Betanin pKₐ Prediction Using DFT Methods
Sergio A. Rodriguez* and Maria T. Baumgartner

ABSTRACT: Betalains can be used in food, drugs, and cosmetic industries and have shown their bioactive potential. These properties are strongly influenced by pH and other physicochemical conditions. The pKₐ values for the polyprotic Betanidin (Bd) molecule are unknown, and they are crucial to elucidate the oxidation mechanism in which its pigment is involved. In the present study, the values of pKₐ for all protic groups of Bd were analyzed using five hybrid density functionals (B3LYP, B3PW91, ωB97XD, PBE0, and M06-2X), five basis sets (6-31+G(d), 6-31+G(d,p), 6-311+G(d,p), and 6-311+G(d,p)) and the solvation model based on density (SMD) implicit solvation model. Moreover, one and three explicit water molecules were added to improve the solvation free energy values. Furthermore, the values of pKₐ of betanin, betalamic acid, and cyclo-dihydroxyphenylalanine (DOPA) were studied. Based on these analyses, we propose the acid–base behavior of Bd in water and develop new tools to understand their chemical reactivity.

1. INTRODUCTION

Betalains are water-soluble nitrogenous pigments found in high amounts in plant species from the order Caryophyllales.¹ Betalains are divided into two subclasses: betaxanthins (yellow-orange pigments) and betacyanins (red-violet pigments).² Betacyanins are derivatives of Betanidin (Bd), an iminium compound derived from betalamic acid and cyclo-dihydroxyphenylalanine (DOPA), whereas betaxanthins are obtained from the reaction of amines or α-amino acids with betalamic acid, for example, indicaxanthin (In) in prickly pear (Opuntia ficus-indica).² Typically, betalains show good solubility in water and alcohols.³ Betalains in aqueous solutions are stable in the pH range from 3.5 to 7;⁴,⁵ however, light, metal cations, high temperature, and oxygen reduce the dye stability.⁶

Betalain pigment can be used in food, drugs, and cosmetic industries as beet powder or beet juice concentrate.⁷ In recent years, betalains have shown their biological potential; they have antimicrobial, antiviral, and antitumoral effects and anti-inflammatory activity.⁸ In addition, the pure natural betalains and semisynthetic analogues have demonstrated to significant reduce of lipoxgenase and cyclooxygenase activity.⁹ The strong free radical scavenging capacity and antioxidant activity of these pigments and their modulation by different structural factors have been well documented.¹⁰ Moreover, the influence of pH and other physicochemical conditions on the antioxidant activity of betalains has been observed.¹⁰,¹¹

Betanin (Bn, E-number E162), the only betalin approved for use in food and almost mainly obtained from red beet crops,¹² is one of the best-known betacyanin dyes found in a few plants from the Caryophyllales family.⁵ Structurally, Bn is composed for Bd, linked β-glycosidic with glucose at C5. The presence of Bn in roots, stems, fruits, leaves, and flower petals results in a characteristic red/purple color.

Temperature is the main factor causing Bn decomposition, and light-induced decay at low temperatures may also be important.¹³ Molecular oxygen is involved in Bn photo-degradation, being undetectable in anaerobic conditions. Dehydrogenated and decarboxylated betanins were the main products formed in aqueous solutions upon UV irradiation.¹⁴

On the other hand, Bd was exposed to oxidation by enzyme tyrosinase, a polyphenol oxidase that plays a key role in the betalain biosynthetic scheme.¹⁵ Moreover, the same oxidation products profile of betalains by peroxidase enzyme was reported.¹⁶ Recently, the enzymatic oxidation of Bd and Bn, followed by chromatographic separation of the oxidation products with spectrophotometric and mass spectrometric detection (LC-DAD-MS/MS) was studied.¹⁷ In that article, dehydrogenated, decarboxylated, and o-quinone oxidation products were identified, and their proportions were closely related to the average medium pH. Besides, the oxidation mechanism was proposed from the fully protonated betalin. Similar results were reported for the electrochemical oxidation of Bd.¹⁸ The chemical oxidation of Bn by 2,2′-azinobis(3-
ethylbenzthiazoline-6-sulfonic acid) (ABTS) radical also resulted in the formation of dehydrogenated and decarboxylated derivatives, presumably followed by a one-electron oxidation product as an intermediate. On the other hand, the same product’s profile was found by the thermal treatment of betalains in aqueous and alcoholic media. Moreover, these compounds were identified as products from Bn oxidation catalyzed by Cu²⁺ cations in aqueous solutions at neutral pH, but an alternative mechanism was proposed. Besides, the oxidation of the excited Bn was studied using electrochemistry techniques and a two-electron oxidation mechanism was proposed without conclusive evidence. Despite the potential of betalains and based on the above discussion, it is clear that systematic research on their oxidation mechanism is still insufficient, particularly the study of pH incidences on the oxidation pathways. Acid dissociation constant (pKₐ) is essential for evaluating the acid−base properties of betalains, but they remain unknown. Direct measurement of the pKₐ values with traditional experimental approaches for each functional group involving complex aqueous samples can be defying. Thus, the accurate calculation of the values of pKₐ with theoretical methods is very interesting. An extensively used method for calculating the values of pKₐ implies the use of quantum chemical calculations, typically density functional theory (DFT), and the solvation model based on density (SMD), as an electron density-based solvation model. Many methodologies have been used to calculate the values of pKₐ with computational chemistry approaches. One of the most common methods involves the use of an appropriate thermodynamic cycle to calculate the aqueous Gibbs free energy of deprotonation. Another ever more common methodology is the “direct” approach, where the geometries are optimized and deprotonation free energies calculated directly in the presence of the continuum solvent without the necessity for gas-phase calculations. A succession of studies comparing methods using thermodynamic cycles and the direct approach showed that the two approaches can accomplish comparable accuracy. On the other hand, including first-shell explicit water molecules, is an additional tool to enhance the accuracy of pKₐ calculations. Using this methodology, it was proved that including explicit water molecules as hydrogen bond donors or acceptors to dicarboxylic acids reduced the error in calculations compared continuum solvation alone. In the present study, the values of pKₐ for all protic groups of Bd were analyzed using many computational approaches to unveil its acid−base behavior and develop new tools to understand their chemical reactivity in oxidation mechanisms in which this important molecule are involved.

