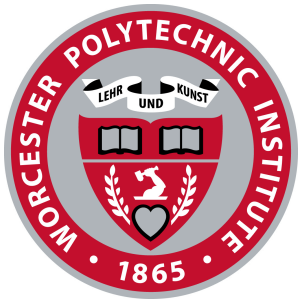
Modeling Interactions Between Carbon Dioxide and Biomolecules



A Major Qualifying Project submitted to the Faculty of

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the degree of Bachelor of Science

by Nathan Demers

Date:

25 April 2024

Report Submitted to:

Professor N. Aaron Deskins

Chemical Engineering Department

Worcester Polytechnic Institute

Abstract

With CO_2 levels on the rise in the earth, air, and water, reducing emissions to combat global warming is more important than ever. Solvent carbon capture is an effective method for removing CO_2 from industrial flue gas streams, but traditional solvents pose issues for the sustainability of the process. New solvents that are environmentally friendly and more efficient are in demand. Biomolecules, specifically amino acids, may be able to fill this demand. Atomistic modeling (density functional theory) was employed to model the interactions of nine amino acids with CO_2 in an aqueous setting. The goals of the project were to determine which density functional theory methods were most effective for modeling the binding interactions between CO_2 molecules and amino acids, and to quantify the binding energies of amino acid-CO_2 complexes. Analysis with t-tests revealed significant effects when solvation modeling was removed, or the basis set was altered. Conversely, excluding dispersion corrections showed no significant impact. Switching the exchange-correlation functional from M06-2X to B3LYP did not significantly alter the internal energies or reaction geometries, except for in the geometries of the amino acid-CO_2 zwitterions. However, changing the base which deprotonates the zwitterion in the reaction notably affected the internal energy change. DFT calculations performed at a 6-311++G** level with the M06-2X functional, solvation modeling with COSMO, and without dispersion correction are recommended as the most effective method, among those explored, for modeling the interactions of amino acids with CO_2. The binding energies for the complexes of nine amino acids with CO_2 were calculated and data supports that the binding energies of these complexes are independent of the amino acid.

Acknowledgements

I would like to express my gratitude to Professor Deskins for his invaluable guidance, support, and encouragement throughout this project. His expertise, patience, and feedback have been instrumental, I am very grateful for the time he dedicated to this project.

I would also like to extend my appreciation to WPI, specifically the Chemical Engineering Department, for providing me with the resources and opportunities necessary to pursue this project. I am thankful for the academic environment fostered at WPI and the enriching experiences and opportunities it has offered me during my undergraduate studies.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 The Role of CO_2 in Global Warming

Climate change casts an ever-darkening shadow over our planet, with carbon dioxide (CO_2) at the forefront as a potent force driving global warming. As one of the most prevalent greenhouse gasses emitted by human activities, CO_2 plays a pivotal role in trapping heat within the Earth's atmosphere, exacerbating the alarming trends of rising temperatures and environmental instability (Smith et al., 2016). The burning of fossil fuels, industrial processes, and deforestation have been primary contributors to the unprecedented increase in atmospheric CO_2 levels in recent decades, surpassing 400 parts per million (ppm) for the first time in recorded history (IPCC, 2005).

The need to address CO_2 emissions has never been more urgent, as the consequences of unchecked climate change manifest themselves in increasingly severe weather patterns, rising sea levels, and unprecedented ecological disruptions. Moreover, the ramifications of heightened CO_2 levels extend across the earth, as heightened CO_2 levels are observed in the land, the oceans, and the atmosphere (Rochelle, 2009). Elevated temperatures caused by CO_2-induced warming exacerbate the frequency and intensity of extreme weather phenomena, including heatwaves, storms, and extreme precipitation. Concurrently, seawater's thermal expansion and the melting of polar ice contribute to rising sea levels, posing imminent threats to coastal populations and ecosystems (Church & Clark, 2013). In light of these urgent challenges, concerted global action to mitigate CO_2 emissions is imperative to safeguarding our planet's future well-being and ensuring the resilience of human societies and natural ecosystems alike.

1.2 Advancing Solvent Carbon Capture

Carbon capture, a pivotal strategy in combating climate change, involves the extraction of CO_2 produced by various human activities before their release into the atmosphere. By intercepting CO_2 at its source, carbon capture technologies offer a crucial means to significantly curtail emissions and transition towards a more sustainable future (IPCC, 2022).

Solvent carbon capture stands out among the various methods for capturing carbon, proving to be an efficient approach for capturing CO_2 emissions. This method relies on chemical

solvents to absorb CO_2 from industrial flue gasses, trapping it for subsequent storage or utilization. Solvent carbon capture systems can be seamlessly integrated into existing industrial processes, making them an attractive option for reducing emissions across diverse sectors, including power generation, manufacturing, and refining (Haszeldine, 2009). Moreover, captured CO_2 can be stored underground or utilized in various applications such as enhanced oil recovery or the production of synthetic fuels, thus reducing emissions and increasing economic viability (IPCC, 2005).

Despite recent advancements in solvent carbon capture technology, challenges persist in optimizing efficiency, reducing costs, and minimizing environmental impacts. Innovations are needed to enhance solvent performance, increase capture rates, and decrease the energy requirements associated with the separation and storage of CO_2. Efforts are underway to develop novel solvents with improved selectivity, stability, and regeneration capabilities to further enhance the overall efficiency of carbon capture systems (Lu et al., 2023). Continued investment in research, development, and deployment efforts is essential to overcoming existing barriers and unlocking the full potential of carbon capture as a cornerstone of global climate change mitigation efforts.

1.3 Amino Acids in Solvent Carbon Capture

Biomolecules encompass a diverse array of organic compounds essential for various biological processes. These molecules, from carbohydrates and lipids to nucleic acids and proteins, play pivotal roles in sustaining life and driving cellular functions. Beyond their biological significance, biomolecules have applications in diverse industrial sectors owing to their unique properties and versatility. Notably, amino acids have garnered attention for their potential in solvent carbon capture due to their ability to interact with CO_2.

In solvent carbon capture, amino acids facilitate the chemical absorption of CO_2 from flue gases emitted by industrial processes. When exposed to CO_2-rich gas streams, amino acid solvents react to absorb the CO_2 molecules. This process removes CO_2 from the gas stream, enabling the purification of industrial emissions and subsequent sequestration or utilization of captured CO_2 (Lu et al., 2023). Amino acids possess distinct characteristics that make them promising options for solvent-based CO_2 capture systems. These properties include reduced

vapor pressures, heightened resilience to oxygen degradation, and diminished toxicity compared to conventional amine solvents. Their versatility arises from the presence of functional groups which can participate in chemical reactions with CO_2 molecules. Additionally, the side chains of amino acids contribute to their diverse properties, enabling tailored designs of solvent systems optimized for specific capture applications. The structural diversity of amino acids can allow for the modulation of solvent properties, such as selectivity, capacity, and regeneration efficiency, to optimize the performance of solvent carbon capture systems. Furthermore, the tunability of amino acid-based solvents enables the design of tailored capture solutions for specific industrial applications and operating conditions (Guo et al., 2013).

1.4 Exploring Molecular Interactions with Density Functional Theory

Density Functional Theory (DFT) is a computational chemistry method used for studying molecular interactions and properties. DFT provides a theoretical framework for accurately predicting the electronic structure and energies of atoms and molecules. Its widespread adoption in research and industry stems from its ability to balance computational efficiency and accuracy, making it applicable to various chemical systems and phenomena (Jones, 2015).

To accurately simulate CO_2 binding in solvents, it is essential to factor in dispersion forces and solvation effects when modeling these molecular interactions. Dispersion interactions, also known as Van der Waals forces, emerge from the temporary fluctuations in electron density within molecules and play a critical role in the interaction between CO_2 and solvent molecules. While often subtle, these forces contribute significantly to the overall binding energy and must be accurately accounted for in computational models (Tkatchenko & Scheffler, 2009). Additionally, solvation effects, which are the interactions between solute molecules like CO_2 and surrounding solvent molecules, influence the behavior and stability of CO_2 in solution. Solvent molecules can modulate the energetics of CO_2 binding, alter its structural conformation, and impact its transport properties in the environment. Therefore, incorporating solvation effects into computational models is necessary for obtaining realistic insights into the behavior of CO_2 in solution (Jalan et al., 2010).

DFT-based simulations enable the exploration of various factors influencing CO_2 solvation, such as solvent effects and Van der Waals forces, thereby providing valuable insights

into CO_2 capture, transport, and storage processes. DFT facilitates an understanding of CO_2 behavior in aqueous solutions, thereby catalyzing innovations in CO_2 capture and storage technologies. DFT is a pivotal tool for exploring the behavior of CO_2 molecules in aqueous solutions. By integrating dispersion and solvation effects into DFT calculations, accurate predictions can be made to describe the thermodynamics and kinetics of CO_2 interactions with water molecules and other solvents.

1.5 Project Goals

This project had two main goals. First, it aimed to determine which DFT methods were most effective for modeling the binding interactions between CO_2 molecules and amino acids. Second, it aimed to quantify the binding energies of amino acid-CO_2 complexes to provide insight into the strength and nature of these interactions, which are pivotal for designing effective CO_2 capture systems.

This research plays a significant role in addressing CO_2 emissions and combating global warming. By exploring the capabilities of DFT and investigating the properties of amino acids, the project aimed to contribute to advancing carbon capture technologies. Through these efforts, this study sought to offer insights that could lead to the development of more sustainable and efficient methods for capturing CO_2, thereby helping to mitigate climate change.

2. Literature Review

2.1 Overview

The aim of this review is to identify and analyze existing molecular modeling techniques employed to study similar complexes, evaluating their advantages and limitations. Additionally, it aims to evaluate the rationale behind selecting Density Functional Theory (DFT) over alternative methods like MP2 and DLPNO-CCSD(T), exploring the strengths and weaknesses of each approach. The review endeavors to compile a list of amino acids studied in the context of solvent-based carbon capture, providing insights into their individual characteristics and effectiveness in CO_2 capture. A subset of amino acids will be selected for in-depth investigation based on their structural and chemical properties, and potential for CO_2 capture. A brief overview of the computational models used in this study will also be provided.

2.2 A Comparison of Molecular Modeling Methods

Previous studies investigating similar reactions have utilized various molecular modeling methods to simulate CO_2 binding to amino acids. Among these methods are MP2 (Møller-Plesset perturbation theory), and DLPNO-CCSD(T) (Domain-based Local Pair Natural Orbital Coupled Cluster with Single, Double, and Perturbative Triple excitations). Additionally, Density Functional Theory (DFT) has been widely used due to its computational efficiency and versatility in describing molecular interactions.

