribution to different carboxylation reaction electronic energies has been carried out by means of several high-level \textit{ab-initio} post-Hartree Fock (post-HF) and Density Functional Theory (DFT) calculations for a set of biochemistry and inorganic systems. We have found that for a correct description of the CO$_2$ electronic correlation energy it is necessary to include post-CCSD(T) contributions (beyond the gold standard). These high-order excitations are required to properly describe the interactions of the four $\pi$-electrons associated with the two degenerated $\pi$-molecular orbitals of the CO$_2$ molecule. Likewise, our results show that in some reactions it is possible to obtain accurate reaction electronic energy values with computationally less demanding methods when the error in the electronic correlation energy compensates between reactants and products. Furthermore, the provided post-HF reference values allowed to validate different DFT exchange-correlation functionals combined with different basis sets for chemical reactions that are relevant in biochemical CO$_2$ fixing enzymes.

Introduction

Carboxylation reactions are a very important class of chemical reactions that are present in a plethora of different biological and inorganic process.$^{1-4}$ In this kind of chemical reactions a carbon dioxide (CO$_2$) molecule is involved as reactive species within the global chemical equation considered. CO$_2$ is a low molecular-weight compound that is found widely
distributed in the Earth’s atmosphere.\textsuperscript{2,4} This neutral molecule exhibits only a quadrupole moment and a very low reactivity. This chemical behavior can be quantified by its high standard Gibbs free energy of formation (-394.4 kJ mol\textsuperscript{-1}) which can partially be associated with its two strong (C=O) double-bonds.\textsuperscript{5} The central carbon atom within the CO\textsubscript{2} molecule is found in its highest oxidation state and can only be reduced by a suitable chemical species with enough reactivity.\textsuperscript{1,3}

For many centuries the atmospheric CO\textsubscript{2} concentration have been regulated by diverse natural cycles.\textsuperscript{2,4} However, in the last decades an increase in the atmospheric CO\textsubscript{2} level has been recorded disturbing the delicate environmental balance.\textsuperscript{2,4} Nowadays, different strategies have been developed in order to reduce the high atmospheric CO\textsubscript{2} concentration and appease the environmental problems associated with this greenhouse gas.\textsuperscript{3,6–10} In this context, employing atmospheric CO\textsubscript{2} as feedstock for chemical synthesis or conversion into hydrocarbon fuels represent a very promising strategy. From a productive point of view, using atmospheric CO\textsubscript{2} excess as feedstock can be seen as an advantage considering its scarce production cost associated. Within this production scheme, carboxylation reactions plays a fundamental role as an initial step within all proposed reaction mechanisms. In the last years, the formation of methanol (CH\textsubscript{3}OH), formic acid (HCOOH) and methane (CH\textsubscript{4}) is beginning to be actively explored.\textsuperscript{7,8,10–12} Several molecular devices have been proposed to obtain added-value chemical products from CO\textsubscript{2} molecules.\textsuperscript{1,3,7–15}

On the other hand, in Nature the CO\textsubscript{2} molecule constitutes the main carbon source for many autotrophic organism which is used as the fundamental building block to produce more complex molecules. Inorganic CO\textsubscript{2} is involved in different cellular metabolic pathways where different carboxylase enzymes catalyze its conversion.\textsuperscript{16} Carboxylases are an important class of enzymes that catalyze the fixation of CO\textsubscript{2} molecules into different organic substrates. Here, in the present report we have addressed a collection of the different carboxylase catalyzed reactions. Specifically, the carboxylation reaction catalyzed by Ribulose-1,5-bisphosphate carboxylase-oxygenase\textsuperscript{17,18} (RuBisCO) enzyme that is involved in
the main step of the Calvin-Benson-Bassham (CCB) cycle, the Crotonyl-CoA carboxylase-reductase\textsuperscript{19} (CCR) enzyme that is involved in assimilatory carboxylation reaction and the Phenyl-phosphate carboxylase\textsuperscript{20,21} (PPC) enzyme have been assessed. Likewise, the carboxylation reaction associated with the biotin molecule has been considered. Several carboxylases are assisted by the biotin cofactor in order to activate the CO\textsubscript{2} molecule.\textsuperscript{22,23}

For a detailed study of the carboxylation reactions associated with the inorganic and biological systems discussed above, it is paramount to find a correct description of the electronic structure of the CO\textsubscript{2} molecule and the other involved chemical species. In this context, high-level \textit{ab-initio} post-Hartree-Fock (post-HF) electronic structure methods like Coupled-Cluster (CC) serve to obtain correct reference values. The CC exponential Ansatz allows to obtain the electronic correlation energy at an arbitrarily high order while maintaining the size-extensibility of the system. Thus, the CC Ansatz provides an accurate electronic structure method for small to medium sized molecules.\textsuperscript{5,24–26} Specially, the Coupled-Clusters Singles and Doubles with a perturbative correction for Triples excitations (CCSD(T)) approach is currently considered as the "gold standard" in computational chemistry.\textsuperscript{5,24–26} The accuracy of this method is often reported to be within 1 kcal mol\textsuperscript{−1} and is located within the denominated "chemically accuracy". For very small molecules, higher-order excitations are often compensated and its contributions are neglected.\textsuperscript{24–26} The success of the CCSD(T) method is that in this scheme the triple excitation are included, which are very important in covalent and non-covalent interactions.\textsuperscript{24} Higher-order CC methods are not often required to determine molecular properties for very small molecular systems.

Here, in the present report a systematic study of the CO\textsubscript{2} electronic structure and its contribution to electronic correlation energy in carboxylation reactions for several biological and inorganic systems have been addressed with different high-level \textit{ab-initio} electronic structure methods of different hierarchy. The obtained post-HF reference values were then used to validate different Density Functional Theory (DFT) exchange-correlation (xc) functionals (DFT xc-functionals) in combination with various basis sets of different nature and cardi-
nality. For the studied systems a suitable balance between variety and simplicity of the methods have been prioritized in order to compute the electronic energy of all chemical species involved in the different carboxylation reactions at the best possible level of theory. This strategy allow us to lay a solid foundation to address carboxylation reactions in more complex environments as caboxylases. These biological systems are not often approachable for post-HF methods due to the high computational cost. In this context, the validation of different DFT xc-functionals can be used as attractive benchmark to assess chemical and electronic properties of these more relevant and complex molecular systems.