2. RESULTS AND DISCUSSION

The oxidation mechanism of Betanidin (Bd) is highly affected by the medium pH. Therefore, it is necessary to know the values of pKₐ and the deprotonation positions of the Bd (Figure 1). In addition, it is generally assumed that the N16H₂ of Bd is deprotonated before the carboxylic acid groups.

2.1. Conformational Analysis. The study began by calculating the energies of different conformations of fully protonated Bd. The most stable conformer was used to calculate the values of pKₐ (Figure S1). This structure shows the presence of three intramolecular hydrogen bonds; one between the two hydroxyl groups attached to C5 and C6. The other two are formed by the oxygen atom from each carboxylic group in C15 and C17 with the hydrogen-bonded to N16; it is evidenced the dihedral angles between the donor and the receptor hydrogen bond groups are closed to 0° (Table S1). Similar optimized conformations were obtained with all basis sets used (Figure S2).
Next, the optimized Bd was deprotonated in each carboxylic group and optimized to obtain the lowest-energy structures for BdC2, BdC15, and BdC17 (Figure 2). It is important to remark that the main structural parameters are preserved compared to the uncharged parental Bd molecule.

### 2.2. Calculation of Values of pKₐ

#### 2.2.1. Effect of Calculation Methods in the Direct Approach

To determine the best methodology to calculate values of pKₐ, five DFT functionals (B3LYP, B3PW91,ωB97XD, PBE0, and M06-2X) and five basis sets (6-31+G(d), 6-31+G(dp), 6-31++G(dp), 6-311+G(dp), and 6-311++G(dp)) were used. These most stable conformers were used for the calculation of the values of pKₐ. Experimental estimated values for Betanin.

| Calculation Method | 6-31+G(d) | 6-31+G(dp) | 6-31++G(dp) | 6-311+G(dp) | 6-311++G(dp) |
|-------------------|-----------|------------|-------------|-------------|--------------|
| B3LYP             | 3.04      | 3.03       | 3.01        | 3.02        | 3.03         |
| B3PW91            | 3.05      | 3.04       | 3.02        | 3.03        | 3.04         |
| PBE0              | 3.06      | 3.05       | 3.03        | 3.04        | 3.05         |
| ωB97XD            | 3.07      | 3.06       | 3.04        | 3.05        | 3.06         |
| M06-2X            | 3.08      | 3.07       | 3.05        | 3.06        | 3.07         |

#### 2.2.2. Exploration of Fitting and Scaled Solvent-Accessible Surface (SAS) Models

To obtain pKₐ values using other different computational protocols, two general strategies were explored. These strategies correlate the experimental pKₐ with free energy values obtained by theoretical calculations. In the first place, the reported empirical fitted equations ("Fitting model") for the calculation of carboxylic acids pKₐ values were used (see Section 2.3). The calculated pKₐ values are reported in Table 1.

Recently, the pKₐ calculation of carboxylic acids was reported using a scaled solvent-accessible surface and SMD solvation model (SMDₛₐₛ). This is the second protocol we have applied (Table 1).

The pKₐ, C17 values calculated are between 0.96 and 1.57 (fitting and SMD values; Table 1). With these methodologies, the pKₐ, C2 and pKₐ, C15 are very similar, 2.11–2.64 and 2.25–2.39, respectively. It is important to remark that these strategies were developed mainly with monoprotic acids.

In conclusion, the pKₐ values reported with the B3LYP functional and mostly all basis set (6-31+G(dp), 6-31++G(dp), and 6-311++G(dp)) correlated better with the Fitting and SAS calculations. On the other hand, the pKₐ values calculated with the M06-2X level of theory were the most inaccurate values.

#### 2.2.3. Electron Density and Electrostatic Potential Surface Analysis

The C17COOH group was found to first deprotonate at all levels of theory, similar to the observation of other authors. This observation is contrary to the the pKₐ value of approximately 1.5 reported for C2COOH group of betanin. To understand this result, it is useful to analyze the electron density of the highest occupied molecular orbital (HOMO) and the electrostatic potential surface of Bd and its various deprotonated forms (Figure S3). The electron density of Bd is delocalized in the molecule’s conjugate system (the DOPA ring and the conjugated double bonds that include C17). After proton removal, the electron density delocalizes toward the site of the missing proton, especially when the H of C2COOH or C15COOH is removed. It is beneficial to take

---

Table 1. pKₐ Values Calculated Using Functional/Basis Set/SMD Implicit Model Level of Theory for First Deprotonation Sites of Betanin

| Calculation Method | 6-31+G(d) | 6-31+G(dp) | 6-31++G(dp) | 6-311+G(dp) | 6-311++G(dp) |
|-------------------|-----------|------------|-------------|-------------|--------------|
| B3LYP             | 3.04      | 3.03       | 3.01        | 3.02        | 3.03         |
| B3PW91            | 3.05      | 3.04       | 3.02        | 3.03        | 3.04         |
| PBE0              | 3.06      | 3.05       | 3.03        | 3.04        | 3.05         |
| ωB97XD            | 3.07      | 3.06       | 3.04        | 3.05        | 3.06         |
| M06-2X            | 3.08      | 3.07       | 3.05        | 3.06        | 3.07         |

Values obtained using reported fitting parameters (Galano et al.). Values calculated modifying the SMD cavity surface (Lian, Johnston, Parks, and Smith). Experimental estimated values for Betanin.
away the C17COOH hydrogen because the electron density is slightly influenced after removing a proton.