MP2 is one of the more simple and accurate expansions of Hartree-Fock. It provides accurate electronic structure calculations by including electron correlation effects beyond the Hartree-Fock approximation. It has been utilized in studies investigating CO_2 binding to amino acids due to its ability to capture dispersion interactions, which are crucial for describing weak interactions such as those involved in CO_2 solvation (Cremer, 2011). However, MP2 calculations can be computationally demanding, particularly for large systems, limiting its applicability to relatively small molecular complexes.

DLPNO-CCSD(T) is a highly accurate method that accounts for electron correlation effects through coupled cluster theory. It offers a compromise between accuracy and computational cost by employing truncation schemes, such as the domain-based approach, to

reduce the computational expense associated with conventional CCSD(T) calculations. DLPNO-CCSD(T) has been used in studies of CO_2 interactions with biomolecules due to its high accuracy in describing non-covalent interactions. However, DLPNO-CCSD(T) can be complex, and its computational cost remains significant, especially for large systems (Sandler et al., 2021).

DFT, on the other hand, offers a computationally efficient approach to modeling molecular interactions by approximating the electron density of a system. It has been widely applied in studies of CO_2 binding to amino acids due to its balance of accuracy and computational feasibility. DFT methods vary in exchange-correlation (XC) functionals, each offering advantages and limitations in describing different types of interactions. The consideration of methods like MP2 and DLPNO-CCSD(T) for this research is motivated by the need for accurate descriptions of CO_2 binding to amino acids. While DFT provides a computationally feasible approach, MP2 and DLPNO-CCSD(T) offer higher accuracy in capturing dispersion and non-covalent interactions, which are crucial for understanding the binding energetics of amino acid-CO_2 complexes.

DFT stands out due to its balance between accuracy and computational cost. While MP2 and DLPNO-CCSD(T) offer higher accuracy in describing molecular interactions, they have significantly higher computational demands. DFT provides a more feasible approach and moreover, it can be configured to incorporate solvation and dispersion. While MP2 and DLPNO-CCSD(T) can be extended to include solvation models, the computational cost increases substantially. This limitation hinders the practicality of using these methods for simulations involving the solvent environments relevant to CO_2 capture processes (Tomasi et al., 2005). DFT offers various solvation models that can be readily integrated into simulations, allowing for more accurate descriptions of CO_2 binding in solution. The time efficiency of DFT compared to MP2 and DLPNO-CCSD(T) further justifies its selection as a computational method for this study. While MP2 and DLPNO-CCSD(T) calculations can be prohibitively expensive, DFT calculations scale more favorably with system size, allowing for the examination of larger molecules or groups within a feasible computational budget (Matta, 2010).

Exchange-correlation functionals in DFT approximate the exchange and correlation effects between electrons in a many-electron system. These functionals provide a framework to describe these complex electron-electron interactions within the system accurately. In this study, the XC functionals M06-2X and B3LYP will be used. The selection of the M06-2X and B3LYP

functionals for modeling the absorption of CO_2 in an aqueous solution is rationalized by their complementary strengths in capturing different aspects of the system. M06-2X is renowned for its accuracy in describing noncovalent interactions and dispersion forces, which are crucial for describing the solvation behavior of CO_2 in water. Its incorporation of a significant fraction of Hartree-Fock exchange allows for a balanced treatment of dynamic and static correlation effects, enhancing accuracy in predicting energetics and geometries (Zhao & Truhlar, 2008). On the other hand, B3LYP, another widely used hybrid functional, excels in describing the electronic structure and transition properties.

2.3 A Review of Amino Acids for Solvent Carbon Capture

Numerous in-vitro and in-silico studies have been conducted to analyze the efficacy of various amino acids for carbon-capture applications. Park et al. tested twelve different amino acids as rate promoters for CO_2 absorption in a potassium carbonate solution. Six primary and six secondary amino acid salts were used; the secondary amino acids generally increased the absorption rates of CO_2 by the K_2CO_3 solution much more. Sarcosine and pipecolic acid were found to be the most effective rate promoters (Park et al., 2014). Hu et al. also studied the efficacies of amino acid salts as rate promoters for CO_2 absorption in a K_2CO_3 solution. Similarly to Park et al, they found the effect of the amino acid salts to be sensitive to the pH of the solution. Sarcosine and proline were found to be the most effective rate promoters. Both of these amino acids were shown to be more effective absorption rate boosters at high pHs than monoethanolamine (MEA), a commercially used solvent for CO_2 capture. This is due to the higher reaction order between CO_2 and sarcosine/proline than between CO_2 and MEA (Song et al., 2012).

Holst et al. studied CO_2 absorption in an aqueous solution with amino acid salts. They found that the amino acid's pK_a affects the CO_2 absorption rate, with a lower pK_a corresponding to a greater absorption rate. The potassium salts of sarcosine and proline were again found to be the most effective at CO_2 absorption, the authors attribute this to higher rate constants and low pK_as for these two amino acids. A difference between the lithium and potassium salts of proline was noted, with the potassium salts showing an experimentally higher rate constant. Interestingly, the absorption rate of CO_2 seemed to be independent of the counter-ion for

sarcosine salts. This discussion of counter-ions is also brought up in a review conducted by Hu et al. with their conclusion being that potassium salts are generally the most effective. Hu et al. reviewed recent experimental data using amino acid salts for CO_2 absorption. Their review provides estimated reaction orders and corrected reaction constants of various amino acids at 298K. The greatest amount of data was collected for glycine, sarcosine, and proline, but experimental data for more than a dozen other amino acids was also analyzed. They conclude that lysine, proline, and sarcosine have been reported to have the largest reaction constants with CO_2.

Based on this research, sarcosine and proline have emerged as the most promising amino acids for study. Their efficacies are supported by experimental data, and their properties, namely low pK_as and high rate constants with CO_2, make them ideal candidates for further study. Moreover, they have a relatively low number of atoms compared to other amino acids. Molecular size is important for this study as the complexity of DFT calculations increases greatly for larger molecules. In addition to sarcosine and proline, alanine, glycine, valine, leucine, serine, phenylalanine, and threonine will be included in the calculations to provide a larger sample set.

2.4 Modeling the Reaction

To accurately model the reaction of an amino acid with CO_2, knowledge of the appropriate reaction mechanism is necessary. Three potential mechanisms can be employed to model the reactions of amines; these are the zwitterion, termolecular, and base-catalyzed hydration mechanisms (Vaidya & Kenig, 2007). Of these, further investigation is warranted into the zwitterion and termolecular mechanisms. The base-catalyzed hydration mechanism describes the reaction of tertiary amines, which does not pertain to this project. Additionally, Yu et al. suggest that a zwitterion-type mechanism may also describe the base catalysis reaction.

Firstly, the zwitterion mechanism is a two-step mechanism. It suggests that the reaction between the amino acid and CO_2 proceeds by forming a zwitterion intermediate.

$$RNH_2 + CO_2 \Leftrightarrow RNH_2^+COO^- \quad (1)$$

A base then deprotonates the zwitterion intermediate to form the product.

$$RNH_2^+COO^- + B^- \Rightarrow RNHCOO^- + BH \quad (2)$$

Vaidya et al. evaluated the kinetics and overall rate of this reaction. They found that the reaction order would be fractional between 1 and 2, dependent on the concentration of salt in the solution. This evaluation agrees with the data provided in the 2018 review by Hu et al. They found the experimental reaction order of eight different amino acids with CO_2 to fall between 1 and 1.81, dependent on the reaction apparatus, salt concentration, and temperature.

The termolecular mechanism is a one-step mechanism that assumes an amino acid reacts with a CO_2 molecule and a base molecule simultaneously through an encounter complex. Most occurrences of this encounter complex dissociate again to form the reactants, but a small amount react with an additional amino acid salt or a water molecule to form the products.

$$RNH_2 + CO_2 + B \Leftrightarrow \{Encounter\ Complex\} \Rightarrow RNHCOO^- + BH^+ \quad (3)$$

Vaidya et al.'s kinetics evaluation of this mechanism suggests that this reaction is first-order with respect to the amino acid salt when water is the dominant base in the solution and second-order when RNH_2 is the dominant base. This is another possible explanation for the fractional reaction orders observed by Hu et al.

In their study of the kinetics of the reaction of aqueous salts of taurine and glycine, Kumar et al. concluded that either the zwitterion or termolecular mechanism could be used to describe the experimental kinetic data. Additionally, they found that more data would be required for the reaction of various aqueous amino acid salts with CO_2 to come to a more certain conclusion on which mechanism best describes the reaction. For the purposes of this study, the zwitterion mechanism will be used to describe the reaction. The primary reason for this is that it allows for the intermediate step of the reaction to be modeled more easily and accurately. In the termolecular mechanism, the exact geometry of the encounter complex intermediate is unknown and may be difficult to elucidate, given that the complex is composed of three molecules. The geometry of the zwitterion intermediate is known and therefore makes it a more compelling choice of mechanism for this study.

2.5. A Brief Overview of DFT, Dispersion, and Solvation Models

DFT offers a practical approach to studying the electronic structure of matter. Proposed by Walter Kohn and Pierre Hohenberg in 1964, DFT simplifies the many-body problem of interacting electrons by focusing on the electron density rather than the wave function. This reduction enables the calculation of various properties of atoms, molecules, and solids. Central to DFT are the Kohn-Sham equations, introduced by Walter Kohn and Lu Sham in 1965. These equations transform the many-electron problem into a set of single-electron equations, where each electron moves in an effective potential determined by the electron density and the exchange-correlation functional (Jones & Gunnarsson, 1989).

DFT calculations rely on accurately approximating the XC functional, which captures the effects of electron exchange and correlation. While various approximations exist, none are exact, and developing more accurate functionals remains an active area of research. Despite its approximations, DFT is an indispensable tool in many subjects, providing insights into the properties of diverse systems. Due to its efficiency and versatility, it is used widely in both academic and industrial research.

In DFT, standard functionals often fail to accurately account for Van der Waals forces, which are crucial molecular interactions. To account for this, dispersion correction (DFT-D3) is included in calculations. DFT-D3 addresses this limitation by introducing an empirical correction to the total energy of a system, specifically targeting dispersion interactions. This correction term is added to the standard DFT energy and is based on the pairwise summation of atom-atom dispersion contributions. The DFT-D3 method incorporates parameters derived from reference data sets; it adjusts the dispersion interaction energy based on the distance between atoms and their chemical environment. By including these corrections, DFT-D3 improves a description of Van der Waals forces in DFT calculations, leading to more accurate predictions of molecular structures, energetics, and properties (Grimme et al., 2010).