Computational details

Coupled-Cluster (CC) Calculation

The electronic energy associated with small molecules (H\(_2\), H\(_2\)O, CH\(_4\), CO, CO\(_2\), CH\(_3\)OH, HCOOH and CH\(_3\)COOH) considered in the present study has been determined by means of a hierarchy of CC wave-functions based model chemistries (WF-models), along with systematically increasing basis sets. Through these WF-models it is possible to progressively capture the electronic correlation energy by means of increasing the excitation rank in the coupled-cluster wave-functions. For the H\(_2\) molecule, SCF, MP2 and CCSD WF-models have have considered whilst in case of H\(_2\)O, CH\(_4\), CO, HCOOH (formic acid up to CCSDT(Q)) and CO\(_2\) molecules SCF, MP2, CCSD, CCSD(T), CCSDT, CCSDT(Q), CCSDTQ and CCSDTQ(P) WF-models have been used for the electronic structure computation. In the case of CH\(_3\)OH and CH\(_3\)COOH only SCF, MP2, CCSD and CCSD(T) WF-models have been employed due to high computational cost required. All these post-HF calculations have been performed by using different basis set with increasing cardinality. In general, for basis sets beyond the triple-\(\zeta\) level only SCF, MP2, CCSD and CCSD(T) WF-models have been considered. All of these calculation have been performed by using a correlation-consistent basis set (cc-pVxZ with x=D,T,Q,5; Dunning’s basis set) in order to systematically describe the basis
set dependence. Moreover, the self-consistent-field (SCF) and electronic correlation (Corl) energy in the complete basis set (CBS) limit have been obtained by means of a two-point extrapolation scheme proposed by Helgaker formula from two point obtained at cc-pVQZ and cc-pV5Z level. For the largest molecular systems (CH$_3$OH and CH$_3$COOH) the electronic energy in CBS limit was obtained by using the same two-point extrapolation scheme but from two point obtained at cc-pVTZ and cc-pVQZ. The reaction electronic energy was determined by means of the Focal point analysis technique (FPA). In this technique the reaction electronic energy at the CBS limit is obtained by summing the increments of the reaction electronic energy of every theory with respect to the preceding level in the hierarchy

SCF $\rightarrow$ MP2 $\rightarrow$ CCSD $\rightarrow$ CCSD(T) $\rightarrow$ CCSDT $\rightarrow$ CCSDT(Q) $\rightarrow$ CCSDTQ $\rightarrow$ CCSDTQ(P).

Electronic energy up to CCSD(T) level has been determined using the PSI4 program. For the post-CCSD(T) computations the MRCC software coupled to PSI4 program has been employed. All small molecules (H$_2$, H$_2$O, CH$_4$, CO, CO$_2$ and CH$_3$OH) have been optimized at CCSD(T)/cc-pVTZ level with the exception of formic and acetic acid (HCOOH, CH$_3$COOH) which have been optimized at CCSD/cc-pVTZ level. For all these molecules the geometry optimizations were carried out by using PSI4 software.

**Density Functional Theory (DFT) calculation**

In order to find a DFT xc-functional that is able to accurately describe the electronic energy of different carboxylation reactions, a representative set of DFT xc-functionals have been benchmarked against the obtained high-level ab-initio post-HF electronic energy values. All DFT computations have been performed using the ORCA (version 4.0) quantum chemistry package. To obtain the electronic energy and optimized geometry of all molecules here considered different DFT xc-functionals (PBE, BLYP, M06-L, B3LYP, PBE0, M06-2X, ωB97x, CAM-B3LYP, B2PLYP and B2GP-PLYP) have been used. The employed DFT xc-functionals can be classified as gradient corrected functionals (PBE and BLYP, GGA functionals), hybrid functionals (B3LYP and PBE0, HGGA functionals), meta-GGA functionals
(M06-L and M06-2X, HmGGA functionals), range-separated hybrid functional (\(\omega\)B97x and CAM-B3LYP) and perturbatively corrected double-hybrid functional (B2PLYP and B2GP-PLYP) in accordance with the Jacob’s ladder classification scheme. Likewise, the influence of the DFT xc-functional on the geometry has been included in order to assess in a more general way the performance of these DFT xc-functionals. For each DFT xc-functional employed an optimized geometry is obtained and its electronic energy is assessed with the remaining DFT xc-functionals in order to produce a matrix arrangement that contain the electronic energy for different geometries. Several basis sets have been used for all these calculation in order to produce each matrix arrangement: def2-SVP, def2-TZVP, def2-TZVPP, def2-QZVP and def2-QZVPP (Karlsruhe’s basis set) and 6-31g, 6-31g(d), 6-311g(d) and 6-311g-(2d,f) (Pople’s basis set). Likewise, the root-mean-square deviation (RMSD) associated with all chemical species involved in the different biochemical systems here considered has been computed, in order to assess the effect of geometry over the obtained carboxylation reaction electronic energy values. In this case, the obtained geometry at \(\omega\)B97x-D3/ma-def-QZVP level has have been used as reference geometry for this purpose. This particular DFT xc-functional have been chosen due to its performance on the carboxylation reaction electronic energy for all small inorganic systems here discussed.

**Result and Discussion**

The study of inorganic and biochemical carboxylation reactions have been addressed by means of different electronic structure methods. From a physico-chemical perspective the observable that can be measured experimentally is the reaction enthalpy. This observable can be calculated considering the change in electronic energy of products and reactants and adding the zero-point energy corrections and thermal energies, which represent new sources of errors in its calculation. To exclude additional error sources in this kind of computations we have decided to focus on the electronic energy contribution to the reaction enthalpy,
which will be referred hereinafter as reaction electronic energy (a quantity that includes nuclei-nuclei repulsion term). The validation has been performed for different electronic structure methods against high level ab-initio calculations at the CBS limit.

In the present report the reaction electronic energy of a variety of carboxylation reactions involved in different biochemical and inorganic processes has been determined. The study of carboxylation reactions have been addressed classifying the chemical reactions in inorganic and biochemical ones. This separation allows us to explore the small inorganic systems to a high-level of theory by means of different high-level ab-initio post-HF electronic structure methods. For example, in the case of the smallest molecular systems an electronic correlation energy treatment up to CCSDTQ(P) level have been included. The obtained main features in these low computational-cost systems will be used to interpret the result of more complex and computationally demanding biochemical systems.

**Inorganic carboxylation reactions**

A small collection of inorganic carboxylation reactions have been selected in order to assess its reaction electronic energy by means of different electronic structure methods including post-HF and DFT methods with different DFT xc-functionals. Some of the selected carboxylation reactions are involved in different elementary steps to obtain valuable chemical products employing CO$_2$ as feedstock. The following inorganic carboxylation reactions have been considered:

\[
\begin{align*}
\text{R1} & : & \text{CO}_2 + 4\text{H}_2 & \rightarrow \text{CH}_4 + 2\text{H}_2\text{O} \\
\text{R2} & : & \text{CO}_2 + \text{H}_2 & \rightarrow \text{CO} + \text{H}_2\text{O} \\
\text{R3} & : & \text{CO}_2 + \text{CH}_4 & \rightarrow 2\text{CO} + 2\text{H}_2 \\
\text{R4} & : & \text{CO}_2 + \text{H}_2 & \rightarrow \text{HCOOH} \\
\text{R5} & : & \text{CO}_2 + \text{CH}_4 & \rightarrow \text{CH}_3\text{COOH}
\end{align*}
\]
In this sequence of chemical reactions it is possible to identify carboxylation reactions associated with the formation of methane (CH$_4$, reaction R1, reverse methane steam reforming), carbon monoxide (CO, reaction R2 and R3, reverse water-gas shift reaction), formic (HCOOH, reaction R4) and acetic (CH$_3$COOH, reaction R5) acid, and methanol (CH$_3$OH, reaction R6, reverse methanol steam reforming).$^{7,8,10-12}$

Post-HF electronic structure methods at the CBS limit

The reaction electronic energy ($\Delta E_R$) is a global property that relies on the electronic energy ($E_j$) and stoichiometric coefficients ($\nu_j$) of each chemical species $j$ considered within the global chemical equation,

$$\Delta E^M_i = \sum_{\text{prod}} \nu_j E^M_i - \sum_{\text{reac}} \nu_j E^M_i$$ \hspace{1cm} (1)

