

This is the peer-reviewed version of the article:

Cvijetić, I.N., Pešić, M.P., Todorov, M.D. et al. Struct Chem (2018) 29: 423.
<https://doi.org/10.1007/s11224-017-1039-3>



This work is licensed under the
[Attribution-NonCommercial-NoDerivatives 4.0 International \(CC BY-NC-ND 4.0\)](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Tautomerism of 4-phenyl-2,4-dioxobutanoic acid. Insights from pH ramping NMR study and quantum chemical calculations

Ilija N. Cvijetić,^{1*} Miloš P. Pešić,² Miljana D. Todorov,² Branko J. Drakulić,^{3&} Ivan O. Juranić,³
Tatjana Ž. Verbić,^{2*} Mire Zloh⁴

¹*Innovation Center of the Faculty of Chemistry, Ltd, Studentski Trg 12-16, 11000 Belgrade, Serbia*

²*Faculty of Chemistry, University of Belgrade, Studentski Trg 12-16, 11000 Belgrade, Serbia*

³*Department of Chemistry-ICHTM, University of Belgrade, Njegoševa 12, 11000 Belgrade, Serbia*

⁴*School of Life and Medical Sciences, University of Hertfordshire, Hatfield, AL10 9AB, United Kingdom*

Abstract

Aryldiketo acids (ADKs) exhibit the variety of biological activities, mainly due to large affinity toward divalent metal ions. Metal complexation ability of ADKs, as well as interactions with proteins, depend on tautomeric form present in solution. The main aim of this study was to fully explore the tautomeric preferences of 4-phenyl-2,4-dioxobutanoic acid (4PDA), as ADKs representative, in aqueous media at different pH values. 1D and 2D NMR spectroscopy in combination with quantum chemical calculations was applied in order to better understand the tautomeric preferences of 4PDA. The data in highly acidic media are especially interesting since there are no such findings in the literature due to low solubility of ADKs in molecular form.

At low pH values, where 4PDA is unionized, the most abundant tautomeric form is enol with keto group closer to phenyl ring. At higher pH values, mixture of two 4PDA ionic forms coexists in solution. Their ratio calculated according to NMR data fits the values predicted using two experimentally determined pK_a values. Based on the complexity of ¹H NMR spectrum of monoanionic 4PDA form, coexistence of two stable rotamers was assumed. In an alkaline media 4PDA is mostly present in dianionic form. As π -electrons of dianion are delocalized over an entire keto-enol moiety, spectral distinction between tautomers was not possible.

Quantum chemical calculations were used to predict relative stability of tautomers. The predictions were in good accordance with experimental results only in case when explicit water molecule was included in calculations.

Keywords: Aryldiketo acid; keto-enol tautomerism; NMR spectroscopy; quantum chemical calculations.

Acknowledgments: Ministry of Education, Science, and Technological Development of Serbia supported this work, Grant No. 172035.

*Corresponding authors: tatjanad@chem.bg.ac.rs (+381 63 255 356, ORCID 0000-0002-6348-1644);
ilija@chem.bg.ac.rs (+381 64 280 9835, ORCID 0000-0002-5568-6172);

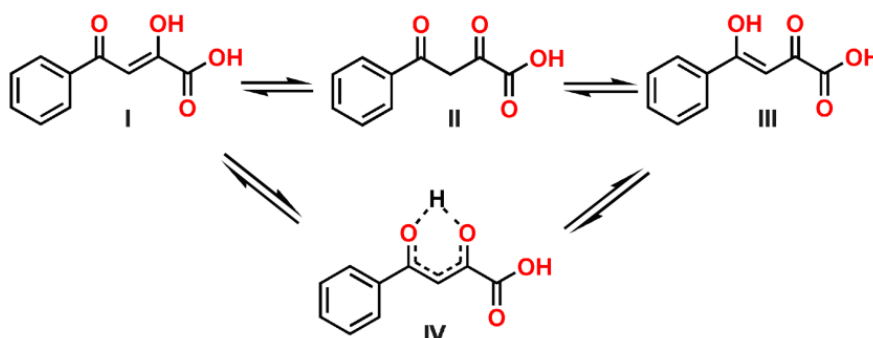
& We regret to inform that Branko Drakulić has passed away since completion of this work.

1. Introduction

Molecular properties and biological activities of aryldiketo acids (ADKs), an important class of molecules with widespread biological activities [1-6], are affected by keto-enol tautomerism. Tautomerism is of particular interest in studies of small organic molecules recognition properties, including protein-ligand interactions, since different tautomers of molecules of this class have different hydrogen-bond acceptor/hydrogen bond donor patterns, as well as different metal complexation abilities [7].