Another factor that standard DFT calculations do not account for is solvation. COSMO (Conductor-Like Screening Model) is a continuum solvation model designed to simulate the behavior of solute molecules in a solvent environment. In the COSMO model, the solvent is treated as a continuous dielectric medium with a permittivity similar to water's. The solute molecules are represented as charged surfaces surrounded by a layer of discrete point charges,

which mimic the polarizability and charge distribution of the real molecules. These point charges interact with the solvent molecules, and the resulting electrostatic interactions are calculated using continuum electrostatic theory, incorporating the solvent dielectric constant and the surface charges of the solute (Klamt, 1995). COSMO has been successfully applied in various areas of computational chemistry, including studies of solvation effects on reaction kinetics, thermodynamics, and spectroscopy.

3. Methodology

3.1 Modeling a Reaction with DFT

The calculations for this study were conducted using WebMO for a graphical interface, and the Northwest Computational Chemistry Package (NWChem) 7.0.2 for execution of the calculations. WebMO is a web-based interface that facilitates the setup, visualization, and analysis of computational chemistry calculations. It provides a platform that enables users to construct molecular models, define computational parameters, and submit jobs to various computational chemistry software packages, including NWChem. NWChem is a high-performance computational chemistry software suite that offers a wide range of quantum chemistry and molecular dynamics capabilities. Solvation modeling was performed with COSMO (Klamt, 1995), and dispersion correction calculations were performed with DFT-D3 (Grimme et al., 2010). The exchange-correlation functions M06-2X (Zhao & Truhlar, 2008), and B3LYP (Becke, 1988) (Lee et al., 1988) were used.

When modeling a reaction using DFT, the first step involves optimizing the molecular structures of the reactants, intermediates, and products individually. This optimization process aims to find the most energetically favorable configurations by adjusting the positions of atoms within the molecules until the forces on each atom approach zero. Through this process, DFT calculates the total energy of each molecule based on the positions of its constituent atoms; it employs the exchange-correlation functional to provide a correction to the electron-electron interactions within the system. Once the optimized structures of the individual molecules are obtained, the next step is to compute the overall reaction energetics, represented by the change in internal energy (ΔE). ΔE is calculated as the difference in the total energies of the products and the reactants, as shown below.

$$\Delta E = \sum E_{DFT}(Products) - \sum E_{DFT}(Reactants)$$

The ΔE values for Reactions 1 and 2 (re-stated below) will be calculated for nine amino acids.

$$RNH_2 + CO_2 \Leftrightarrow RNH_2^+COO^- \quad (1)$$

$$RNH_2^+COO^- + B^- \Rightarrow RNHCOO^- + BH \quad (2)$$

3.2 Optimizing Molecular Structures using DFT in WebMO

1. To begin, create a new job and, using the build tool, draw a skeleton structure of the molecule you wish to model. Include charges on any atoms that require them.

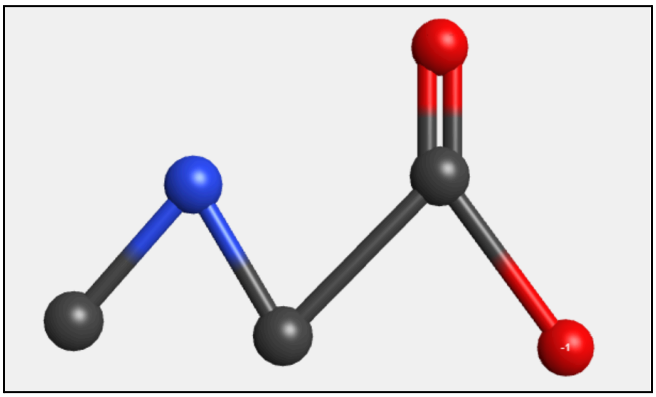


Figure 1: A skeleton structure of a sarcosine molecule created in WebMO. Grey spheres represent carbon, red spheres represent oxygen, and blue spheres represent nitrogen.

2. Use the Cleanup → Comprehensive Mechanics tool to add hydrogens and perform rough corrections for bond angles and lengths.

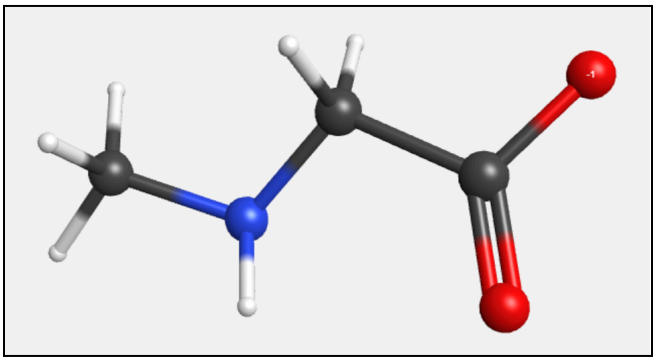


Figure 2: A corrected geometry for sarcosine generated with WebMO. This image uses a similar color scheme to Figure 1. White spheres represent hydrogen.

3. Before proceeding, ensure that the molecule's geometry does not present any issues that may prevent calculations from proceeding. Examples of this may be bond distances that are too short or long, unreasonable bond angles, or atoms that are too close together.
4. Proceed using the ‘Continue’ arrow in the lower right corner of the interface.
5. Select the correct computational engine and queue. Proceed using the ‘Continue’ arrow.
6. Name the job and select ‘Geometry Optimization’ from the calculation drop-down. Select ‘DFT’ from the theory drop-down and ensure the molecule's charge is correct.

7. Proceed to the ‘Advanced’ tab and provide the desired calculation parameters.
8. The input script must be modified to include dispersion and solvation. To do this, proceed to the ‘Preview’ tab and click ‘Generate’ to generate an input script that can be edited.
9. Within the DFT input block, include the keywords ‘disp vdw 3’ to include DFT-D3 dispersion corrections.
10. Next, add a DRIVER input block to increase the maximum number of geometry steps for the calculation. An example DRIVER block is shown here.

```
driver
  maxiter #
end
```

11. Add a COSMO input block to include solvation modeling with COSMO. The COSMO input block requires the dielectric constant of the solvent and the Van der Waals radii of each atom in the molecule. Van der Waals radii must be listed in the same order as their corresponding atoms are listed in the geometry input. Below is an example COSMO input block for sarcosine.

```
cosmo
  dielec 78.0
  radius 1.7
  1.55
  1.7
  1.7
  1.52
  1.52
  1.2
  1.2
  1.2
  1.2
  1.2
  1.2
  lineq  0
end
```

12. Specify the amount of memory allocated for the calculation. Include the following input line and change the memory amounts as desired. The sum of stack, heap, and global memories must equal the total.

    memory total 1024 mb stack 256 mb heap 256 mb global 512 mb

13. Submit the calculation using the 'Continue' arrow. The Appendix contains example input scripts for H_2O, OH, CO_2, sarcosine, the sarcosine-CO_2 zwitterion, and the sarcosine-CO_2 product.

3.3 Variables

Several variables must be considered when modeling the reactions of amino acids with CO2. The variables that will be controlled for this experiment are amino acid type, basis set, functional, inclusion of dispersion corrections, type of solvation, and the base in Reaction 2. The values of each variable are shown in Table 1 below.

Table 1: The values of controlled variables used for modeling the reactions of amino acids and CO_2.

Amino Acid	**Basis Set**	**Functional**	**Dispersion**	**Solvation**	**Base (Reaction 2)**
Sarcosine Glycine Alanine Proline Valine Leucine Serine Phenylalanine Threonine	Routine (6-31G*) Accurate (6-311++G**)	M06-2X B3LYP	Dispersion No Dispersion	COSMO No COSMO	OH^- $RNHCOO^-$

The base case set of variables used for calculations will be the following: Routine, M06-2X, Dispersion, COSMO, and OH^-. Reactions 1 and 2 will be modeled with this set of variables for a

base calculation set. A 6-31G* basis set was chosen for the base case due to its compromise between efficiency and accuracy. M06-2X was chosen as it is able to describe Van der Waals interactions and medium-range electron correlations much better than B3LYP (Zhao & Truhlar, 2008). Dispersion and solvation corrections were included in the base case to increase the accuracy of the calculations. OH^- was chosen as a base due to its relative abundance in solution compared to $RNHCOO^-$. The total charge of these molecules (OH^- and $RNHCOO^-$) was -1 when modeling them.

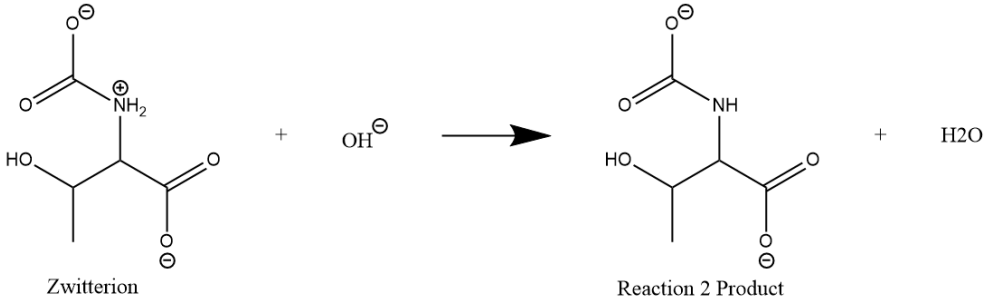
Table 2: Different combinations of modeling variables (different calculation sets) used to model Reactions 1 and 2. The variables for each modeling set are specified.

Name of Calculation Set	**Basis Set**	**Functional**	**Dispersion**	**Solvation**	**Base (Reaction 2)**
Base	Routine	M06-2X	Dispersion	COSMO	OH^-
Accurate	Accurate	M06-2X	Dispersion	COSMO	OH^-
B3LYP	Routine	B3LYP	Dispersion	COSMO	OH^-
No Disp	Routine	M06-2X	No Dispersion	COSMO	OH^-
No COSMO	Routine	M06-2X	Dispersion	No COSMO	OH^-
Change of Base	Routine	M06-2X	Dispersion	COSMO	$RNHCOO^-$

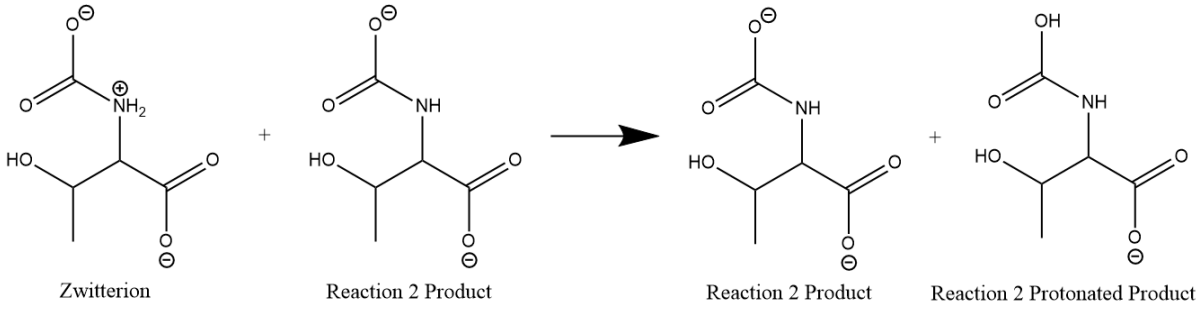
The effects of each variable are isolated by only changing one variable from its base value in each calculation set. The number of calculations required for the base calculation set is 30. The Accurate, B3LYP, No Disp, and No COSMO calculation sets each require an additional 30 calculations. Only 9 calculations are needed to model the change of base. Additionally, vibrational frequency analysis will be conducted on the base geometries for the reactions of Sarcosine, Glycine, Alanine, and Proline to perform a thermodynamic analysis. The thermodynamic analysis will be used to support the presence of Reactions 1 and 2. In total, 174 calculations will be performed.