Here, $E^M_j$ is the electronic energy associated with chemical specie $j$ calculated with a ($M_i$) particular electronic structure method and $\Delta E^M_i$ corresponds to the obtained reaction electronic energy combining and weighting the calculated electronic energy values. In general, the computed reaction electronic energy value ($\Delta E^M_i$) is compared with a suitable reference value ($\Delta E^\text{Ref}_R$) that is obtained from high-level ab-initio calculations extrapolated at the CBS limit (or an experimental result). The reference value can be obtained by using the same previous equation but replacing the electronic energy value by extrapolated reference electronic energy value ($E^\text{Ref}_j$),

$$\Delta E^\text{Ref}_R = \sum_{\text{prod}} \nu_j E^\text{Ref}_j - \sum_{\text{reac}} \nu_j E^\text{Ref}_j$$ \hspace{1cm} (2)
The difference between both values corresponding to the reaction electronic energy error \( \Delta(\Delta E_R) \), which is a metric of how well the reaction electronic energy is described by means of the \( M_i \) electronic structure method, and can be written in these two equivalent forms:

\[
\Delta(\Delta E_R) = \Delta E_{R}^{M_i} - \Delta E_{R}^{Ref} \tag{3}
\]

or

\[
\Delta(\Delta E_R) = \sum_{prod} \nu_j (E_{M_i}^{j} - E_{j}^{Ref}) - \sum_{reac} \nu_j (E_{M_i}^{j} - E_{j}^{Ref}) \tag{4}
\]

Defining \( \delta E_j = E_{M_i}^{j} - E_{j}^{Ref} \) as a new variable that quantifies the quality of a \( M_i \) particular electronic structure method describing the electronic energy of each chemical species \( j \), the previous equation can be written as,

\[
\Delta(\Delta E_R) = \sum_{prod} \nu_j \delta E_j - \sum_{reac} \nu_j \delta E_j \tag{5}
\]

At this point, it is important to stress that previous general analysis can be performed through both theoretical frameworks, namely, by means of both DFT and post-HF electronic structure methods. However, an important feature, which can be taken advantage only in post-HF methods, is its systematic nature to include the electronic correlation energy. In this theoretical framework, the recovered electronic correlation energy by means of a particular electronic structure method \( (M_i) \) is added to the Hartre-Fock (or SCF) energy making it possible to separate the computed electronic energy \( (E_{j}^{M_i}) \) for each \( j \) chemical specie in its two-components: self-consistent-field (SCF) and electronic correlation (Corl) energy.

\[
E_{j}^{M_i} = E_{j}^{SCF,M_i} + E_{j}^{Corl,M_i} \tag{6}
\]
If the electronic energy for each chemical specie $j$ is separated in its two contributions the following expression can be obtained for $\delta E_j$,

$$\delta E_j = (E_j^{SCF,M_i} + E_j^{Corl,M_i}) - (E_j^{SCF,Ref} + E_j^{Corl,Ref})$$  \hspace{1cm} (7)

and the errors can be separated in $\delta E_j^{Corl} = E_j^{Corl,M_i} - E_j^{Corl,Ref}$ and equivalently $\delta E_j^{SCF}$. Equation 7 can be now expressed as,

$$\delta E_j = \delta E_j^{SCF} + \delta E_j^{Corl}$$  \hspace{1cm} (8)

Here, $\delta E_j^{Corl}$ is a metric of how well the electronic correlation energy of each chemical species $j$ is described by means of the M$_i$ electronic structure method. Similar meaning is ascribed to $\delta E_j^{SCF}$. Therefore, the reaction electronic energy error in the equation 5 can be decomposed as:

$$\Delta(\Delta E_R) = \left\{ \sum_{prod} \nu_j \delta E_j^{SCF} - \sum_{reac} \nu_j \delta E_j^{SCF} \right\} + \left\{ \sum_{prod} \nu_j \delta E_j^{Corl} - \sum_{reac} \nu_j \delta E_j^{Corl} \right\}$$  \hspace{1cm} (9)

or

$$\Delta(\Delta E_R) = \Delta(\Delta E_R^{SCF}) + \Delta(\Delta E_R^{Corl})$$  \hspace{1cm} (10)

A schematic representation of both contributions can be found in the Figure 1. Here, $\Delta(\Delta E_R^{Corl})$ depends on the description of electronic correlation energy for each chemical species $j$ involved in the analyzed chemical reaction.

As example of this procedure, let us consider reaction R1. In this case, the following expression can be obtained for determining the carboxylation reaction electronic energy
Figure 1: Schematic representation of the two components (\(\Delta(\Delta E_{R}^{SCF})\) and \(\Delta(\Delta E_{R}^{Corl})\)) that contribute to the reaction electronic energy error (\(\Delta(\Delta E_{R})\)) in the post-HF methods. The SCF and correlation reaction electronic energy recovered by means of the \(M_{i}\) electronic structure method have been highlighted in blue and orange color, respectively (left). The height of each box drawn with solid line represent the reference reaction electronic energy value for both components (center) and dotted box represent the SCF and electronic correlation reaction electronic energy error associated to the \(M_{i}\) electronic structure method (right).

\[
\Delta(\Delta E_{R}) = \Delta(\Delta E_{R}^{SCF}) + \Delta(\Delta E_{R}^{Corl})
\]

(11)

Basically, the quality in the SCF and electronic correlation energy description for each \(j\) chemical specie represents the key factor to determine the reaction electronic energy (\(\Delta(\Delta E_{R})\)) error (see equation 11). Through a division of electronic energy in these two contributions it is possible to identify different sources of error that might contribute to this metric to assess different electronic structure method. In the present study a systematic analysis on the contribution of the different components to the reaction electronic energy error associated with the post-HF methods for typical inorganic carboxylation reactions has been carried out.
Focal point analysis of reactions R1-R6

As a first step in our study, the electronic energy ($E^{M_i}_j$) of all chemical species involved in six inorganic carboxylation reactions (reaction R1-R6) have been computed by means of several CC calculation with different WF-models. The most accurate electronic energy values have been combined to compute the reference reaction electronic energy for all inorganic carboxylation reactions here considered. The inclusion of CC higher rank excitation terms allowed us to explore the convergence in electronic correlation energy for each molecule, and therefore for each chemical reaction. For the sake of clarity, reactions R1 and R4 have been considered in more details since they represented two limiting cases, where the compensation in the description of electronic correlation energy is the main factor for the accuracy of the obtained value. Likewise, reactions R2 and R3 have been discussed as intermediate cases between the two former mentioned ones. Table 1 and 2 show the reaction electronic energy values calculated with different WF-models for reactions R1 and R4 whilst for reactions R2 and R3 they are summarized in Table S18 and S19 in supplementary materials. The extrapolated value of the reaction electronic energy for reaction R1 is -61.457 kcal mol\(^{-1}\) whilst the values obtained for R2, R3 and R4 were: +6.392 kcal mol\(^{-1}\), +74.241 kcal mol\(^{-1}\) and -3.696 kcal mol\(^{-1}\) at the CBS limit, respectively.

Table 1: Focal point analysis of the reaction electronic energy of R1 (values are in kcal mol\(^{-1}\)).