2.2.4. Explicit Water Molecule Evaluation. Due to the good relation between computational cost, time consumption, and pKₐ results expressed in Table 1, we decided to use 6-31+G(d,p) as a basis set to analyze the effects of explicit water molecules (Table 2).

Table 2. Calculated pKₐ Using Functional/6-31+G(d,p)/SMD and Explicit Water Molecules

|                     | one water                  | three waters               |
|---------------------|----------------------------|----------------------------|
|                     | pKₐC2COOH | pKₐC15COOH | pKₐC17COOH | pKₐC2COOH | pKₐC15COOH | pKₐC17COOH |
| B3LYP               | 2.97       | 2.51       | 1.51       | 1.26       | 1.00       | 0.84       |
| B3PW91              | 3.86       | 3.44       | 2.10       | 3.14       | 1.67       | 1.11       |
| PBE0                | 3.64       | 3.53       | 1.66       | 1.69       | 1.22       | 0.14       |
| ωB97XD              | 4.89       | 4.43       | 2.12       | 1.67       | 1.13       | 0.92       |
| M06-2X              | 2.48       | 2.90       | 0.62       | 0.11       | 1.06       | 0.07       |

The arrangement of one water molecule close to the protonation/deprotonation site improved the pKₐ values over calculations without an explicit water molecule (Table 2). The effect observed in comparison with the results without water molecules is moderate; most pKₐ values are shifted to lower values. Still, B3LYP/6-31+G(d,p) is a better approach. The SMD solvation model has some difficulties in accounting for the solvation of ions, just a single water molecule is capable of

Table 3. Calculated Values of pKₐ for All Protic Groups

|                     | pKₐ₁     | pKₐC17C₂<sup>a</sup> | pKₐC17C₁₅<sup>b</sup> | pKₐ₃ | pKₐ₄<sup>c</sup> | pKₐ₅<sup>d</sup> | pKₐ₆<sup>e</sup> |
|---------------------|----------|-----------------------|-----------------------|------|------------------|------------------|------------------|
| B3LYP               | 6-31+G(d)| −1.36                 | 1.17                  | 0.99 | 1.38             | 17.55            | 21.99            |
|                     | 6-31+G(d,p)| 1.11                 | 2.93                  | 3.38 | 5.71             | 10.59            | 19.97            |
|                     | 6-31+G(d,p)| 1.13                 | 2.96                  | 2.80 | 3.59             | 11.71            | 19.84            |
|                     | 6-31+G(d,p)| 0.65                 | 3.39                  | 2.77 | 3.82             | 11.72            | 19.75            |
|                     | 6-31+G(d,p)| 0.95                 | 3.01                  | 3.28 | 3.50             | 11.55            | 19.57            |
|                     | 6-31+G(d)| −1.33                 | 2.31                  | 1.70 | 2.64             | 9.88             | 18.21            |
|                     | 6-31+G(d)| 1.79                 | 4.09                  | 4.20 | 4.64             | 12.15            | 20.74            |
| B3PW91              | 6-31+G(d,p)| 1.19                 | 4.10                  | 4.21 | 4.52             | 12.18            | 20.77            |
|                     | 6-31+G(d,p)| 1.33                 | 4.58                  | 3.84 | 4.85             | 12.31            | 20.53            |
|                     | 6-31+G(d,p)| 1.61                 | 3.55                  | 3.60 | 4.63             | 11.93            | 20.51            |
|                     | 6-31+G(d)| −1.44                 | 1.05                  | 1.24 | 1.88             | 9.18             | 17.64            |
|                     | 6-31+G(d)| 0.94                 | 4.09                  | 3.31 | 4.44             | 11.88            | 19.79            |
| PBE0                | 6-31+G(d,p)| 1.20                 | 3.46                  | 3.72 | 4.53             | 11.58            | 20.28            |
|                     | 6-31+G(d,p)| 1.10                 | 3.50                  | 3.73 | 6.18             | 8.78             | 16.98            |
|                     | 6-31+G(d,p)| 1.39                 | 3.28                  | 3.49 | 4.57             | 11.55            | 20.25            |
|                     | 6-31+G(d)| −1.11                 | 1.62                  | 2.06 | 3.86             | 10.28            | 20.01            |
|                     | 6-31+G(d)| 1.32                 | 3.84                  | 4.27 | 5.26             | 13.81            | 23.35            |
|                     | 6-31+G(d)| 1.35                 | 3.87                  | 4.26 | 4.89             | 13.80            | 22.13            |
| ωB97XD              | 6-31+G(d,p)| 0.86                 | 3.94                  | 4.03 | 5.16             | 13.50            | 22.34            |
|                     | 6-31+G(d,p)| 1.13                 | 4.86                  | 3.82 | 4.93             | 13.61            | 21.99            |
|                     | 6-31+G(d)| −2.66                 | 0.62                  | 1.29 | 1.14             | 9.62             | 17.75            |
|                     | 6-31+G(d)| −0.18                 | 2.90                  | 2.92 | 3.43             | 12.02            | 20.14            |
| M06-2X              | 6-31+G(d,p)| 0.12                 | 3.03                  | 2.02 | 3.40             | 12.07            | 20.18            |
|                     | 6-31+G(d,p)| −1.16                 | 3.08                  | 2.49 | 3.71             | 11.45            | 19.87            |
|                     | 6-31+G(d,p)| −0.69                 | 2.68                  | 2.06 | 3.33             | 11.35            | 19.50            |
|                     | B3LYP     | 1.51                 | 3.48                  | 3.27 | 4.24             | 10.80            | 15.95            |
|                     | B3PW91    | 2.10                 | 4.72                  | 4.37 | 5.28             | 10.67            | 17.80            |
|                     | PBE0      | 1.66                 | 4.55                  | 4.13 | 4.96             | 10.80            | 17.20            |
|                     | ωB97XD    | 2.12                 | 4.77                  | 4.78 | 5.72             | 12.70            | 18.89            |
|                     | M06-2X    | 0.62                 | 2.96                  | 3.47 | 3.76             | 10.68            | 17.93            |
|                     | B3LYP     | 1.57                 | 2.65                  | 3.00 | 8.62             | 11.77            | 18.99            |
|                     | PBE0      | 1.30                 | 2.49                  | 2.89 | 8.27             | 11.64            | 19.35            |
|                     | ωB97XD    | 1.02                 | 2.78                  | 3.80 | 8.07             | 12.21            | 17.26            |
|                     | M06-2X    | 0.96                 | 2.49                  | 2.83 | 8.46             | 11.96            | 17.88            |
|                     | M06-2X SAS| 0.93                 | 3.43                  | 4.10 | 11.31            | 18.94            | 18.93            |