3.4 'Change of Base' Reaction

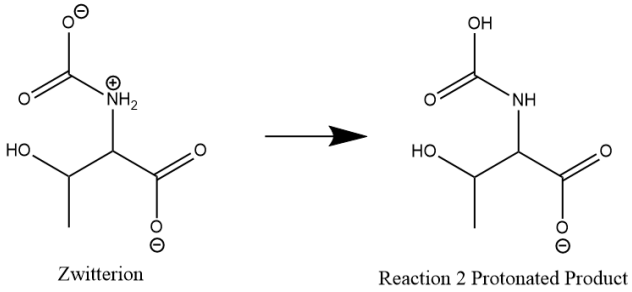
For Reaction 2, the change of base from OH^- to $RNHCOO^-$ is considered. The base case for Reaction 2 of threonine, an example, is shown below.



When instead the Reaction 2 product, $RNHCOO^-$, acts as the base in the reaction, the reaction changes to the following:



However, the Reaction 2 product is present on both sides of this reaction, and it simplifies to:



This simplified reaction, shown for threonine in Figure 10, is applicable to study the change in internal energy of the system. However, it does not accurately represent the reaction kinetics and therefore cannot be applied to the reaction mechanism or to study transition states.

3.5 Statistical Analysis

A t-test comparison was chosen to analyze and compare sample sets comprising data from nine amino acids. The t-test is particularly well-suited for small sample sizes, making it appropriate for scenarios where the number of observations is limited, as in this case. By calculating the t-value, which measures the difference between the means of the two sample sets relative to the variability within the samples, the t-test determines whether the observed differences are statistically significant or simply due to random variation. This statistical test provides a reliable means to infer whether the observed distinctions between the sample sets are likely to reflect genuine differences in the population means or are merely artifacts of sampling variability. Therefore, employing a t-test facilitates robust and reliable comparisons between small sample sets, enabling meaningful interpretations of the data. A t-value below 0.05 indicates that the observed difference between the means of two sample sets is statistically significant at the 95% confidence level. Therefore, a t-value below 0.05 suggests that the observed difference is unlikely to be a result of chance variation and is more likely to reflect a genuine distinction between the populations being compared. Excel was used to perform the t-tests for this study.

4. Results

4.1 Modeling Reaction 1

Calculations were performed to model Reaction 1 for nine amino acids with five different calculation sets. Figure 3 depicts the final geometries for valine and the valine-CO_2 zwitterion, a representative reactant and product. Calculated ΔE values for Reaction 1 are organized in Table 3, and a visual representation is provided in Figure 4.

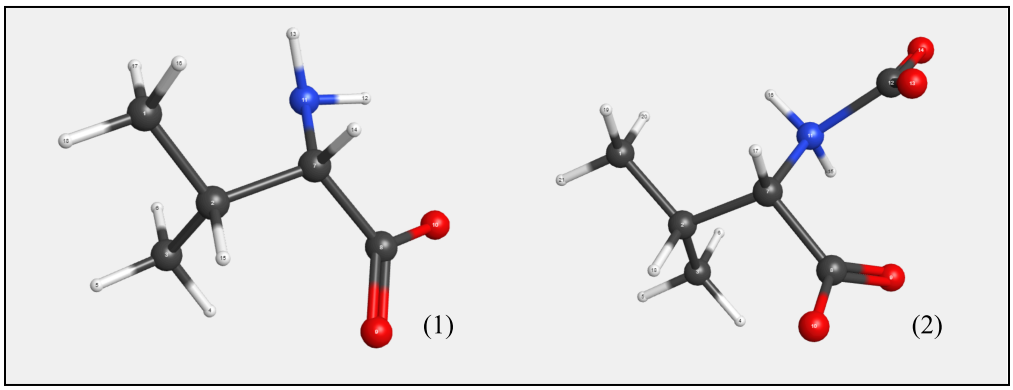


Figure 3: Final geometries for (1) Valine and the (2) Valine-CO_2 zwitterion. An example of a reactant and product.

Table 3: Calculated ΔE values for Reaction 1 by calculation set and amino acid.

		Reaction 1 ΔE (kcal/mol)				
Amino Acid	*Sarcosine*	-14.57	-19.94	-12.28	-13.44	-12.03
	Glycine	-14.51	-17.47	-10.47	-14.41	-12.01
	Alanine	-14.89	-18.19	-12.15	-14.74	-11.78
	Proline	-13.91	-19.90	-11.33	-13.37	-11.25
	Valine	-13.98	-17.28	-12.28	-13.76	-12.52
	Leucine	-13.88	-18.77	-13.35	-13.63	-13.38
	Serine	-11.42	-13.98	-8.21	-10.02	-10.29
	Phenylalanine	-13.09	-15.31	-9.77	-12.83	-11.52
	Threonine	-7.88	-10.08	-7.66	-8.08	-7.16
	*Average Value**	-13.13*	-16.77*	-10.83*	-12.70*	-11.33*
Calculation Set		*Base*	*No COSMO*	*Accurate*	*No Disp*	*B3LYP*

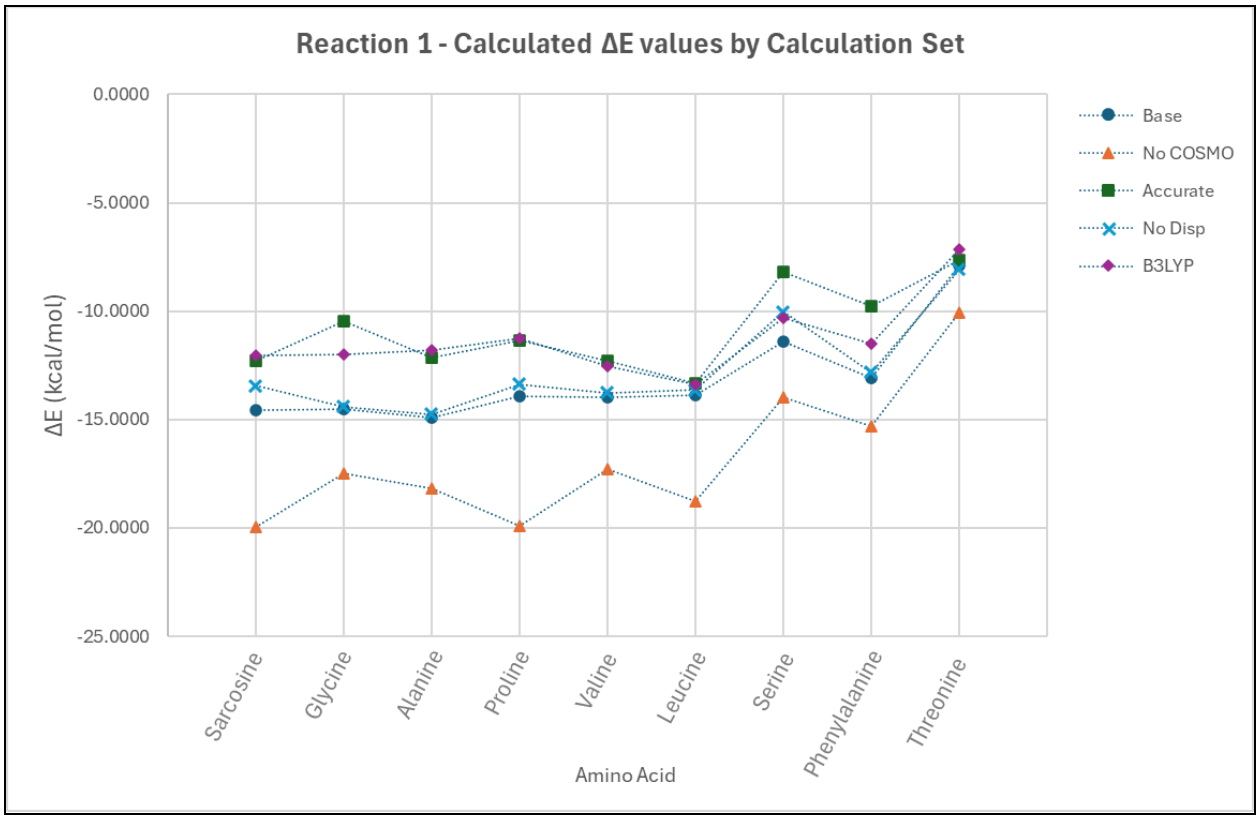


Figure 4: Calculated ΔE values for Reaction 1 for nine amino acids and five calculation sets.

An analysis using t-tests was performed to determine if there was a significant difference between the groups of ΔE values found using each of the different calculation sets. Table 4 shows the t-values for comparing the base calculation set to the other four calculation sets.

Table 4: t-values for comparing groups of Reaction 1 ΔE values. The t-values shown compare ΔE values from the base calculation set to the other individual calculation sets.

Reaction 1: t-values			
Base/No COSMO	*Base/Accurate*	*Base/No Disp*	*Base/B3LYP*
0.014	0.034	0.686	0.077

4.2 Modeling Reaction 2

Reaction 2 was modeled for nine amino acids with six calculation sets. Figure 5 depicts an example group of products and reactants for the reaction of phenylalanine. Calculated ΔE values for Reaction 2 are organized in Table 5, Figure 6 presents and a visual representation.