<table>
<thead>
<tr>
<th>Basis set</th>
<th>$\Delta E_{\text{RHF}}$</th>
<th>+$\delta$MP2</th>
<th>+$\delta$CCSD</th>
<th>+$\delta$CCSD(T)</th>
<th>+$\delta$CCSDT</th>
<th>+$\delta$CCSDT(Q)</th>
<th>+$\delta$CCSDTQ</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>cc-pVDZ</td>
<td>-54.716</td>
<td>+14.362</td>
<td>-7.763</td>
<td>+5.239</td>
<td>-0.294</td>
<td>+0.620</td>
<td>-0.168</td>
<td>-42.772</td>
</tr>
<tr>
<td>cc-pVTZ</td>
<td>-56.150</td>
<td>+2.729</td>
<td>-6.880</td>
<td>+4.455</td>
<td>-0.469</td>
<td>+0.732</td>
<td>-</td>
<td>-55.583</td>
</tr>
<tr>
<td>cc-pVQZ</td>
<td>-57.165</td>
<td>+0.953</td>
<td>-6.281</td>
<td>+4.186</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-58.306</td>
</tr>
<tr>
<td>cc-pV5Z</td>
<td>-57.039</td>
<td>+0.108</td>
<td>-5.979</td>
<td>+4.098</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-59.812</td>
</tr>
<tr>
<td>CBS</td>
<td>-58.251</td>
<td>-1.482</td>
<td>-5.663</td>
<td>+4.005</td>
<td>-0.502</td>
<td>+0.655</td>
<td>-0.168</td>
<td>-61.457</td>
</tr>
</tbody>
</table>

The symbol $\delta$ denotes the increment in the relative energy ($\Delta E_e$) with respect to the preceding level of theory in the hierarchy SCF $\rightarrow$ MP2 $\rightarrow$ CCSD $\rightarrow$ CCSD(T) $\rightarrow$ CCSDT $\rightarrow$ CCSDT(Q) $\rightarrow$ CCSDTQ(PS). Final extrapolated value are boldfaced.

Table 2: Focal point analysis of the reaction electronic energy of R4 (values in kcal mol\(^{-1}\)).

<table>
<thead>
<tr>
<th>Basis set</th>
<th>$\Delta E_{\text{RHF}}$</th>
<th>+$\delta$MP2</th>
<th>+$\delta$CCSD</th>
<th>+$\delta$CCSD(T)</th>
<th>+$\delta$CCSDT</th>
<th>+$\delta$CCSDT(Q)</th>
<th>+$\delta$CCSDTQ</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>cc-pVDZ</td>
<td>-2.198</td>
<td>+6.349</td>
<td>-4.705</td>
<td>+2.929</td>
<td>-0.126</td>
<td>+0.290</td>
<td>-</td>
<td>-2.005</td>
</tr>
<tr>
<td>cc-pVTZ</td>
<td>-1.807</td>
<td>+2.330</td>
<td>-4.283</td>
<td>+1.661</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.941</td>
</tr>
<tr>
<td>cc-pVQZ</td>
<td>-1.764</td>
<td>+1.995</td>
<td>-4.129</td>
<td>+1.567</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-2.570</td>
</tr>
<tr>
<td>cc-pV5Z</td>
<td>-1.936</td>
<td>+1.735</td>
<td>-4.098</td>
<td>+1.540</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-2.758</td>
</tr>
<tr>
<td>CBS</td>
<td>-1.978</td>
<td>+1.325</td>
<td>-4.065</td>
<td>+1.513</td>
<td>-0.126</td>
<td>+0.290</td>
<td>-</td>
<td>-3.696</td>
</tr>
</tbody>
</table>

The symbol $\delta$ denotes the increment in the relative energy ($\Delta E_e$) with respect to the preceding level of theory in the hierarchy SCF $\rightarrow$ MP2 $\rightarrow$ CCSD $\rightarrow$ CCSD(T) $\rightarrow$ CCSDT $\rightarrow$ CCSDT(Q). Final extrapolated value are boldfaced.
The electronic correlation energy increments show that the reaction electronic energy is converged to chemical accuracy at the CCSD(T)/CBS level of theory, since quadruple excitation correction do not significantly affect the reaction electronic energy. Nonetheless, these were necessary to assure the convergence of the reaction electronic energy, since especially in R1 the CCSD(T) electronic correlation energy correction is significant and convergence of the correlation energy cannot be guaranteed at that level of theory. It is also worth noting that in R1 there is a noticeable basis set dependence in the MP2 electronic correlation energy contribution to the reaction electronic energy. This is to be expected due to the poor description of the electron-electron cusp when using mono electronic basis sets. The dependence of the basis set in the electronic correlation energy diminishes when corrections due to correlation of three or more electrons are considered, as in CCSD and higher order corrections. But, even for this highly correlated methods the observed error is still significant at double-ζ and triple-ζ level, ~ 19 kcal mol\(^{-1}\) and ~6 kcal mol\(^{-1}\) respectively. This behavior is different in reaction R4 where the CCSD(T) contribution is smaller together with a reduced basis set dependence (similar to reactions R2-R3). Finally, an oscillating behavior in the electronic correlation energy corrections from MP2 up to CCSD(T) can be observed.

In pursuance of the culprit for the observed poor convergence, a detailed analysis of the electronic correlation energy of each species participating in these chemical reactions is done.

**Electronic correlation energy analysis of reactants and products**

The analysis of the electronic correlation energy of the chemical specie involved in the six inorganic carboxylation reactions is performed first for the CO\(_2\) molecule because it is represented in each one of them. The different corrections to the CO\(_2\) electronic correlation energy calculated up to CCSDTQ(P) excitation at the CBS limit are shown in Table 3. As in the focal point analysis each correction to the electronic correlation energy corresponds to the difference among the successive excitation ranks implying that a small correction cor-
responds to a converged electronic correlation energy. Table 3 shows that the contribution

to the electronic correlation energy at MP2 level is significant and depends markedly on the
cardinality of the basis set (\textit{vide supra}). Slow convergence of the electronic correlation energy
with the excitation rank is observed compared to the other chemical species participating
in the reactions as H$_2$O and CH$_4$ (see Tables S3-5 in supplementary materials). As will be
shown below this can be attributed to the delocalized nature of the $\pi$-system associated with
the ground state electronic configuration of CO$_2$. Therefore, it is preponderant to incorpo-
rate terms in the CC ansatz which describe the correlation of three and up to four electrons,
as in the CCSDT and CCSDTQ methods.

From the other chemical species carbon monoxide (CO), HCOOH and CH$_3$COOH exhibit
faster convergence with respect to the CC excitation rank than CO$_2$ but slower than H$_2$O
and CH$_4$ as can be seen in Table S1,S2 and S7 in the supplementary material. An important
difference between CO$_2$ and CO or the carboxylic acids is found in the relative contributions
of the CCSD(T) and CCSDT(Q) corrections to the electronic correlation energy. In the case
of CO$_2$ both relative contributions turned out to be markedly bigger.

Interestingly, the slow convergence and noticeable basis set dependence of the electronic
correlation energy corrections to the reaction electronic energy that have been discussed in
the focal point analysis above are also reflected in the trends found in the case of CO$_2$ and
partly also in HCOOH. In both cases, this behavior is verified with basis set size and CC
excitation rank considered.