ADKs act by functional sequestration of Mg^{2+} ion in the active center of HIV-1 integrase (IN) [8], an enzyme responsible for the integration of viral DNA into the host genome. It was shown that ADKs' metal complexation ability depends on tautomeric form that is predominant in solution [9], and that Mg^{2+} preferentially reacts with enolate form **I** (**Scheme 1**) [10]. Similarly, in crystal structures of influenza virus PA protein (PDB entries 4E5G and 4E5H) and La Crosse orthobunyavirus L-protein (PDB entry 2XI7) ADKs are involved in interactions with two Mn^{2+} ions [11, 12]. In crystal structures of ADKs complexed with various bacterial enzymes, both diketo and enol forms are found [13, 14]. Furthermore, hydrolytic C–C bond cleavage of 1,3-diketones (β -diketones) by β -ketolases (EC 3.7.1.1 to EC 3.7.1.10) is sensitive to tautomeric form in which a 1,3-diketone is present in solution [10, 15, 16].

ADKs simultaneously exist in two enolate forms (**Scheme 1, I and III**), conformationally locked by the pseudo-ring, and one diketo form (**II**) having two rotatable bonds responsible for their conformational flexibility [15, 17]. As a continuation of our group's work on structure-property relationships [18-20] and biological activities of ADKs [6], we aimed to further explore their tautomeric preferences.



Scheme 1 Tautomerization of 4-phenyl-2,4-dioxobutanoic acid (4PDA) in aqueous solution

It was shown previously that the enol form of 1,3-diketones is thermodynamically favored compared to diketo form due to stabilization *via* intramolecular hydrogen bonding. In the ground state of avobenzene (1,3-diketone used in sunscreen products as a UV light absorber), the presence of other forms such as non-chelated enol and its rotamers is also hypothesized [21, 22].

Keto-enol tautomerization of 1,3-diketones has been studied extensively in various chemical systems using a range of analytical techniques (NMR, IR, HPLC, gas electron diffraction) [15, 23, 24]. It is enhanced in polar, protic solvents. Density functional theory (DFT) calculations showed that tautomerization is highly enhanced in water, as two water molecules assist process *via* transition state analog to E2 mechanism TS [25]. Tautomerization may be difficult without highly activated proton relays of water hydrogen bonds, *i.e.* the methylene C–H bond in the keto form cannot be cleaved readily by the nucleophilic sources other than a water molecule connected by hydrogen-bond networks. The newly formed enol O–H bond participates in the intermolecular hydrogen bond rather than in the intramolecular one. NMR spectroscopy and DFT calculations of phenindione (cyclic 1,3-diketone) and derivatives showed that the predominant tautomer for these compounds in DMSO solution is enol form [26].

In a quantum chemical study of the structure and stability of diketo acid HIV-1 IN inhibitors, 5-CITEP and L-731,988, enol forms were more stable than diketo form, with the energy differences for these molecules ranging from 15 to 28 kJ/mol [17]. Two enol forms had similar energies with only 3 kJ/mol differences. Because practically no energy barrier between these two enol forms was observed, it was proposed that they can interconvert easily, and delocalized transition state (six-membered ring) was suggested.

Detailed study of keto-enol tautomerism of eleven 4-alkyl- and 4-aryl-2,4-diketobutanoic acids has shown that enolate **I** (**Scheme 1**) is a predominant form (98%) in a protic solvent (CDCl₃) [15]. The equilibrium ratios of aqueous solution structures of aliphatic 2,4-diketobutanoic acids (2,4-diketo, 2-enol-4-keto, and 2-hydrate-4-keto) were markedly affected by solution pH value within wide pH range (1.5–10.5). At pH 7.5 ratio of these structures was approximately 4:5:1, at low pH values 2-hydrate predominated (\approx 50%) and at high pH values 2-enolate carboxylate was dominant (\approx 80%) while 2-hydrate was not detected. As aromatic 2,4-diketobutanoic acids are less soluble in acidic media, tautomerization study was performed only in solutions with pH \geq 5.5. The authors reported that three tautomeric forms (**I–III**) were not distinguished in the NMR spectra due to the formation of pseudodienolate form (**IV**) and fast interconversion between two enolate forms. To the best of our knowledge, no experimental data on keto-enol tautomerism of ADKs in aqueous solution with pH \leq 5.5 was published so far.

In order to fully explore a potential of ADKs as biologically active molecules with metal complexing ability, it is important to have knowledge about their tautomeric forms in aqueous media at different pH values. NMR experiments at a wide range of pH values, starting from extreme acidic to basic conditions, are performed for assessing the tautomeric preferences of 4PDA

as a representative scaffold for active ADKs. Results are explained and rationalized with the aid of theoretical calculations and previously determined acid-base properties of the compound of interest.