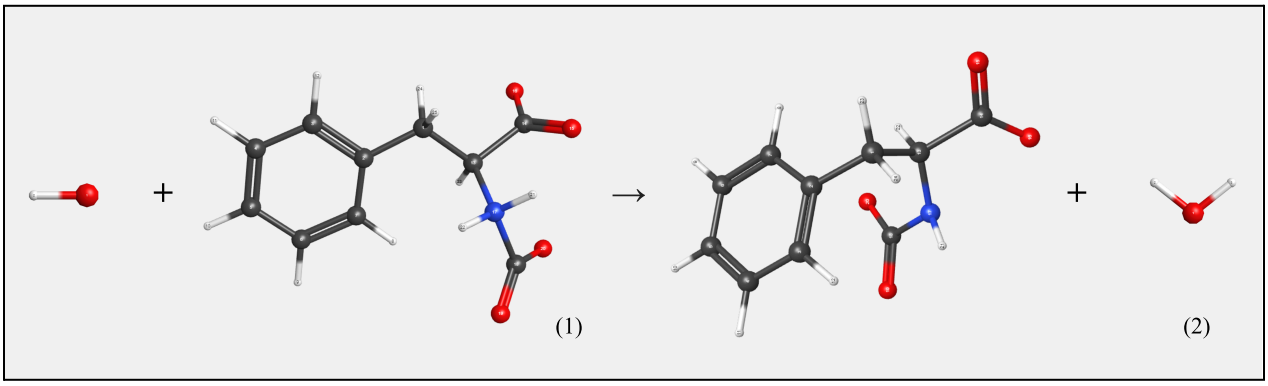


Figure 5: Final phenylalanine geometries for the Reaction 2 (1) Reactants and (2) Products.

Table 5: Calculated ΔE values for Reaction 2 by calculation set and amino acid.

		Reaction 2 ΔE (kcal/mol)					
Amino Acid	*Sarcosine*	-36.43	11.78	-20.58	-37.46	-34.42	-1.36
	Glycine	-33.90	9.38	-19.88	-33.88	-32.04	0.36
	Alanine	-34.71	8.91	-19.66	-34.69	-33.74	0.55
	Proline	-37.57	13.48	-22.68	-37.92	-35.35	-2.63
	Valine	-35.21	4.53	-20.17	-35.21	-32.74	0.29
	Leucine	-33.27	20.51	-16.01	-33.89	-30.20	2.09
	Serine	-40.55	-7.00	-23.69	-40.54	-38.19	-1.88
	Phenylalanine	-37.71	-0.88	-22.48	-37.67	-34.64	0.73
	Threonine	-41.08	-4.75	-21.37	-40.67	-37.84	-3.65
	*Average Value**	-36.71*	6.22*	-20.72*	-36.88*	-34.35*	-0.61*
Calculation Set		*Base*	*No COSMO*	*Accurate*	*No Disp*	*B3LYP*	*Change of Base*

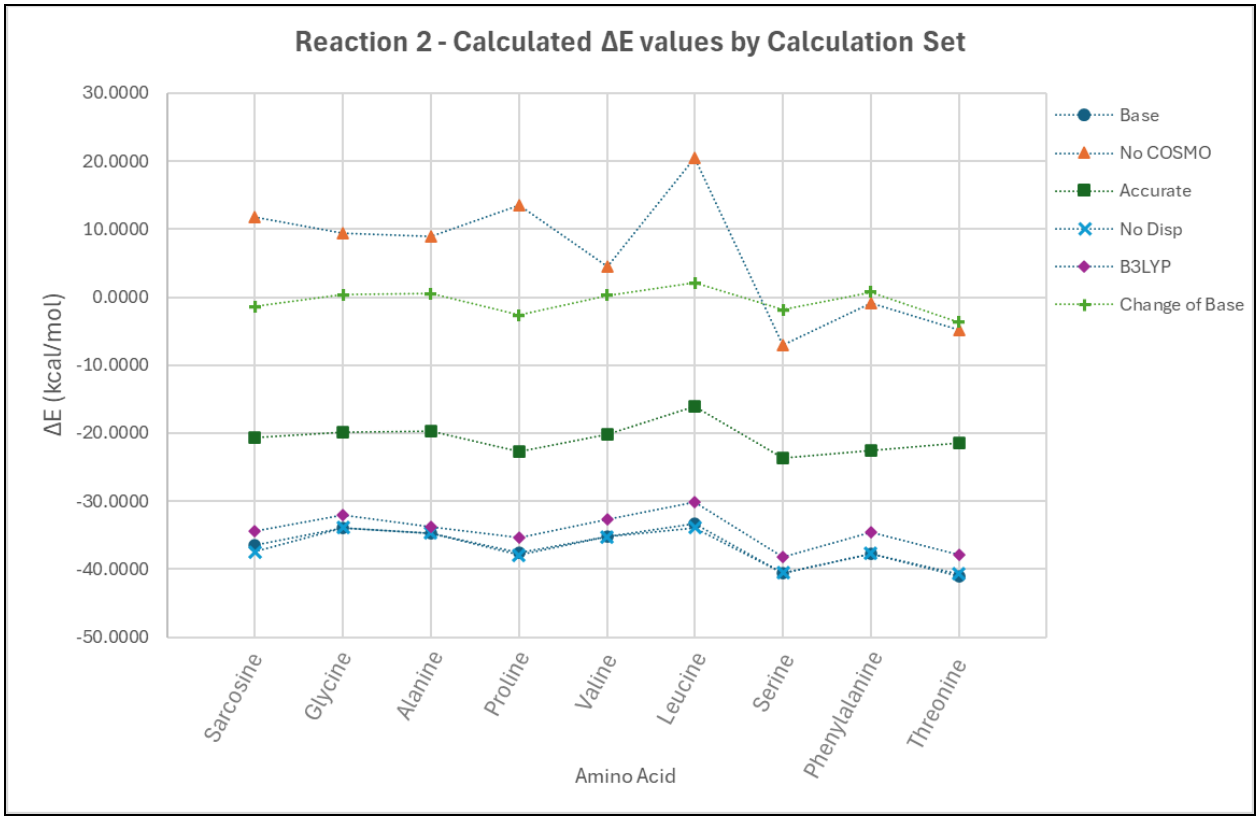


Figure 6: Calculated ΔE values for Reaction 2 for nine amino acids and six calculation sets.

Again, t-tests were used to determine if there was a significant difference between the groups of ΔE values. The base calculation set was compared to the other five calculation sets, t-values are shown in Table 6.

Table 6: t-values for comparing groups of Reaction 2 ΔE values. The t-values shown compare ΔE values from the base calculation set to the other individual calculation sets.

Reaction 2: t-values				
Base/No COSMO	*Base/Accurate*	*Base/No Disp*	*Base/B3LYP*	*Base/Change of Base*
1.51E-07	6.81E-10	0.899	0.080	1.46E-14

4.3 Other Computational Results

Vibrational frequency calculations were performed for the reactions between four amino acids and CO_2. Calculations were performed at 298K with the base case method. Table 7 includes the calculated standard enthalpy, entropy, and Gibbs free energy values for sarcosine, glycine, alanine, and proline in Reactions 1 and 2.

Table 7: Calculated ΔH^o, ΔS^o, and ΔG^o values of Reactions 1 and 2 for four amino acids.

	Reaction 1			**Reaction 2**		
Amino Acid	ΔH^o (kcal/mol)	ΔS^o (cal/mol-K)	ΔG^o (kcal/mol)	ΔH^o (kcal/mol)	ΔS^o (cal/mol-K)	ΔG^o (kcal/mol)
Sarcosine	2.82	-30.85	12.02	-1.34	-1.92	-0.77
Glycine	3.14	-36.15	13.92	-0.03	8.91	-2.69
Alanine	2.23	-33.68	12.27	-0.26	-1.92	0.31
Proline	1.18	-42.76	13.92	-0.38	4.24	-1.65

An analysis of the C9-N2 bond length was performed. The C9-N2 bond, shown below in Figure 7, is generalized as the bond between the carbon of the CO_2 molecule and the participating nitrogen of the amino acid. Bond lengths were analyzed for the base case and four additional calculation sets.

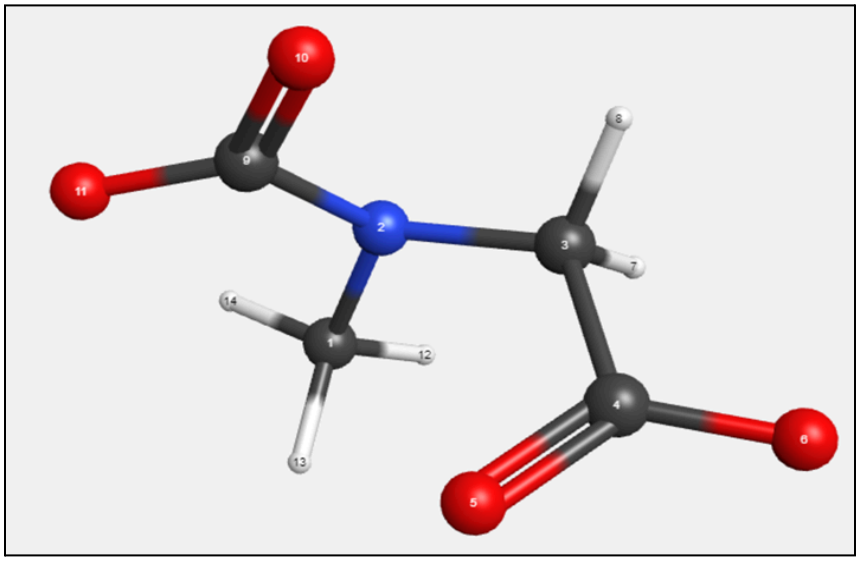


Figure 7: A numbered geometry of the sarcosine-CO_2 Reaction 2 product.

Table 8: C9-N2 bond length by calculation set, amino acid, and molecule.

Calculation Set	Molecule	C9-N2 Bond Length (Å)								
Base	*Zwitterion*	1.582	1.566	1.570	1.564	1.573	1.562	1.559	1.555	1.574
	Product	1.411	1.405	1.399	1.398	1.400	1.402	1.394	1.413	1.405
No COSMO	*Zwitterion*	1.658	1.648	1.647	1.646	1.641	1.627	1.616	1.656	1.692
	Product	1.450	1.434	1.427	1.461	1.443	1.446	1.418	1.442	1.446
No Disp	*Zwitterion*	1.579	1.565	1.569	1.563	1.573	1.562	1.558	1.555	1.574
	Product	1.408	1.405	1.399	1.398	1.400	1.402	1.394	1.413	1.405
Accurate	*Zwitterion*	1.555	1.556	1.557	1.552	1.561	1.546	1.555	1.546	1.566
	Product	1.393	1.376	1.385	1.373	1.393	1.382	1.384	1.392	1.384
B3LYP	*Zwitterion*	1.615	1.590	1.579	1.588	1.596	1.588	1.582	1.572	1.603
	Product	1.417	1.411	1.404	1.391	1.405	1.410	1.398	1.413	1.411
Amino Acid		*Sarcosine*	*Glycine*	*Alanine*	*Proline*	*Valine*	*Leucine*	*Serine*	*Phenylalanine*	*Threonine*

To compare the base case to the four other calculation sets, t-tests were used. The bond lengths in the zwitterion and product molecules are compared separately.