Yet, in general it is not possible to compare absolute electronic correlation energy values
between different chemical species because of its extensive nature. At first glance, it sounds
reasonable that those molecules that contain electrons in a $\pi$-systems as CO$_2$ and CO may give rise to large electronic correlation energies since the electrons share the same region of space. To be able to make a qualitative comparison between the small molecules of this study a normalized electronic correlation energy per molecule is proposed. This normalized correlation energy results from the division of the total electronic correlation energy obtained at the CBS limit for each chemical specie with its total number of valence electrons. The obtained values are in atomic units and decreasing order: H$_2$: -0.0205 , CH$_4$: -0.0356 , CH$_3$OH: -0.0428 , H$_2$O: -0.0447 , CH$_3$COOH: -0.0466 , HCOOH: -0.0486 , CO: -0.0505 and CO$_2$: -0.0519 . CO and CO$_2$ exhibit the largest absolute values in accordance with the correlated $\pi$-electrons that concentrate in a small region between the atoms and could result in larger electron-electron correlation. These two molecules are followed by the two carboxylic acids, and the molecules with single bonds present the smallest values.

In this context, it is possible to explain the need to include higher CC excitation to describe the electronic energy of molecules with $\pi$ bonds as CO$_2$, CO or the carboxylic acids. An incomplete description would alter the calculated reaction electronic energies from the focal point analysis.

Another important feature that can be analyzed in addition to the amount of electronic correlation energy is its nature for each molecule. In general, the main contribution to the electronic correlation energy stems from electron pair interaction in tune with the Pauli exclusion principle. For some correlated systems higher order (three and four body interactions) contributions become more relevant, as it has been discussed above for CO and CO$_2$. To quantify the different kind of contributions to the electronic correlation energy the following quotient $\alpha = \frac{E_{CBS}}{E_{CBS}}$ has been defined. This quotient identifies the relative contribution of the triple excitation to the total electronic correlation energy. The obtained $\alpha$ values for all molecules involved in the inorganic carboxylation reaction are: $\alpha_{CH_4} = 0.02843$, $\alpha_{CH_3OH} = 0.03140^*$, $\alpha_{H_2O} = 0.02941$, $\alpha_{CH_3COOH} = 0.03685^*$, $\alpha_{HCOOH} = 0.03788$, $\alpha_{CO} = 0.04151$ and $\alpha_{CO_2} = 0.04179$ (the value with a asterisk corresponds to the
CCSD(T) level). This sequence shows that the contribution of triple excitation to the electronic correlation energy is larger in the case of CO and CO$_2$. A possible explanation is that in these molecules more than two electrons are located in the region between two atoms due to double/triple bonds, in which case the contribution of electronic correlation from three and four electrons becomes more important. Interestingly, $\alpha_{CH_3COOH}$ and $\alpha_{HCOOH}$ with only one double bond present values located in the middle range compared to the other molecules with only single bonds.

As a partial conclusion from this analysis one can conclude that both molecules CO$_2$ and HCOOH exhibit similar electronic correlation energy contribution which is only described correctly by triple and quadruple excitations. This electronic behavior is also verified for the CO molecule but its absolute value is smaller due to the reduced number of electrons. Therefore, it is possible that in reactions involving these species an incomplete electronic correlation energy description of some electronic structure methods give rise to very large values of $\delta E^{\text{Corl}}_{\text{HCOOH}}$, $\delta E^{\text{Corl}}_{\text{CO}_2}$ and $\delta E^{\text{Corl}}_{\text{CO}}$ which result in a large reaction electronic energy error $\Delta(\Delta E_R)$ (see supplementary material Table S11-13).

For small molecules with single bonds (H$_2$, CH$_4$ or H$_2$O) it is not necessary to include post-CCSD(T) contributions to obtain a suitable description of electronic correlation energy as the gold standard suggests.

**Error analysis of the reaction electronic energy**

With the analysis of the electronic correlation energy for each chemical specie $j$ involved in the six inorganic carboxylation reaction, the different error sources associated with the reaction electronic energy for each post-HF *ab-initio* electronic structure method in the focal point analysis were identified. Now, the reaction electronic energy error ($\Delta(\Delta E_R$)) will be separated into two contributions: SCF ($\Delta(\Delta E^{\text{SCF}}_R$)) and correlation ($\Delta(\Delta E^{\text{Corl}}_R$)) energy contributions (see equation 9 or 10). As it has been mentioned before, this strategy turn out to be an interesting way for detecting different error sources in this kind of calculations.
Each of these terms is determined by the contribution of each \( j \) chemical species involved in the global chemical equation (see equation 11 as example of reaction \( R_1 \)). The quality in the SCF and electronic correlation energy description for different WF-models and for all molecules involved in the chemical reactions \( R_1-R_4 \) are shown in the Table S8(H\(_2\)), S9(H\(_2\)O), S10 (CH\(_4\)), S11 (CO\(_2\)), S12 (HCOOH) and S13 (CO) in the supplementary material. The best description to the SCF and electronic correlation energy at the CBS limit has been taken as reference to determine the quality of all WF-models and basis sets used,

\[
\delta E_{j}^{\text{Corl}} = E_{j}^{\text{corl,M}_i} - E_{j}^{\text{corl,CBS}} \\
\delta E_{j}^{\text{SCF}} = E_{j}^{\text{SCF,M}_i} - E_{j}^{\text{SCF,CBS}}
\]

At first glance, all values converge monotonically to zero when the cardinality of the basis set is increased and this tendency is observed for all species. Likewise, in all cases the error of the SCF energy difference is smaller than the error of the electronic correlation energy difference.

\[
\Delta(\Delta E_{R}^{\text{SCF}}) << \Delta(\Delta E_{R}^{\text{Corl}})
\]

Therefore, the focus will be put on the analysis of the quality of the electronic correlation energy description for all species. Observed deviations that have been identified for all chemical species (\( \delta E_{j}^{\text{Corl}} \)) associated with six inorganic carboxylation reactions will help to rationalize the reaction electronic energy error (\( \Delta(\Delta E_{R}) \)).

This will be first exemplified with reaction \( R_4 \) where the structural similarity criterion between reactants and products is most pronounced. Assuming that the electron correlation energy stemming from the electron involving hydrogen atoms cancel out, one could think that the product, HCOOH, should present an electronic correlation energy similar to a bent CO\(_2\) molecule with an angle \( \theta \) similar to the one in the HCOOH molecule. To quantify the change in the CO\(_2\) electronic correlation energy with the angle, the difference between the bent and the linear CO\(_2\) geometry at the CCSD(T)/CBS level are shown in Figure 2.
At $\theta = 127^\circ$ the CO$_2$ electronic correlation energy is decreased by -12.5 kcal mol$^{-1}$. The increase of electronic correlation can be related to the electrons in the $\pi$ molecular orbitals coming closer and therefore increasing its value.

Figure 2: Variation of the CO$_2$ electronic correlation energy calculated at CCSD(T)/CBS level as a function of the O=C=O angle with respect to the value in the CO$_2$ linear configuration.