<sup>a</sup>pKₐC17C₂ = value for the second deprotonation in C2COOH. <sup>b</sup>pKₐC17C₁₅ = value for the second deprotonation in C15COOH. <sup>c</sup>pKₐ₂ = deprotonation in C6OH. <sup>d</sup>pKₐ₅ = deprotonation in C5OH. <sup>e</sup>pKₐ₆ = deprotonation in N16H. <sup>f</sup>Basis set 6-31+G(d,p). <sup>g</sup>Parameters described in the Materials and Methods section.

The arrangement of one water molecule close to the protonation/deprotonation site improved the pKₐ values over calculations without an explicit water molecule (Table 2). The effect observed in comparison with the results without water molecules is moderate; most pKₐ values are shifted to lower values. Still, B3LYP/6-31+G(d,p) is a better approach. The SMD solvation model has some difficulties in accounting for the solvation of ions, just a single water molecule is capable of
improving the pKₐ values, and the influence is higher when the water is hydrogen-bonded to the deprotonation site. However, the addition of three water molecules produces an over-stabilization of ionic species, and all pKₐ values decrease markedly moving away from the established reference values. Based on these results, we can postulate that in Bd first deprotonation occurs in the C17 carboxylic group.

2.3. Second and Subsequent Deprotonations. For evaluation, further deprotonations were performed following the same procedures except that only implicit SMD and explicit one water molecule models were used (Table 3). Once deprotonated, the C17COOH group calculations were not conclusive with respect to the second deprotonation of C2COOH (pKₐC17C2) or C15COOH (pKₐC17C15; Table 3 and Scheme 1).

The values obtained with Fitting and M06-2X SAS methods suggest that the second deprotonation occurs in C2COOH. The predicted pKₐ values were between 2.49 and 2.78 in the fitting calculation; besides, a higher value of 3.43 was obtained by M06-2X SAS. The main difference at this point is that M06-2X SAS was developed only with monoprotic species and Fitting calculation with few diprotic acids.31

On the other hand, when the pKₐ values were computed using the direct approach with the SMD implicit model, some theories predict the second deprotonation in C2COOH and the others in C15COOH with values in the ranks of 2.68−4.86 and 2.92−4.93, respectively (Table 3). Once again, 6-31+G(d) basis set calculations yielded the worst results. Besides, four of the five functionals combined with 6-31+G(d,p) first predict the C2COOH deprotonation. On the contrary, when the basis sets were 6-31++G(d,p), 6-311+G(d,p), and 6-311++G(d,p), the C15COOH group was deprotonated second, with the exception of the PDBE0. The average pKₐ value for C17C2 calculated without the 6-31+G(d) was 3.56 ± 0.60 (mean ± standard deviation (SD)) and that calculated for C17C15 using the implicit solvent model was 3.41 ± 0.70. However, calculations with one explicit water suggest that pKₐC17C2 and pKₐC17C15 are practically the same with each method used (Table 3). The average values are 4.10 ± 0.80 and 4.00 ± 0.60 for pKₐC17C2 and pKₐC17C15, respectively; the computational methodology cannot differentiate between them.

Finally, the third carboxylic deprotonation was studied to obtain pKₐ3 values (Table 3 and Scheme 1). The fitting pKₐ values were dispersed between 2.83−3.80, and M06-2X SAS gives 4.1 pKₐ units; at this point, the reference values are only guidance due mainly to the polyprotic nature of Bd that was not contemplated in these models as was mentioned previously. Also, all calculated pKₐ3 values were in the rank of 3.33−5.71 (excluding PBE0/6311++G(d,p) value), approximately 1 unit bigger than the second pKₐ values; the average value of the explicit approach was 4.8 ± 0.8. As a partial conclusion, we can propose the next deprotonation order pKₐC17 < pKₐC17C2 ≈ pKₐC17C15 < pKₐ3 (Scheme 1).