Table 9: t-values for the comparison of C9-N2 bond lengths. The t-values compare bond lengths from the base calculation set to the other individual calculation sets.

C9-N2 Bond Length: t-values							
Base/No COSMO		***Base/No Disp***		***Base/Accurate***		***Base/B3LYP***	
Zwitterion	*Product*	*Zwitterion*	*Product*	*Zwitterion*	*Product*	*Zwitterion*	*Product*
6.63E-07	4.99E-06	0.842	0.907	0.003	2.89E-05	5.37E-04	0.297

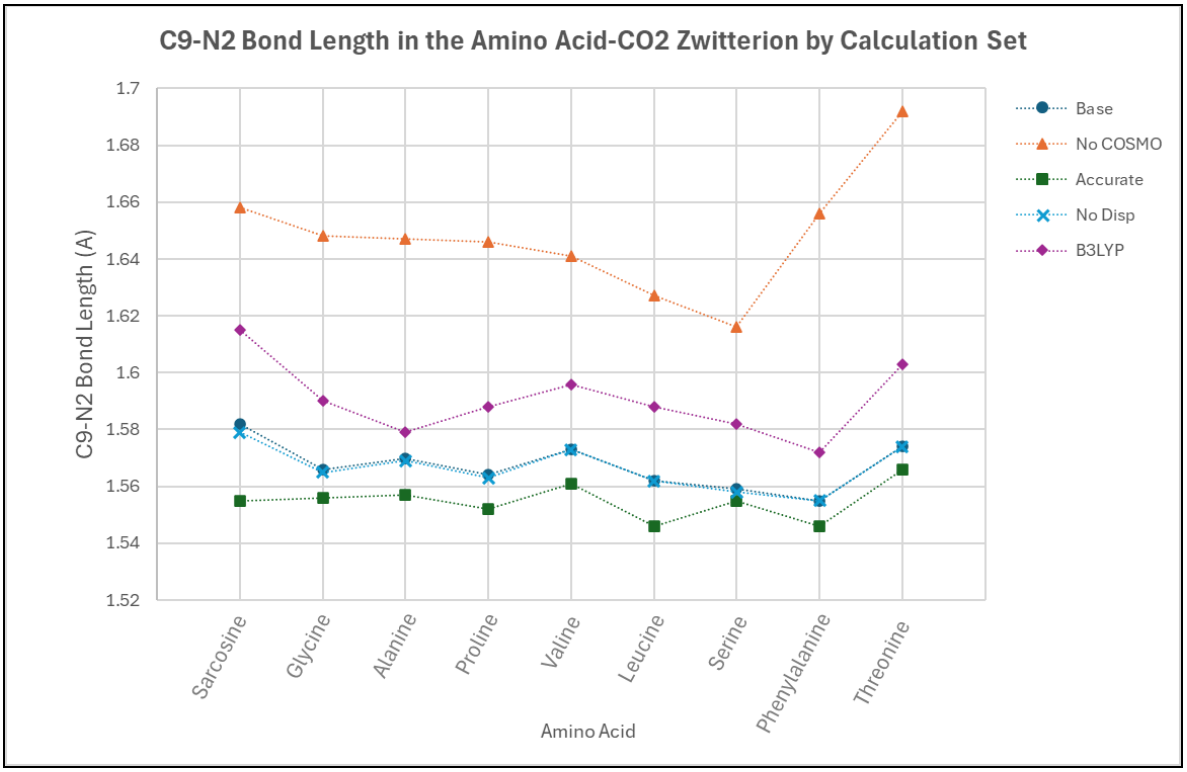


Figure 8: C9-N2 bond lengths in the amino acid-CO_2 zwitterion for nine amino acids and five calculation sets.

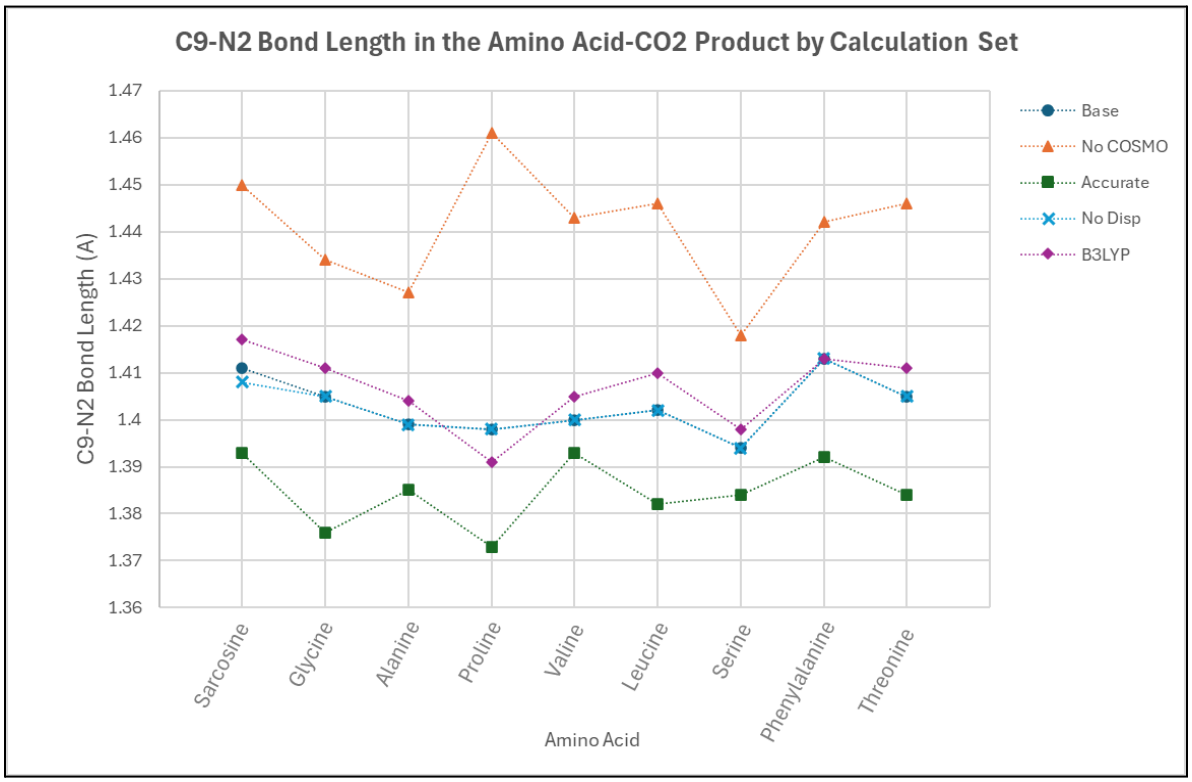


Figure 9: C9-N2 bond lengths in the amino acid-CO_2 product for nine amino acids and five calculation sets.

5. Discussion

The ΔE of Reaction 1 was calculated for nine different amino acids with five different sets of calculation variables. From Figure 4, it is clear that the variable change that had the largest effect on the calculated ΔE values was the removal of COSMO. This caused an average shift in the calculated ΔE values of -3.64 kcal/mol. Removing dispersion correction appears to have had the smallest effect, with the 'Base' and 'No Disp' calculation sets appearing very similar in Figure 4. Both the 'Accurate' and 'B3LYP' calculation sets appear to have differences from the base, but it is unclear from Figure 4 whether the differences are significant. A t-value of 0.014 for comparing the 'Base/No COSMO' data sets confirms that the two differ significantly. This is also the lowest t-value for Reaction 1, confirming that the removal of COSMO caused the largest statistical difference. The removal of dispersion correction did not have a significant effect on the calculated internal energy values, with a t-value of 0.686 when compared to the base case. Changing the basis set from 6-31G* to 6-311++G** had a significant effect on the internal energies (t-value 0.034). With a t-value of 0.077, changing the XC functional from M06-2X to B3LYP did not cause a statistically significant difference in the ΔE values. However, it is clear that the XC functional has a discernible effect on the calculations, as this t-value is close to 0.05.

ΔE values for Reaction 2 were calculated for nine amino acids with six sets of calculation variables. The magnitude of the observed difference in ΔE values between calculation sets was much larger for Reaction 2 than for Reaction 1. For Reaction 1, the removal of COSMO had the largest average effect of -3.64 kcal/mol. For Reaction 2, calculation sets 'Accurate,' 'Change of Base,' and 'No COSMO' had average effects of +16.0, +36.1, and +42.9 kcal/mol, respectively. The t-values for comparing those three calculation sets to the base case are $6.81*10^{-10}$, $1.46*10^{-14}$, and $1.51*10^{-7}$, respectively. The t-value for the comparison of the 'No COSMO' data set is higher than the other two, despite having a larger average change, due to the data set having a comparatively large standard deviation (See Figure 6). Similar to Reaction 1, the removal of dispersion corrections did not have a significant effect on the reaction ΔE; the t-value for the comparison of this data set to the base case was calculated to be 0.899. Another similarity to Reaction 1 was observed with the change of the XC functional. The comparison of the 'B3LYP' data set to the base case resulted in a t-value of 0.080, a value extremely close to that of the same comparison for Reaction 1.

A similar pattern that was observed for the t-test comparisons of the ΔE values was seen in the t-tests comparisons of the C9-N2 bond length. For both the zwitterion and product geometries, the removal of COSMO and the change in the basis set significantly impacted the C9-N2 bond length. T-values for comparing these bond length data sets to the base case are shown in Table 9. The removal of dispersion correction did not have a significant impact on the C9-N2 bond length for either the zwitterion or product geometries. T-values for comparing these 'No Disp' geometries to the base case are 0.842 and 0.907, respectively. Interestingly, while the change of the XC functional did have a significant effect on the zwitterion geometries, the effect was not significant for the product geometries. This is due to a larger average change in the C9-N2 bond length in the zwitterion when the XC functional is changed; the average change in the C9-N2 bond length in the zwitterion for the 'B3LYP' data set is +0.023Å, whereas for the product the average change is only +0.004Å.

T-values for the comparisons of both the Reaction 1 and 2 ΔE values and the C9-N2 bond length in the zwitterion and product geometries indicate that the removal of COSMO and the change in the basis set from 'Routine' to 'Accurate' both had statistically significant impacts on the calculated ΔE values and molecule geometries. The removal of dispersion corrections were not statistically significant to either the calculated ΔE values or the molecule geometries. The change of XC functional had a clear impact on both the calculated ΔE values and molecule geometries, though the change was not significant in all cases.