However, this geometrical changes in the product are also accompanied by the change of the hybridization of the carbon atom. A carbon-oxygen double-bond of CO$_2$ is broken to form two single-bonds (C-O and C-H) within HCOOH. This hybridization change from a $\pi$-bond to two $\sigma$-bond breaks down the delocalized $\pi$—system present in bent CO$_2$ molecule and based on the analysis of the electronic correlation energy above should reduce the total electronic correlation energy in HCOOH.

Comparing $\delta E^\text{Corl}_j$ of CO$_2$ and HCOOH (see Table S11-12) a similar contributions for each WF method and basis set is observed. This indicates that the electronic correlation of reactants (CO$_2$) and products (HCOOH) would cancel out, and that the electronic correlation increase in the bending is partly compensated by the lost in electronic correlation energy from the broken double bond. The similar contribution of electronic correlation energy of reactants and products is reflected further in the contribution of the electronic correlation to the total reaction electronic energy that amounts to only -1.23 kcal mol$^{-1}$. One can conclude
that in reaction R4 the reactants, especially CO₂, and the products share the same amount of electron correlation. Any electronic structure method that describes with the similar quality both molecules would yield similar electronic correlation energy error ($\delta E_{\text{HCOOH}}^{\text{Corl}} \approx \delta E_{\text{CO}_2}^{\text{Corl}}$), which will cancel for the reaction electronic energy as shown in Figure 3 and in Table 2.

However, this does not hold for the chemical species involved in the reaction R1 as it is schematically shown in the Figure 3. From the analysis of the electronic correlation energy it has been concluded, that for CO₂ quadruple excitation and basis sets larger than triple-ζ are required for a proper quantitative description of the electronic correlation energy, whereas for the rest of the single bonded reactants and products a lower excitation rank and basis set suffices. The neglect of higher excitation rank in the calculation of the reaction electronic energy would result in large errors, which are mostly attributable to the error of the electronic correlation energy in CO₂. The errors associated with the neglect of higher excitation ranks observed in reaction R4 are smaller than for R1 although CO₂ is still present as reactant. The reason is a compensation of the error in the electronic correlation energy in CO₂ with the error in the HCOOH product which is similar in magnitude as discussed above (see Figure 3). The very large error detected in the reaction electronic energy for reaction R1 is explained given the large global correlation energy contribution ($\Delta(\Delta E_R^{\text{Corl}})$) for this reaction stemming from the CO₂ molecule. In this case, there is no compensation of the individual electronic correlation energy ($\delta E_j^{\text{Corl}}$) description among all chemical species. An incomplete description of electronic correlation energy can, therefore, result in a good description of the reaction electronic energy value only if a compensation is reached as in the chemical reaction R4 (see Figure 5) even for computational less demanding WF-models like MP2.

The compensation of errors in calculating reaction enthalpies to subchemical accuracy has already been analyzed previously for reactions involving alkanes. In this context rigorous categories were defined as isodesmic and different variants of homodesmotic reactions. The categories are based on equal number of bonds and bond types, defined only by the elements
participating in the bond or including their hybridization as in carbon atoms. Recently, this concept has been also applied to molecules containing oxygen and nitrogen. However, the carboxylation reactions in this study do not belong to any of these definitions because in none of them bond types are conserved. Yet, all reactions are assigned to the original proposed class of isogyric reactions with equal number of electron pairs in reactants and products. For this kind of reactions our analysis provides additional insights to possible error cancellations depending on the description of the electronic correlation energy of reactants and products. In this respect $\sigma$ and $\pi$ bonds of reactant and products contribute differently to the correlation energy errors, which can cancel out in the calculation of the reaction electronic energy as reaction $\textbf{R3}$ reveals. The detailed analysis provided above, therefore, permits a rational analysis of the factors which contribute to the absolute error and might represent a first step to extend the concept of isodesmic and homodesmic reactions to reactions involving carbon and oxygen with varying number of $\sigma$ and $\pi$ bonds.

![Figure 3: Schematic representation of the electronic correlation energy recovered by means of $M_i$ electronic structure method for each $j$ chemical species involved in inorganic carboxylation reactions $\textbf{R1}$ and $\textbf{R4}$. The electronic correlation energy associated with products and reactants have been represented with different colors. The height of each box represent the reference electronic correlation energy value for each chemical specie. For each chemical equation the reaction correlation energy error can be calculated from: $\Delta E^\text{Corl}_R = (2\delta E^\text{Corl}_{H_2O} + \delta E^\text{Corl}_{CO_2} + 4\delta E^\text{Corl}_{H_2}) - (\delta E^\text{Corl}_{H_2} + 4\delta E^\text{Corl}_{H_2O})$ and $\Delta E^\text{Corl}_R = (\delta E^\text{Corl}_{H_2O} + \delta E^\text{Corl}_{HCOOH}) - (\delta E^\text{Corl}_{HCOOH})$, respectively.](image)

Up to now, the focus has been put on the compensation of the total electronic correlation energy. Another important feature that is worth to be assessed is the compensation for different types of electronic correlation energy. The CCSD(T) contribution associated with
reactions R1-R4 turned out to be: +4.005 kcal mol\(^{-1}\), +2.3005 kcal mol\(^{-1}\) and +0.605 kcal mol\(^{-1}\) and +1.513 kcal mol\(^{-1}\), respectively. For the reaction R1, the only chemical specie contributing to this kind of electronic correlation is CO\(_2\) that exhibits a very large contribution of triple excitation which is not compensated by the remaining chemical species. In the case of reactions R2 and R3 a partial or complete compensation for this type of electronic correlation energy takes place because in the CCSD(T) excitations only CO and CO\(_2\) present a significant contribution although with different absolute value. In the case of reaction R2 the compensation is not reached completely but in reaction R3 where two molecules of CO are formed an almost complete cancellation is observed.

Although the previous discussion has been focused on R1-R4 the observed errors in the reaction electronic energies show the same behavior for the reactions R5-6, which are similar to reaction R4 and R1 respectively. Their values are reported in Table S16-17 in the supplementary material.

From this first analysis of several inorganic carboxylation reaction, we can conclude that there is an important electronic correlation energy correction in the species involving delocalized \(\pi\)-systems which can produce large errors in the calculation of the reaction electronic energy values if no compensation between reactants and products take place. In that line, the CCSD(T) and even higher rank contributions are essential to obtain chemical accuracy with errors below 1 kcal mol\(^{-1}\), especially when the double bonds in CO\(_2\) are transformed to single bonds as in reaction R1.

In the next step this accurate high-level \textit{ab-initio} electronic structure methods will be used to address the performance of different DFT methods for the same chemical reactions.