On the other hand, the carboxylic deprotonations are followed by C6OH ionization (Scheme 1). In all calculations, the deprotonation order of the OH groups was C6OH first, followed by C5OH deprotonation (see Table S3 for one...
example). The $pK_a$ was estimated by the implicit SMD model at an average value of $11.5 \pm 1.4$, while this range is more precise in case one explicit water molecule is used, $10.7 \pm 0.1$, except that the $\omega$B97XD values are always bigger. As expected, the SMD implicit model presents more inaccurate results due to the failure to solvate anionic and polyprotic species. A similar behavior was observed for $pK_a$ and $pK_6$ calculations, where the values present high discrepancies due to the complexity in reproducing correct solvation of ionic species. The SMD calculations suggest markedly that $pK_a$ corresponds to CSOH deprotonation, followed by N16H deprotonation, having average values of $19.9 \pm 1.5$ and $23.3 \pm 1.9$, respectively; with one explicit water molecule, the values of $17.2 \pm 0.9$ and $16.5 \pm 1.8$ are almost the same. The $pK_a$ values of monoprotic alcohols and amines calculated using only one explicit water molecule and SMD solvation are found to have average errors of $3$ for $pK_a$ and $1$ for $pK_6$ units, respectively. Clearly, this background supports the results obtained in the hydroxyl and amine deprotonations. Finally, the $pK_a$ values of CSOH and C6OH for Bt were obtained using B3LYP/6-311+ +G(d,p) and three explicit water molecules, as reported by Thapa and Schlegel. The Bt C6OH deprotonation has a $pK_a$ of 6.64, and the subsequent deprotonation of CSOH gives a $pK_a$ of 11.20 (Table S3); these values are close to 8.5 assigned for C6OH for Bn.

### 2.4. Carboxylic Acid Deprotonation Order

At this point, the more relevant issue of the present article is the prediction that the first deprotonation of Bt occurs in C17COOH, while it was postulated previously that this happens in the C2COOH group. The isoelectric point of Bn was determined between 1.5 and 2.0. The value of $pK_a$ of the two less acidic carboxylic groups of Bn was 3.5; both facts imply that the $pK_a$ of the most acidic carboxylic group in Bn should be lower than 2.

On the other hand, Bn hydrolysis affords betalamic acid (Bt) and cyclo-DOPA-5-O-β-D-glucoside (Figure S4). The cyclo-DOPA-5-O-β-D-glucoside $pK_a$ values were experimentally determined to be $pK_a$COOH = 1.58, $pK_a$NH2 = 4.75, and $pK_a$OH6 = 9.42. Recently, values of $pK_a$ of Bt (unknown) were calculated using three theoretical methods (Marvin, Jaguar, and Epik) in the range of 2.82–3.43 and 4.16–5.18 for the C10 and the C7, respectively; also, the $pK_a$C15COOH (2.96–4.0) and $pK_a$C17COOH (2.90–5.19) for Bn were calculated. Moreover, the deprotonation of Bn C2COOH ($pK_a$ = 1.5), C15COOH ($pK_a$ = 3.7), C17COOH ($pK_a$ = 3.0), OH6 ($pK_a$ = 8.5), and N16H ($pK_a$ = 10.3) were obtained using the Marvin method.

Intending to unveil the mechanism of the first deprotonation of betalains (Bn and Bt), we decided to study the behavior of cyclo-DOPA-5-O-β-D-glucoside, Bt, and Bn by calculating the $pK_a$ values for each molecule. For this purpose and based on the results presented here, a very good approach could be obtained using the B3LYP/6-31+ +G(d,p)/SMD with one explicit water model with a low computational cost. Following this approach, cyclo-DOPA-5-O-β-D-glucoside gave the values of $pK_a$COOH = 1.39, $pK_a$NH2 = 4.52, and $pK_a$OH6 = 14.78 and Bt gave the values of $pK_a$C10COOH = 2.95 and $pK_a$C7COOH = 4.10; these results are in very good agreement with the above-described results. Furthermore, Bn showed the same behavior as Bt (Figure S5).

On the other hand, an exhaustive analysis of the reported $^1$H NMR of Bn and Bt was made to unveil that the carboxylic group deprotonated first; also, $^1$H NMR was estimated for different deprotonation states of Bt using the DFT/SMD/B3LYP 6-31G+(d,p) methodology (Table S4). The H2 chemical shift ($\delta$) of Bn at pH < 1 was reported to be 5.23 ppm, and a similar value was observed at pH 2.4. The H2 atom is more deshielded at pH 5 ($\Delta \delta$ of ca 0.29 ppm) due to the deprotonation of the carboxylic group attached to C2; a similar result was observed for Bt (Table S4). It is clear that C2COOH cannot be the carboxylic group of the first deprotonation. The effect of pH on other hydrogen atoms is less pronounced. In Bn, the H15 chemical shift cannot be assigned a lower pH due to the signal overlapping with deuterated water, but the $\delta$ at other pH values are similar. However, the Bt H15 presents a $\delta$ of 4.97 at a very acidic pH, but no experimental information is available at higher pH. All these results support the fact that the methodology proposed for carboxylic acid $pK_a$ determination is well-grounded and reliable and reproduces the experimental evidence. It is clear that the first deprotonation of Bt and Bn occurs in C17COOH, followed by C2COOH and C15COOH.