For both reactions, the removal of COSMO and the change of basis set from 6-31G* to 6-311++G** had statistically significant effects on the calculated ΔE values and molecule geometries. The removal of COSMO represents a fundamental change to the reaction system. Without COSMO, the solvent-solute interactions are not accounted for in calculations; removing these interactions is why the geometries and internal energies differ significantly from the base case. Several differences exist between the 6-31G* and 6-311++G** basis sets. Most notably, the latter includes polarization functions for heavy elements and diffuse functions for all elements included in the basis set, neither of which are present in 6-31G*. Likely, the addition of diffuse functions in the 6-311++G** set is largely responsible for the statistical difference of these calculations from the base case. Diffuse functions are particularly important for modeling anions, which are present in the reactants and products of both Reactions 1 and 2. The removal of dispersion corrections did not significantly impact the internal energies of either reaction; this is

likely because the M06-2X functional is parameterized for dispersion. Had B3LYP been used for the base case, dispersion corrections likely would have had a more significant impact on calculations. The change of the XC functional had an apparent but not statistically significant effect on the internal energies and geometries of both reactions. This difference may be due to the ability of M06-2X to describe dispersion interactions better than B3LYP, or it may be because M06-2X includes a higher percent of the Hartree-Fock exact exchange than B3LYP (54% vs 20%). In Reaction 2, the change of the base from OH^- to $RNHCOO^-$ had a significant impact on the reaction ΔE values. This is expected as the change of base in the reaction represents a change to the reaction mechanism. This highlights the importance of the chosen base when modeling the reaction. The change of base reaction for threonine is shown below.

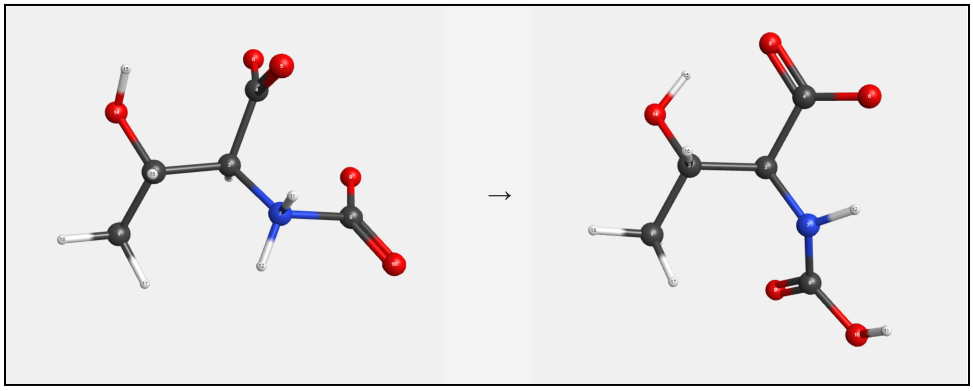


Figure 10: Final reactant and product geometries for the 'Change of Base' Reaction 2 of threonine.

The calculated ΔG values shown in Table 7 indicate that Reaction 1 is not spontaneous at 298K for the four amino acids analyzed. An additional driving force for the reaction, such as a catalyst or a change in pH, is be necessary for the reaction to proceed. Strongly negative ΔS values for Reaction 1 show that the reaction will become less favorable with increasing temperature for all four amino acids. Reaction 2 is spontaneous at 298K for the four amino acids, with the exception of alanine. It remains spontaneous at higher temperatures for glycine and proline but becomes less favorable for sarcosine and alanine. This analysis supports the formation of the Reaction 2 amino acid-CO_2 product for some amino acids; others may exist in solution as a zwitterion if Reaction 1 proceeds. Like Reaction 1, an additional driving force may be necessary to shift Reaction 2 toward the products. The decrease in length of the C9-N2 bond from the amino acid-CO_2 zwitterion to the product across all calculation sets belies increased stability and supports the formation of the Reaction 2 product.

6. Conclusions

The t-test comparisons of five data sets for Reaction 1 and six data sets for Reaction 2 show that removing solvation modeling (COSMO) and, separately, changing the basis set from 6-31G* to 6-311++G** both significantly affected the calculated reaction internal energies and geometries. The removal of dispersion correction did not significantly affect either, though this is likely because M06-2X is capable of accurately describing dispersion effects without the correction of DFT-D3. The change of XC functional from M06-2X to B3LYP did not have a significant effect on the calculated internal energies or geometries of Reactions 1 and 2; except for in the geometries of the amino acid-CO_2 zwitterions, where B3LYP calculated the C9-N2 bond length to be an average of 0.023Å longer than M06-2X. The change of base in Reaction 2 presented a significant change to the ΔE of the reaction. Based on these results and information from the literature review, DFT calculations performed at a 6-311++G** level with the M06-2X functional, solvation modeling with COSMO, and without dispersion correction are recommended as the most effective method, among those explored, for modeling the interactions of amino acids with CO_2. A 6-31G* basis set may be used to increase the speed of calculations at the expense of accuracy. The binding energies of the nine amino acid-CO_2 complexes were calculated as the sum of the ΔE values for Reactions 1 and 2. They were calculated as follows: sarcosine (-32.86 kcal/mol), glycine (-30.35 kcal/mol), alanine (-31.81 kcal/mol), proline (-34.02 kcal/mol), valine (-32.46 kcal/mol), leucine (-29.36 kcal/mol), serine (-31.90 kcal/mol), phenylalanine (-32.25 kcal/mol), and threonine (-29.03 kcal/mol). These values represent binding energies calculated using DFT with COSMO, M06-2X, a 6-311++G** basis set, and DFT-D3 dispersion correction. These calculated binding energies are very similar, with a standard deviation of 1.65 kcal/mol. This suggests that the binding energies of amino acid-CO_2 complexes are independent of the amino acid.

References

Becke, A. D. (1988). Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A*, *38*(6), 3098–3100. https://doi.org/10.1103/PhysRevA.38.3098

Church, J. A., Clark, P. U., Cazenave, A., Gregory, J. M., Jevrejeva, S., Levermann, A., Merrifield, M. A., Milne, G. A., Nerem, R. S., Nunn, P. D., Payne, A. J., Pfeffer, W. T., Stammer, D., & Unnikrishnan, A. S. (2013). *Sea level change* [Technical Report]. P.M.Cambridge University Press. https://drs.nio.res.in/drs/handle/2264/4605

Cremer, D. (2011). Møller–Plesset perturbation theory: From small molecule methods to methods for thousands of atoms. *WIREs Computational Molecular Science*, *1*(4), 509–530. https://doi.org/10.1002/wcms.58

Grimme, S., Antony, J., Ehrlich, S., & Krieg, H. (2010). A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics*, *132*(15), 154104. https://doi.org/10.1063/1.3382344

Guo, D., Thee, H., Tan, C. Y., Chen, J., Fei, W., Kentish, S., Stevens, G. W., & da Silva, G. (2013). Amino Acids as Carbon Capture Solvents: Chemical Kinetics and Mechanism of the Glycine + CO2 Reaction. *Energy & Fuels*, *27*(7), 3898–3904. https://doi.org/10.1021/ef400413r

Haszeldine, R. S. (2009). Carbon Capture and Storage: How Green Can Black Be? *Science*, *325*(5948), 1647–1652. https://doi.org/10.1126/science.1172246

Holst, J. van, Versteeg, G. F., Brilman, D. W. F., & Hogendoorn, J. A. (2009). Kinetic study of CO2 with various amino acid salts in aqueous solution. *Chemical Engineering Science*, *64*(1), 59–68. https://doi.org/10.1016/j.ces.2008.09.015

Hu, G., Smith, K. H., Wu, Y., Kentish, S. E., & Stevens, G. W. (2017). Screening Amino Acid Salts as Rate Promoters in Potassium Carbonate Solvent for Carbon Dioxide Absorption. *Energy & Fuels*, *31*(4), 4280–4286. https://doi.org/10.1021/acs.energyfuels.7b00157

Hu, G., Smith, K. H., Wu, Y., Mumford, K. A., Kentish, S. E., & Stevens, G. W. (2018). Carbon dioxide capture by solvent absorption using amino acids: A review. *Chinese*

Journal of Chemical Engineering, *26*(11), 2229–2237. https://doi.org/10.1016/j.cjche.2018.08.003

IPCC. (2005). *IPCC special report on carbon dioxide capture and storage*. Cambridge University Press.

IPCC. (2022). *Global Warming of 1.5°C: IPCC Special Report on Impacts of Global Warming of 1.5°C above Pre-industrial Levels in Context of Strengthening Response to Climate Change, Sustainable Development, and Efforts to Eradicate Poverty* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009157940

Jalan, A., W. Ashcraft, R., H. West, R., & H. Green, W. (2010). Predicting solvation energies for kinetic modeling. *Annual Reports Section "C" (Physical Chemistry)*, *106*(0), 211–258. https://doi.org/10.1039/B811056P

Jones, R. O. (2015). Density functional theory: Its origins, rise to prominence, and future. *Reviews of Modern Physics*, *87*(3), 897–923. https://doi.org/10.1103/RevModPhys.87.897

Jones, R. O., & Gunnarsson, O. (1989). The density functional formalism, its applications and prospects. *Reviews of Modern Physics*, *61*(3), 689–746. https://doi.org/10.1103/RevModPhys.61.689

Klamt, A. (1995). Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *The Journal of Physical Chemistry*, *99*(7), 2224–2235. https://doi.org/10.1021/j100007a062

Kumar, P. S., Hogendoorn, J. A., Versteeg, G. F., & Feron, P. H. M. (2003). Kinetics of the reaction of CO2 with aqueous potassium salt of taurine and glycine. *AIChE Journal*, *49*(1), 203–213. https://doi.org/10.1002/aic.690490118

Lee, C., Yang, W., & Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, *37*(2), 785–789. https://doi.org/10.1103/PhysRevB.37.785

Lu, G., Wang, Z., Bhatti, H., Fan, X., Umair, H., & Bhatti. (2023). Recent progress in carbon dioxide capture technologies: A review. *Clean Energy Science and Technology*, *1*. https://doi.org/10.18686/cest.v1i1.32

Matta, C. F. (2010). How dependent are molecular and atomic properties on the electronic structure method? Comparison of Hartree-Fock, DFT, and MP2 on a biologically relevant set of molecules. *Journal of Computational Chemistry*, *31*(6), 1297–1311. https://doi.org/10.1002/jcc.21417