2. Experimental

2.1. Materials and methods

Synthesis and characterization of 4PDA were already described [20]. The same sample was used in all NMR measurements described herein. All other chemicals were purchased from Fluka, Aldrich, or Merck, having >98% purity, and were used as received.

^1H , ^{13}C , COSY, HMQC, and HMBC NMR spectra of 4PDA in aqueous solution at different pH values were acquired using Bruker Avance 500/125 MHz NMR spectrometer. pH Values were measured using Corning 120 pH-meter equipped with Corning Ag/AgCl microelectrode.

2.2. 1D and 2D NMR spectra of 4PDA

All NMR spectra of 4PDA were acquired at $t=25\pm 1$ °C, and constant ionic strength ($I=0.1$ M (NaNO_3)). TSP was used as the internal standard for spectra calibration, and chemical shifts (δ) are given in *ppm*. Measured pH values are converted to pD according to relation: $\text{pD}=\text{pH}_{\text{measured}}+0.4$ [27]. The sample was dissolved in an appropriate solvent (CF_3COOD for $\text{pH}<0$, $\text{D}_2\text{O}/\text{CD}_3\text{COOD}$ for pH 2.09; acetate buffer-*d* – for pH 4.41; and carbonate buffer for pH 7.80 and 9.20), and spectra were calibrated using TMS as an internal standard.

The 1D ^1H spectrum was acquired using sweep width of 10330 HZ and 65536 data, giving a digital resolution of 0.157 Hz/point and an acquisition time of 3.17 s. Pulse width of 11.50 μs and relaxation delay of 2 s were used for 256 transients.

The 1D ^{13}C spectrum was acquired using sweep width of 30030 Hz and 131072 data points giving a digital resolution of 0.229 Hz/point and an acquisition time of 2.18 s. Pulse width of 10 μs and relaxation delay of 2 s were used for 6000 transients.

2D HMQC (pulse program *inv4gpqf*) and HMBC (pulse program *inv4gplplrndqf* - optimized for coupling constant 10 Hz) spectra were acquired with 1024 points in direct dimensions (sweep width 4882 Hz) and 256 complex points (sweep width 31440 Hz) giving digital resolutions of 4.76 and 122 Hz/point respectively. Digital resolution for complex dimension has increased to 30.7 Hz/point after zero filling to 1024 data points.

2.3. Solubility studies

Intrinsic solubility (solubility of the molecular form (H_2A)) of 4PDA was determined by ‘shake-flask’ method in 1 M HCl. An excess of the accurately weighed compound was added to 5.0 mL of 1 M HCl; 5 samples were prepared. All samples were stirred and thermostated overnight at $t=25\pm 1$ °C. After equilibration, samples were filtered, aliquots appropriately diluted with HCl, and the concentration of 4PDA in saturated solution determined spectrophotometrically at the wavelength of the absorption maximum, $\lambda_{max}=308.8$ nm. Conformity with Beer’s law had previously been verified.

2.4. Computational chemistry studies

The full geometry optimizations of different 4PDA tautomers, in neutral (H_2A) and monoanionic (HA^-) forms, were performed at the MP2 level of theory, using the 6-31G(d,p) basis set, calculating the force constants at every point (Opt=CalcAll), and tightening the cutoff for forces and step size to determine the convergence (Opt=Tight).

The NMR shifts of the three tautomers of molecular form of 4PDA with the carboxyl H in *out* orientation were predicted applying the Gauge-Independent Atomic Orbital (GIAO) method with MP2/6-311++G(d,p) by SP calculations on fully optimized geometries [28]. Implicit water solvation model (IEF-PCM) was applied. The GIAO predicted magnetic shielding of tetramethylsilane (TMS) was taken as a reference.

For the optimization of the systems comprising one water molecule and different tautomeric forms of 4PDA we used optimized geometries of tautomers and manually added one explicit water molecule in the proximity of carboxyl OH. For each tautomer, the system was initially optimized holding 4PDA rigid and allowing movement of the water molecule by semiempirical MO PM6 method in MOPAC2016 [29]; then full optimization without constraints was performed by MP2/6-31G(d,p). The influence of solvent was also simulated applying the implicit (water) solvation model, IEF-PCM, as default in Gaussian09 [30].

3. Results and Discussion

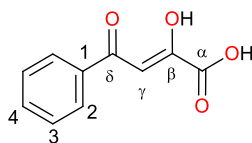
3.1. NMR study of aqueous 4PDA solutions at different pH values

Brecker *et al.* [15] used the NMR spectroscopy to study the aqueous structures of 4PDA in phosphate buffer at pH 7.5. The authors reported broad signals of C_γ H atom, as well as broad signals of C_β, C_γ, and C_δ carbon atoms (for atom labeling see **Table 1**). Interconversion between two enolic forms is very fast, and therefore it is impossible to ascribe signals to any distinct tautomer. Signal broadening could be explained by the finding of Guthrie *et al.* [31] that enolization rate constants of 2,4-diketo acids are significantly increasing as pH value increases.