### 3. CONCLUSIONS

In the present article, we have developed a systematic study of Bt acid–base behavior using different theoretical approaches. In the first place, two models, Fitting and M06-2X_SAS, were used to have reference $pK_a$ values because there is no experimental information. Further, the direct approach was used to compute the values of $pK_a$ using a wide range of DFT methods with the SMD model, one and three explicit water molecules. The results show similar $pK_a$ values for all methods using a more complex basis set of 6-31+ +G(d,p). Besides, we have shown that the inclusion of one explicit water molecule increases the accuracy of the calculated $pK_a$ values mainly by improving the accuracy of the solvation free energies of the anions. The DFT B3LYP/631+ +G(d,p)/SMD with one water molecule allowed to analyze the system behavior quite well with a low computational cost. It is important to remark that Bt is a polyprotic molecule, whose experimental values of $pK_a$ are still unknown, and systematic theoretical methods reported only allow the calculation of $pK_a$ of monoaions and a few diaions. This study evaluated a key parameter, $pK_a$, which is needed to understand the structure and reactivity of Bt in solution.

### 4. MATERIALS AND METHODS

#### 4.1. Computational Details

Calculations were performed by the Gaussian 09 revision E01 series of programs using five hybrid density functionals (B3LYP, B3PW91, $\omega$B97XD, PBE0, and M06-2X), five basis sets (6-31+ +G(d), 6-31+ +G(dp), 6-31+ +G(dp), 6-311+ +G(dp), and 6-311+ +G(dp)), and the SMD implicit solvation model. Geometries were fully optimized in aqueous solution. Harmonic frequencies were calculated to confirm that the structures were minima on the potential energy surface and to obtain the thermal and entropic contributions to the free energies. A conformational analysis was carried out considering all of the possible torsion angles of carboxylic and hydroxyl groups and the angles of C2N1C11H and C11C12C13C14 (Figure S1).

#### 4.2. Direct Approach

The calculation of $pK_a$ was based on the use of the direct approach, given by the proton dissociation reaction shown in eq 1. The $pK_a$ value of the molecule HA was calculated according to eq 2, where $G^{\gamma}_{\text{aq}}$ is
and \( G_{\text{aq,H+}} \) are the standard free energies of the deprotonated and protonated species, respectively, calculated directly in aqueous solution at 298.15 K.

\[
\text{HA}_{\text{aq}} + \text{H}_2\text{O}^* \rightarrow \text{A}^- + \text{H}_3\text{O}^+ \quad \Delta G^o \quad \text{K}_a = \frac{\Delta G^o}{2.303 RT} = \frac{G_{\text{aq,A}}^o + G_{\text{aq,H+}}^o - G_{\text{aq,H+}}^o}{2.303 RT}
\]

The Gibb's free energy of a proton in the aqueous phase is calculated using the following equations:

\[
G_{\text{aq,H+}}^o = G_{\text{aq,H}^+}^o + \Delta G_{\text{aq,solv}}^o + \Delta G_{\text{1 atm}→1 \text{ M}}^o \quad (3)
\]

\( G_{\text{aq,H+}}^o = -6.287 \text{ kcal/mol at 298 K} \)

\( \Delta G_{\text{aq,solv}}^o \text{H}^+ = -265.9 \text{ kcal/mol is the aqueous phase solvation free energy of the proton, taken from the literature.} \)

\( \Delta G_{\text{1 atm}→1 \text{ M}}^o = RT (24.46) = 1.89 \text{ kcal/mol is a correction term for the change in a standard state of 1 atm to 1 mol/L.} \)

The symbols \( ^o \) and \( ^* \) denote the standard state of 1 mol/L and 1 atm, respectively.

\[
\begin{align*}
G_{\text{aq,A}}^o &= E + \text{ZPVE} + \text{(electronic and thermal free energy correction)} \\
G_{\text{aq,H+}}^o &= E + \text{ZPVE} + \text{(electronic and thermal free energy correction)}
\end{align*}
\]

where \( E \) is the electronic energy for \( \text{A}^- \) or \( \text{HA} \) obtained by structure optimization, \( \text{ZPVE} \) is the zero-point vibrational energy in solution. All \( \text{pK}_a \) calculations are given in the Supporting Information.

### 4.3. Fitting Model

The reported empirical fitted equations for the calculation of carboxylic acids \( \text{pK}_a \) values were used (eq 4).

\[
pK_a = m(G_{\text{aq,A}}^o - G_{\text{aq,H+}}^o) + C
\]

The parameters \( m \) and \( C \) and the levels of theory used are listed in Galano et al. (see Table S2). The Gibbs free energies were obtained as described in Section 4.2. In the Supporting Information, the calculations can be found.

### 4.4. Scaled Solvent-Accessible Surface (SAS) Model

The calculation is similar to that we presented for the M06-2X/6-31+G(d,p) level of theory in Section 4.2; the same equations were used. In the geometry optimizations, the surface type and the scaling factor options in the SCRF section were tuned. By choosing SAS as the solute–solvent boundary, the solvent radius (1.385 Å for water) is added to the intrinsic Coulomb radii to construct the cavity. It is necessary to add the scale factor \( \alpha = 0.485 \) and the surface of SAS in the input file (see an example in Supporting Information). No explicit water molecules were used.