**DFT results (benchmark)**

A representative collection of ten DFT xc-functionals have been chosen to study the six inorganic carboxylation reactions with different types of basis sets. For these calculations
the reference values discussed above were used in order to determine the reaction electronic energy error ($\Delta(\Delta E_R)$) for all these chemical reactions. Pople’s (6-31g, 6-311g, 6-31g(d), 6-311g(d) and 6-311g(2d,f)) and Karlsruhe’s (def2-SVP, def2-TZVP, def2-QZVP, def2-TZVPP, def2-QZVPP) basis set with different amount of polarization function have been assessed. Both geometry and energy have been considered in order to obtain more information about the quality of the assessed DFT xc-functionals to describe the reaction electronic energy. The possible combinations of energy and geometry ($E_{i,j}$) have been sorted in matrix form with a color code to indicate the signed error associated with the respective reference reaction electronic energy (see Figure 4). The same color scale has been used to quantify all signed errors associated with the six inorganic carboxylation reactions. The energy values located over diagonal matrix elements ($E_{j,j}$) represent the reaction electronic energy error calculated with same DFT xc-functional $j$ that has been used to obtain the optimized geometry for all chemical species involved. On the other hands, the signed errors located over the same row $j$ represent the reaction electronic energy error ($\Delta(\Delta E_R)$) calculated with DFT xc-functional $i$ and the optimized geometry from DFT xc-functional $j$ ($E_{j,i}$). The DFT results associated with the chemical reactions $R_1$ and $R_4$ with the def2-QZVP basis set are shown in Figure 4 and 5, respectively. Likewise, the results for the same reactions with different Pople’s and Karlsruhe’s basis sets are shown in Figure S1-4 within the supplementary material.

In general, all DFT xc-functionals give rise to very similar optimized geometries because the reaction electronic energies in one column are very similar (similar color tone) with some subtle differences. Due to the small size of the molecules and the simple geometries this is expected for the different DFT xc-functionals. As in the post-HF methods case, the reaction electronic energy error for reaction $R_1$ turns out to be much larger than the reaction electronic energy error obtained for reaction $R_4$. If the DFT methods would present the same behavior as post-HF method the reason for the larger error in reaction $R_1$ is the important contribution of the electronic correlation energy which is not correctly described by some DFT methods. A systematic reduction in the reaction electronic energy error can be
Figure 4: Signed error in DFT reaction electronic energy calculated with different xc-functionals at def2-QZVP level with respect to reference reaction electronic energy value determined at CCSDTQ(P)/CBS level (−61.457 kcal mol\(^{-1}\)) for the chemical reaction \(R_1\).

detected when the Karlsruhe’s basis set is used compared to the Pople’s basis sets and when the cardinality of the basis set is increased in the Karlsruhe’s basis set (see supplementary materials). This systematic behavior with the size of the basis set is not found for the Pople’s basis set. In this case, the inclusion of polarization functions produces an increase in the error as can be seen in the supplementary material. One concludes that the Karlsruhe’s basis set follows the tendency that has been found for the Dunning’s basis set with post-HF methods with regards to the effect of the basis set size, although in DFT no clear improvement with basis set size can be expected a priori. Another interesting feature that can be observed in the DFT result is that double-hybrid DFT xc-functionals give rise to large errors in the reaction electronic energy. This behavior could be related with the poor description of the electronic correlation energy that has been found at MP2 level in the post-HF methods.

The tables of the remaining chemical reactions \(R_2-3\) and \(R_5-6\) show some similar features but no systematic study of the error is possible (see Figure S5 in the supplementary material). Reaction \(R_3\) for example presents the largest error in some functionals and the total error in most cases is larger than in reaction \(R_2\), although in the post-HF methods the electronic correlation error was shown to cancel out in the former reaction. Therefore, it is very difficult
Figure 5: Signed error in DFT reaction electronic energy calculated with different xc-functionals at a def2-QZVP level with respect to reference reaction electronic energy value determined at CCSDT(Q)/CBS level (-3.696 kcal mol$^{-1}$) for the chemical reaction R4.

To anticipate how a certain DFT method will perform since no error cancellation is observed. In most cases, recent DFT xc-functionals including HF exchange or range separation perform better than pure DFT xc-functionals. Comparing all results the M06-2X functional followed by the $\omega$B97x DFT xc-functional with the def2-QZVP basis set show the best performance of the studied set. The latter, however, shows the smallest absolute deviation for all six inorganic carboxylation reaction which might be related to its smaller number of parameters.

**Biochemical carboxylation reactions**

To extend our study of the reaction electronic energy towards more complex molecular systems, a set of biochemical carboxylation reactions have been assessed. Biochemical carboxylation reactions are catalyzed by carboxylases, a set of enzymes highly specialized in CO$_2$ fixation. The selected carboxylases for this study are involved in several cellular metabolic pathway as RuBisCO in the Calvin (CCB) cycle. For these more complex molecular systems, reaction electronic energies have been computed by means of different DFT methods employing the same DFT xc-functionals benchmarked in the inorganic carboxylations reactions discussed above. The reference values have been computed at CCSD(T)/aug-cc-pVDZ level.
with geometries optimized with the ωB97x-D3/ma-def2-QZVP method to assess the performance of different DFT xc-functionals. In the case of the largest molecular systems (BIOT and CCR system, see Figure 6) only excitation associated with the valence electrons have been considered to compute the CCSD(T) reaction electronic energy due to the computational cost. From similar inorganic carboxylation reactions where carboxylated products are formed it is possible to estimate an error margin of the reference values for these biochemical carboxylations reactions of 2-3 kcal mol\(^{-1}\) due to use of a smaller basis set and the missing quadruple and higher excitation in the CC ansatz. For all DFT calculations an unique basis set with diffuse functions have been used (ma-def2-TZVPP) in order to describe the anionic species involved in this kind of systems. Moreover, due to the larger size of the molecules involved in these biochemical systems, the performance of different DFT xc-functionals to provide correct molecular structures has been considered calculating the root-mean-square deviation (RMSD) to the reference geometry (ωB97x-D3/ma-def2-QZVP).

The studied biochemical carboxylation reactions have in common that all are associated with the addition of a neutral CO\(_2\) molecule to anionic substrates to produce an anionic adduct. The anionic substrates and products involved in the biochemical carboxylation reactions catalyzed by selected enzymes are shown in Figure 6. Only model systems of the natural substrates have been considered and are highlighted through black lines in the Figure 6. Here, it is important to stress that in RuBisCO, CO\(_2\) addition in the enzymatic environment occurs on the carbon atom in the carbonyl group of Ribulose 1,5 biphosphate. In our models of this substrate the addition occurs on the carbon atom with the hydroxyl group since this is the most reactive site in vacuum.

Only biochemical carboxylation reactions associated with the biotin (BIOT) molecule and phenyl-phosphate carboxylase (PPC) enzyme will be analyzed with more detail. The other two reactions present a similar behavior to the studied DFT xc-functionals with respect to the \textit{ab-initio} reference values as the inorganic carboxylation reactions and are shown in the supporting material. For BIOT and PPC the deviation of the reaction electronic
energies with different DFT xc-functionals are shown in Figure 7 and 8, respectively. For these two reactions the same trend with respect to the GGA, the hybrid and the Meta-GGA functionals is observed, but there are certain rows that present larger deviations that are independent of the DFT xc-functional used to evaluate the reaction electronic energy. A row corresponds to calculations which share the same geometry of the assigned DFT xc-functional but varying the DFT xc-functional used to obtain the reaction electronic energy value. Therefore, the observed deviation might be related to a difference in the minimum energy structure of reactants or products. The optimized geometries from certain DFT xc-functional, therefore, can give rise to a notable effect on the obtained reaction electronic energy error. The effect of the geometry is often reduced in small molecular systems as it has been shown for the six inorganic carboxylation reactions. However, in the case of larger molecular systems, as all biochemical systems here considered, different DFT xc-functional may lead to different optimized geometry and thus to deviations in the reaction electronic
energies. For better comparison, the deviations of the biochemical carboxylation reactions are shown with the same color scale. The DFT reaction electronic energy errors associated with BIOT carboxylation reaction (second reaction, Figure 6), are shown in the Figure 7. Some DFT xc-functionals (PBE0, M06-2X, ωB97x and CAM-B3LYP) provide very accurate reaction electronic energies with errors of ~ 2 kcal mol\(^{-1}\). But, the row with the M06-L optimized geometry present larger deviations for most DFT xc-functionals and also for the best DFT xc-functionals identified before. Interestingly, for the reaction in PPC a similar