Park, S., Song, H.-J., Lee, M.-G., & Park, J. (2014). Screening test for aqueous solvents used in CO2 capture: K2CO3 used with twelve different rate promoters. *Korean Journal of Chemical Engineering*, *31*(1), 125–131. https://doi.org/10.1007/s11814-013-0200-y

Rochelle, G. T. (2009). Amine Scrubbing for CO2 Capture. *Science*, *325*(5948), 1652–1654. https://doi.org/10.1126/science.1176731

Sandler, I., Chen, J., Taylor, M., Sharma, S., & Ho, J. (2021). Accuracy of DLPNO-CCSD(T): Effect of Basis Set and System Size. *The Journal of Physical Chemistry A*, *125*(7), 1553–1563. https://doi.org/10.1021/acs.jpca.0c11270

Smith, P., Davis, S. J., Creutzig, F., Fuss, S., Minx, J., Gabrielle, B., Kato, E., Jackson, R. B., Cowie, A., Kriegler, E., van Vuuren, D. P., Rogelj, J., Ciais, P., Milne, J., Canadell, J. G., McCollum, D., Peters, G., Andrew, R., Krey, V., … Yongsung, C. (2016). Biophysical and economic limits to negative CO2 emissions. *Nature Climate Change*, *6*(1), 42–50. https://doi.org/10.1038/nclimate2870

Song, H.-J., Park, S., Kim, H., Gaur, A., Park, J.-W., & Lee, S.-J. (2012). Carbon dioxide absorption characteristics of aqueous amino acid salt solutions. *International Journal of Greenhouse Gas Control*, *11*, 64–72. https://doi.org/10.1016/j.ijggc.2012.07.019

Tkatchenko, A., & Scheffler, M. (2009). Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Physical Review Letters*, *102*(7), 073005. https://doi.org/10.1103/PhysRevLett.102.073005

Tomasi, J., Mennucci, B., & Cammi, R. (2005). Quantum Mechanical Continuum Solvation Models. *Chemical Reviews*, *105*(8), 2999–3094. https://doi.org/10.1021/cr9904009

Vaidya, P. D., & Kenig, E. Y. (2007). CO2-Alkanolamine Reaction Kinetics: A Review of Recent Studies. *Chemical Engineering & Technology*, *30*(11), 1467–1474. https://doi.org/10.1002/ceat.200700268

Vaidya, P. D., Konduru, P., Vaidyanathan, M., & Kenig, E. Y. (2010). Kinetics of Carbon Dioxide Removal by Aqueous Alkaline Amino Acid Salts. *Industrial & Engineering Chemistry Research*, *49*(21), 11067–11072. https://doi.org/10.1021/ie100224f

Yu, W.-C., Astarita, G., & Savage, D. W. (1985). Kinetics of carbon dioxide absorption in solutions of methyldiethanolamine. *Chemical Engineering Science*, *40*(8), 1585–1590. https://doi.org/10.1016/0009-2509(85)80101-9

Zhao, Y., & Truhlar, D. G. (2008). The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts*, *120*(1), 215–241. https://doi.org/10.1007/s00214-007-0310-x

Appendix

Example Input Files

H_2O:

```
title "Water Base"

echo

charge 0

geometry
zmatrix
O
H 1 B1
H 1 B2 2 A1
variables
B1 0.941999857
B2 0.941999857
A1 105.4833155
end
end

basis noprint
 * library 6-31G*
end

dft
 maxiter 2000
 XC M06-2X
 mult 1
 disp vdw 3
end
set grid:eaf_size_in_dbl 9999999

driver
 maxiter 2000
end

cosmo
```

```
 dielec 78.0
 radius 1.52
1.2
1.2
 lineq  0
end

task dft optimize
task dft energy

property
  dipole
  mulliken
end

task dft property
```

OH-:

```
title "Hydroxide Base"

echo

charge -1

geometry
zmatrix
O
H 1 B1
variables
B1 1.049999654
end
end

basis noprint
 * library 6-31G*
end

dft
```

```
 maxiter 2000
 XC m06-2x
 mult 1
 disp vdw 3
end
set grid:eaf_size_in_dbl 9999999

driver
 maxiter 2000
end

cosmo
 dielec 78.0
 radius 1.52
1.2
 lineq  0
end

task dft optimize
task dft energy

property
  dipole
  mulliken
end

task dft property
```

CO_2:

```
title "CO2 Base"

echo

charge 0

geometry
zmatrix
C
```

```
O 1 B1
O 1 B2 2 A1
variables
B1 1.313000464
B2 1.313000464
A1 180.0000000
end
end

basis noprint
 * library 6-31G*
end

dft
 maxiter 2000
 XC m06-2x
 mult 1
 disp vdw 3
end
set grid:eaf_size_in_dbl 9999999

driver
 maxiter 2000
end

cosmo
 dielec 78.0
 radius 1.70
1.52
1.52
 lineq  0
end

task dft optimize
task dft energy

property
  dipole
  mulliken
end
```

task dft property

Sarcosine:

```
title "Sarcosine Base"

echo

charge -1

geometry
zmatrix
C
N 1 B1
C 2 B2 1 A1
C 3 B3 2 A2 1 D1
O 4 B4 3 A3 2 D2
O 4 B5 3 A4 2 D3
H 3 B6 2 A5 1 D4
H 3 B7 2 A6 1 D5
H 2 B8 1 A7 3 D6
H 1 B9 2 A8 3 D7
H 1 B10 2 A9 3 D8
H 1 B11 2 A10 3 D9
variables
B1 1.446476535
B2 1.447303093
A1 112.7145102
B3 1.514368149
A2 110.6604817
D1 -177.4008638
B4 1.343861571
A3 112.3166218
D2 73.42412516
B5 1.206508232
A4 126.9266639
D3 -105.8452025
B6 1.115714586
```

```
A5 110.7125113
D4 -56.47505029
B7 1.114987097
A6 109.4687711
D5 62.93047230
B8 1.020643866
A7 108.9531340
D6 121.5644570
B9 1.114152908
A8 110.4989873
D7 -61.06340265
B10 1.114623688
A9 110.9640262
D8 60.84530363
B11 1.113950467
A10 109.0578528
D9 -179.8708414
end
end

basis noprint
 * library 6-31G*
end

dft
 maxiter 2000
 XC m06-2x
 mult 1
 disp vdw 3
end
set grid:eaf_size_in_dbl 9999999

driver
 maxiter 2000
end

cosmo
 dielec 78.0
 radius 1.7
1.55
```

```
1.7
1.7
1.52
1.52
1.2
1.2
1.2
1.2
1.2
1.2
 lineq  0
end

task dft optimize
task dft energy

property
  dipole
  mulliken
end

task dft property
```

Sarcosine-CO_2 Zwitterion:

```
title "Sarcosine Zwitter Base"

echo

charge -1

geometry
zmatrix
C
N 1 B1
C 2 B2 1 A1
C 3 B3 2 A2 1 D1
O 4 B4 3 A3 2 D2
O 4 B5 3 A4 2 D3
```

H 3 B6 2 A5 1 D4
H 3 B7 2 A6 1 D5
C 2 B8 1 A7 3 D6
O 9 B9 2 A8 1 D7
O 9 B10 2 A9 1 D8
H 2 B11 1 A10 3 D9
H 1 B12 2 A11 3 D10
H 1 B13 2 A12 3 D11
H 1 B14 2 A13 3 D12
variables
B1 1.513722544
B2 1.527953130
A1 110.6729213
B3 1.519732722
A2 112.7070854
D1 -177.1058585
B4 1.314154027
A3 112.0359251
D2 -139.3327318
B5 1.203378978
A4 121.1326955
D3 38.37696964
B6 1.114508426
A5 109.6102233
D4 -58.60497930
B7 1.112258772
A6 109.9618022
D5 59.30882469
B8 1.595135769
A7 107.8400528
D6 120.3461733
B9 1.316181932
A8 122.7310752
D7 -107.1186891
B10 1.210184221
A9 114.6304501
D8 62.74404876
B11 1.121424919
A10 109.4251024
D9 -122.6943878

```
B12 1.112290058
A11 110.6779770
D10 58.98121469
B13 1.115289781
A12 111.3728063
D11 179.3206006
B14 1.114036558
A13 110.4078086
D12 -60.59211050
end
end

basis noprint
 * library 6-31G*
end

dft
 maxiter 2000
 XC m06-2x
 mult 1
 disp vdw 3
end
set grid:eaf_size_in_dbl 9999999

driver
 maxiter 2000
end

cosmo
 dielec 78.0
 radius 1.7
1.55
1.7
1.7
1.52
1.52
1.2
1.2
1.7
1.52
```

```
1.52
1.2
1.2
1.2
1.2
 lineq  0
end

task dft optimize
task dft energy

property
  dipole
  mulliken
end

task dft property
```

Sarcosine-CO_2 Product:

```
title "Sarcosine Product Base"

echo

charge -2

geometry
zmatrix
C
N 1 B1
C 2 B2 1 A1
C 3 B3 2 A2 1 D1
O 4 B4 3 A3 2 D2
O 4 B5 3 A4 2 D3
H 3 B6 2 A5 1 D4
H 3 B7 2 A6 1 D5
C 2 B8 1 A7 3 D6
O 9 B9 2 A8 1 D7
O 9 B10 2 A9 1 D8
```

H 1 B11 2 A10 3 D9
H 1 B12 2 A11 3 D10
H 1 B13 2 A12 3 D11
variables
B1 1.454327715
B2 1.460969975
A1 122.6786761
B3 1.523833456
A2 112.1299509
D1 -63.44073493
B4 1.205728147
A3 127.6009580
D2 -4.045073190
B5 1.316868397
A4 108.9774773
D3 172.4681585
B6 1.114961319
A5 107.3904348
D4 56.21812419
B7 1.114203394
A6 112.6268697
D5 173.8411159
B8 1.394452691
A7 117.4352065
D6 -178.4677010
B9 1.205636731
A8 122.6818482
D7 177.9283594
B10 1.326234036
A9 118.4723546
D8 -4.416948220
B11 1.113612139
A10 110.6105407
D9 -2.241620385
B12 1.114223118
A11 110.4928891
D10 117.2103527
B13 1.114053556
A12 110.4771907
D11 -121.1042505

```
end
end

basis noprint
 * library 6-31G*
end

dft
 maxiter 2000
 XC m06-2x
 mult 1
 disp vdw 3
end
set grid:eaf_size_in_dbl 9999999

driver
 maxiter 2000
end

cosmo
 dielec 78.0
 radius 1.7
1.55
1.7
1.7
1.52
1.52
1.2
1.2
1.7
1.52
1.52
1.2
1.2
1.2
 lineq  0
end

task dft optimize
task dft energy
```

```
property
  dipole
  mulliken
end

task dft property
```