### 4.5. Explicit Water Molecules Evaluation

To improve the calculation of the solvation effects that are important for \( \text{pK}_a \) calculations, mainly in charged structures, we included one and three explicit water molecules directly hydrogen-bonded to the site being protonated/deprotonated. For each hydrogen-bonding site, several orientations of the added water were considered and only the lowest-energy structure was used. For optimal cancellation of errors in \( \text{pK}_a \) calculations (as Section 4.2), the water molecule was hydrogen-bonded to the same site in the molecule and its deprotonated form. Typical arrangements of water near the protonation/deprotonation site are shown in Figure 3a,b.

![Figure 3. Arrangement of explicit water molecules near the COOH and COO⁻ groups: (a) one explicit water molecule and (b) three explicit water molecules. Some of the hydrogen bond lengths are shown (in Å).](https://pubs.acs.org/doi/10.1021/acsomega.0c00904)

For \( \text{Bd} \) with one explicit water molecule, the COOH group provides a H atom to form a hydrogen bond with the O of the water, whereas in the deprotonated form, the COO⁻ forms a hydrogen bond with the H atom of the water molecule (Figure 3a). When there are three explicit water molecules near the COOH group, the O of one accepts the H from the COOH group and the other two donate H to form three hydrogen bonds. Besides, in the deprotonated form, COO⁻ forms a hydrogen bond with one hydrogen atom from each of the three water molecules (Figure 3b). Similar conformations and the same number of hydrogen bonds in the parent molecules were kept as consistent as possible to avoid any bias that may result in unsystematic contributions to the calculated free energy. For the molecule having more than one stable conformer, the one with the lowest energy was used in the final calculations.

### ASSOCIATED CONTENT

**Supporting Information**

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.0c00904.

Most stable conformation of \( \text{Bd} \); structure parameters for the most stable conformation of \( \text{Bd} \) optimized at SMD/functional/6-31+G(d); optimized structures obtained for \( \text{Bd} \) with all computational methods; electron density (isovalue = 0.04) and electrostatic potential surface (isovalue = 0.02 density = 0.04. red = −0.1, blue = 0.65) of the HOMO of \( \text{Bd} \) and deprotonated forms calculated at DFT/B3LYP/6-31+G(d,p); the parameters “m” and “C” extracted from Galano et al., optimized with 6-31+G(d) basis set; CSOH and C6OH \( \text{pK}_a \) values obtained with SMD/B3LYP/6-311++G(d,p); computed and reported \( ^1 \text{H} \) NMR of betanin and Betanidin; betanin hydrolysis products; \( \text{Bn} \) \( \text{pK}_a \) values determined with B3LYP/6-31+G(d,p)/SMD with one explicit water
molecule; coordinates of the geometry optimized with SMD/B3LYP/6-31+G(d,p); geometries with one explicit water molecule optimized with SMD/B3LYP/6-31+G(d,p); geometries with three explicit water molecules optimized with SMD/B3LYP/6-31+G(d,p) in a docx document; and all calculated values of pKa in an excel document. Example of SAS input file in a docx document (PDF)

Method (6-31+G(d,p)/SMD,3W) and method (B3LYP/6-31++G(d,p)/SMD,3W) (XLSX)

■ AUTHOR INFORMATION

Corresponding Author
Sergio A. Rodriguez – CONICET, Instituto de Ciencias Químicas, F.AyA, UNSE, Santiago del Estero 4200, Argentina; orcid.org/0000-0002-0904-5817; Email: drsergiorod@gmail.com

Author
Maria T. Baumgartner – INFIQC-Dpto. Química Orgánica, Fac. Ciencias Químicas, UNC, Ciudad Universitaria, Córdoba 5000, Argentina

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.0c00904

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors are grateful to Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), FOHCYT, SECYT UNSE, and SECYT UNC for funding. This work used computational resources from CCAD-UNC, which is part of SNCAD-MinCyT, Argentina.