![Figure 7: Deviation in DFT reaction electronic energies calculated with different xc-functionals and the ma-def2-TZVPP basis set with respect to reference reaction electronic energy value determined at CCSD(T)/aug-cc-pVDZ (frozen core electron) level (-21.038 kcal mol\(^{-1}\)) for the biotin carboxylation reaction.](image)

effect is observed in the case of optimized geometries with the PBE, BLYP and M06-L DFT xc-functionals (see Figure 8). Also for this reaction PBE0, M06-2X, ωB97x and CAM-B3LYP DFT xc-functionals are closest to the reference values, although for this reaction the deviation is larger than in BIOT.

One way to quantify the geometric differences in the optimized structures is to calculate the root mean square deviation (RMSD) of the position of all atoms after translational and rotational alignment of the molecules have been performed. The quality of the optimized geometries for all evaluated DFT xc-functionals have been quantified by means of its RMSD
Figure 8: Deviation in DFT reaction electronic energy calculated with different xc-functionals at ma-def2-TZVPP level with respect to reference reaction electronic energy value determined at CCSD(T)/aug-cc-pVDZ level (-1.876 kcal mol\(^{-1}\)) for the ppc carboxylation reaction.

value in Figure 9. In general a very small RMSD is observed for most substrates and products of the biochemical carboxylation reactions. In the case of the substrates, the Minnesota DFT xc-functionals (M06-L and M06-2X) give rise to optimized geometries that exhibit large RMSD values for the case of CCR (M06-2X and M06-L) and the BIOT (M06-L) system. On the other hands, in the case of the products PBE, BLYP, M06-L and M06-2X DFT xc-functionals give rise to optimized geometries with large RMSD values for PPC (PBE, BLYP and M06-L) and CCR (PBE, BLYP and M06-2X) system. To separate the DFT xc-functionals an arbitrary RMSD value of 0.3 Å has been chosen and is shown as red line in both Figures. Optimized geometries that present RMSD values larger than 0.3 Å also exhibit large errors in the reaction electronic energy. In general, the obtained difference in the reaction electronic energy between any pair of DFT xc-functionals can be thought as a function of the quality of the functional to reproduce the electronic energy and the geometry. If both functionals give rise to the same optimized geometry its difference may stem from the intrinsic nature associated with both DFT xc-functionals. However, if the optimized geometry is not the same an additional term may contribute aside from its intrinsic nature, the geometric term. The large RMSD value obtained for the BIOT substrate with the M06-L
Figure 9: RMSD of substrates and products involved in biochemical carboxylation reactions with respect to the reference structure. The reference geometry has been obtained at ωB97x-D3-def2-QZVP level. The dotted red line represents a RMSD value of 0.3 Å, which is adopted to separate the geometries with the largest deviation. A schematic representation of highlighted chemical structure is shown below. The carbon atom in reference structure is shown in brown colors.

xc-functional, therefore, explains the unusual obtained reaction electronic energy values in third row of Figure 7. Likewise, the large RMSD values obtained for PPC product with PBE,
BLYP and M06-L xc-functional also explain partially the obtained reaction electronic energy values in first, second and third row of Figure 8, respectively. A schematic representation of the obtained product and reactant structures from these particular DFT xc-functional are compared with its respective reference geometry at the bottom of Figure 9. Here, it is important to mention that our obtained RMSD values depend on the selected reference geometry. Our selection criteria is based on obtained results with inorganic carboxylation reactions where $\omega$B97x-D3 has shown a very good performance.

In summary, the DFT benchmark of the biological systems identifies the $\omega$B97x-D3 DFT xc-functional as best suited to study carboxylation reactions in biological systems and special care has to be taken with specific GGA functionals as they can provide significant alterations in the minimum energy structure which is reflected in increased errors of the reaction electronic energy.

**Conclusions**

Carboxylation reactions represent an important class of chemical reactions that play a role in several biochemical and inorganic processes. Here, a systematic study of the reaction electronic energies in various carboxylation reactions have been undertaken focusing on the electronic structure of CO$_2$ and the chemical species involved using high level *ab-initio* and DFT electronic structure methods. We have found that for reactions that maintain the number of $\pi$-bonds in reactants and products, reaction electronic energies are well described by less computational demanding electronic structure method due to compensation of the electron correlation energy error on both sides of the chemical equation. For reactions where the products present no $\pi$-systems involving carbon and oxygens atoms a correct description of the electronic correlation energy in CO$_2$ has to be reached, which demands methods beyond the CCSD(T) gold standard.

The obtained reference values from post-HF methods were used to assess different DFT
exchange-correlation functional. DFT xc-functionals present the same trend in the reaction electronic energies considering the compensation of the electronic correlation energy error. GGA functionals perform worse than hybrid functionals and the ωB97x functional presents the best option when combined with the Karlsruhe def2 basis set. For biological relevant reactions the effect of the minimum energy structure from different DFT xc-functionals was also considered giving rise to a new dimension that should be used in the validation of DFT methods. The present study, therefore, lays a solid foundation to address carboxylation reaction in a more complex molecular system, as carboxylases, where post-HF method can not be used due to their high computational cost.

Acknowledgement

The authors acknowledge financial support by Comisión Nacional de Investigaciones Científicas y Tecnológicas (CONICYT-Chile). O.A.D.-G. thanks Fondo Nacional de Desarrollo Científico y Tecnológico (Fondecyt) for his postdoctoral fellowship No. 3170029 and E. V.-M. for his Fondecyt grant No. 1160197 and the financing by the Max-Planck Society through a Partner Group.

Supporting Information Available

Post-HF calculations for all small molecules involved in the different inorganic carboxylation reactions here considered. DFT calculations associated with all inorganic carboxylation reactions discussed in the main text which have been computed by means of several basis set of different nature and cardinality.

This material is available free of charge via the Internet at http://pubs.acs.org/.
References


(8) Rungtaweevoranit, B.; Baek, J.; Araujo, J. R.; Archanjo, B. S.; Choi, K. M.; Yaghi, O. M.; Somorjai, G. A. Copper Nanocrystals Encapsulated in Zr-based


(34) Hehre, W. J.; Radom, L.; v. R. Schleyer, P.; Pople, J. A. *Ab Initio Molecular Orbital Theory*. 36
Some journals require a graphical entry for the Table of Contents. This should be laid out “print ready” so that the sizing of the text is correct. Inside the `tocentry` environment, the font used is Helvetica 8 pt, as required by *Journal of the American Chemical Society*. The surrounding frame is 9 cm by 3.5 cm, which is the maximum permitted for *Journal of the American Chemical Society* graphical table of content entries. The box will not resize if the content is too big; instead it will overflow the edge of the box. This box and the associated title will always be printed on a separate page at the end of the document.