■ REFERENCES

(1) Gandía-Herrero, F.; García-Carmona, F. Biosynthesis of Betalains: Yellow and Violet Plant Pigments. Trends Plant Sci. 2013, 18, 334−343.
(2) Delgado-Vargas, F.; Jiménez, A. R.; Paredes-López, O. Natural Pigments: Carotenoids, Anthocyanins, and Betalains - Characteristics, Biosynthesis, Processing, and Stability. Crit. Rev. Food Sci. Nutr. 2000, 40, 173−289.
(3) Bartoloni, F. H.; González, L. C. P.; Rodríguez, A. C. B.; Dörr, F. A.; Pinto, E.; Bastos, E. L. Photophysics and Hydrolytic Stability of Betalains in Aqueous Trifluoroethanol. Monatsh. Chem. 2013, 144, 567−571.
(4) Nilsson, T. Studies into the Pigments in Beetroot (Beta Vulgaris L. Ssp. Vulgaris Var. Rubra L.). Lantbrukshögsk. Ann. 1970, 36, 179−219.
(5) Strack, D.; Vogt, T.; Schliemann, W. Recent Advances in Betalain Research. Phytochemistry 2003, 247−269.
(6) Herbach, K. M.; Stintzing, F. C.; Carle, R. Betalain Stability and Degradation - Structural and Chromatic Aspects. J. Food Sci. 2006, 71, R41−R50.
(7) Georgiev, V. G.; Weber, J.; Knesche, E.-M.; Denev, P. N.; Bley, T.; Pavlov, A. I. Antioxidant Activity and Phenolic Content of Betalain Extracts from Intact Plants and Hairy Root Cultures of the Red Beetroot Beta Vulgaris Cc. Detroit Dark Red. Plant Foods Hum. Nutr. 2010, 65, 105−111.
(8) Esatbyegolu, T.; Wagner, A. E.; Schini-Kerth, V. B.; Rimbach, G. Betanin-A Food Colorant with Biological Activity. Mol. Nutr. Food Res. 2015, 59, 36−47.
(9) Vidal, P. J.; López-Nicolás, J. M.; Gandía-Herrero, F.; García-Carmona, F. Inactivation of Lipoxigenase and Cyclooxygenase by Natural Betalains and Semi-Synthetic Analogues. Food Chem. 2014, 154, 246−254.
(10) Gandía-Herrero, F.; Escrivano, J.; García-Carmona, F. Structural Implications on Color, Fluorescence, and Antiradical Activity in Betalains. Plant. 2010, 232, 449−460.
(11) Gandía-Herrero, F.; Escrivano, J.; García-Carmona, F. The Role of Phenolic Hydroxy Groups in the Free Radical Scavenging Activity of Betalains. J. Nat. Prod. 2009, 72, 1142−1146.
(12) Gaertner, V. L.; Goldman, I. L. pigment distribution and total dissolved solids of selected cycles of Table Beet from a Recurrent Selection Program for Increased Pigment. J. Am. Soc. Hortic. Sci. 2005, 130, 424−433.
(13) Attoe, E. L.; Von Elbe, J. H. Photochemical Degradation of Betanine and Selected Anthocyanins. J. Food Sci. 1981, 46, 1934−1937.
(14) Skopińska, A.; Tuwalsa, D.; Wybraniec, S.; Starzak, K.; Mitka, K.; Kowalski, P.; Szaleniec, M. Spectrophotometric Study on Betain Photodegradation. Challenges Med. Technol. 2012, 3, 34−58.
(15) Gandía-Herrero, F.; Escriván, J.; García-Carmona, F. Characterization of the Activity of Tyrosinase on Betadin. J. Agric. Food Chem. 2007, 55, 1546−1551.
(16) Martinez-Parra, J.; Muñoz, R. Characterization of Betacyanin Oxidation Catalyzed by a Peroxidase from Beta Vulgaris L. Roots. J. Agric. Food Chem. 2001, 49, 4064−4068.
(17) Wybraniec, S.; Michalowski, T. New Pathways of Betanin and Betain Enzymatic Oxidation. J. Agric. Food Chem. 2011, 59, 9612−9622.
(18) Ho, J.; Ertem, M. Z. Calculating Free Energy Changes in Continuum Solvation Models. J. Phys. Chem. B 2016, 120, 1319−1329.
(19) Wybraniec, S.; Starzak, K.; Skopińska, A.; Nemzer, B.; Pietrzkowski, Z.; Michalowski, T. Studies on Nonenzymatic Oxidation Mechanisms in Neobetain, Betain, and Decarboxylated Betains. J. Agric. Food Chem. 2013, 61, 6465−6476.
(20) Kumorwiczew, A.; Szmy, N.; Popenda, L.; Pietrzkowski, Z.; Wybraniec, S. Alternative Mechanisms of Betacyanin Oxidation by Complexation and Radical Generation. J. Agric. Food Chem. 2019, 67, 7455−7465.
(21) Knorr, F. J.; McHale, J. L.; Clark, A. E.; Marchioro, A.; Moser, J.-E. Dynamics of Interfacial Electron Transfer from Betain to Nanocrystalline TiO2: The Pursuit of Two-Electron Injection. J. Phys. Chem. C 2015, 119, 19030−19041.
(22) Wendel, M.; Nizinski, S.; Tuwalsa, D.; Starzak, K.; Szt, D.; Pukula, D.; Sikorski, M.; Wybraniec, S.; Burdzinski, G. Time-Resolved Spectroscopy of the Singlet Excited State of Betain in Aqueous and Alcoholic Solutions. Phys. Chem. Chem. Phys. 2015, 17, 18152−18158.
(23) Ho, J.; Ertem, M. Z. Calculating Free Energy Changes in Continuum Solvation Models. J. Phys. Chem. B 2016, 120, 1319−1329.
(24) Ho, J.; Ertem, M. Z. Are Thermodynamic Cycles Necessary for Continuum Solvent Calculation of PKas and Reduction Potentials? Phys. Chem. Chem. Phys. 2015, 17, 2859−2868.
(25) Ho, J.; Coote, M. L. A Universal Approach for Continuum Solvent PKA Calculations: Are We There Yet? Theor. Chem. Acc. 2010, 125, 3−21.
(26) Marenich, A. V.; Ding, W.; Cramer, C. J.; Truhlar, D. G. Resolution of a Challenge for Solvation Modeling: Calculation of Dicarboxylic Acid Dissociation Constants Using Mixed Discrete-Continuum Solvation Models. J. Phys. Chem. Lett. 2012, 3, 1437−1442.
(27) Frisch, M. J.; Trucks, G. W.; Schlegel, B. H.; Scuseria, E. G.; Robb, A. M.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, P. H.; Izmaylov, F. A.; Bloino, J.; Zheng, G.; Sonnenberg, L. J.; Frisch, M. J.; Trucks, W. G.; Schlegel, B. H.; Scuseria, E. G.; Robb, A. M.; C. G. B.; Scalmani, Barone, V.; Mennucci, B.; H.; P. G. A.; Nakatsuji, Caricato, M.; Li, X.; Hratchian, P. H.; Izmaylov, F. A.; Bloino, J.; Zheng, G.; Sonnenberg, L. J.; M. H. K.; E. M.; Toyoda, Fukuada, R.; 13758