Maurin *et al.* [32] studied NMR spectra of 4PDA in pure water, and in buffers pH 7.5 and 10.0. In water, they observed only one species, and attributed signals to fully protonated 4PDA in enol **I** tautomeric form (**Scheme 1**). However, as 4PDA in aqueous media act as weak diprotic acid ($pK_{a1} = 2.06$; $pK_{a2} = 7.56$) sparingly soluble in water [20], it will partially dissociate, giving a mixture of molecular and monoanionic form. Therefore, we recorded NMR spectra in highly acidic medium (**Fig. 1-3, Fig. S1**) in order to suppress the dissociation and obtain signals of the pure molecular (H₂A) form of 4PDA.

The low intrinsic solubility of 4PDA ($1.08 \pm 0.05 \times 10^{-3}$ M) explains why long (overnight) signal acquisition was necessary for 1D and 2D NMR spectra recording in the highly acidic medium. Full structure-spectra assignments were achieved using COSY, HMQC (ESM, **Fig. S1**), and HMBC spectra (**Fig. 3**) of 4PDA recorded in CF₃COOD. Experimental and calculated ¹H and ¹³C NMR chemical shifts of 4PDA are given in **Table 1**.

Table 1 Experimentally obtained and calculated ^1H and ^{13}C NMR chemical shifts for three 4PDA tautomeric forms (**Scheme 1**) in CF_3COOD .



δ (ppm)												
Atom	C_α	C_β	C_γ	H_γ (H_γ')	C_δ	C_1	C_2	H_2 (H_2')	C_3	H_3 (H_3')	C_4	H_4
experim.	163.41	161.59	95.06	7.23 ^a	189.46	129.85	124.05	7.96 ^b	125.05	7.50 ^c	131.13	7.64 ^d
enol I calc.	152.58	149.45	85.60	7.70	181.50	126.42	110.75 (111.11)	8.41 (8.74)	116.75 (117.20)	8.07 (8.23)	116.61	7.72
enol III calc.	151.09	167.25	82.28	7.64	167.28	123.27	109.56 (110.45)	8.22 (8.29)	116.66 (116.88)	8.08 (8.12)	116.23	7.72
diketo II calc.	150.28	177.19	36.38	3.89 (5.14)	182.29	122.98	110.47 (111.11)	8.12 (8.91)	117.31 (118.49)	8.14 (8.29)	117.49	7.79

^a)(s, 1H); ^b)(d, $J = 7.37$ Hz, 2H); ^c)(t, $J = 7.90$ Hz, 2H); ^d)(t, $J = 7.47$ Hz, 1H)

Singlet at 7.23 ppm in ^1H NMR spectrum (**Fig. 1**) is the signal of the vinyl group H atom (H_γ , **Table 1**). As the exchange rate between H_γ and D atom from CF_3COOD is fast, the integral of this signal is smaller than expected. Furthermore, the lack of a singlet around 4.5 ppm in ^1H NMR spectrum (this part of the spectra is not shown in **Fig. 1**), as well as lack of a signal around 50 ppm in ^{13}C NMR spectrum (**Fig. 2**), both expected for $>\text{CH}_2$ group (C_γ) in diketo form **II**, indicate that diketo form does not exist in solution at concentrations detectable by NMR. The presence of hydrate form is also excluded as no signals characteristic for this form were observed. The C_γ atom of hydrate form would also give a signal around 50 ppm in ^{13}C NMR spectrum, as well as the singlet at chemical shift below 4 ppm in ^1H NMR spectrum. This finding is in accordance with findings of Brecker et al. [15], who found this form in significant amount only in 4-alkyl-2,4-dioxobutanoic acids, but not in 4-phenyl-2,4-dioxobutanoic acid.

Due to H/D exchange, signal at 95 ppm in ^{13}C NMR spectrum consists of C-H singlet (C-H decoupled spectrum) as well as C-D triplet, as shown in an inset of **Fig. 2**. This proves that the ^1H signal at 7.23 ppm is not an impurity but the real signal of C-H group of enolic tautomer.

Signals at 7.33 ppm (s, 1H), 7.80 ppm (t, 1H), and 8.12 ppm (d, 2H), although very weak, indicate the existence of enol **III** form, but its concentration is negligible. In highly acidic media (pH<0), enolization rate constant is very low [31], allowing the coexistence of two enol tautomers in solution.

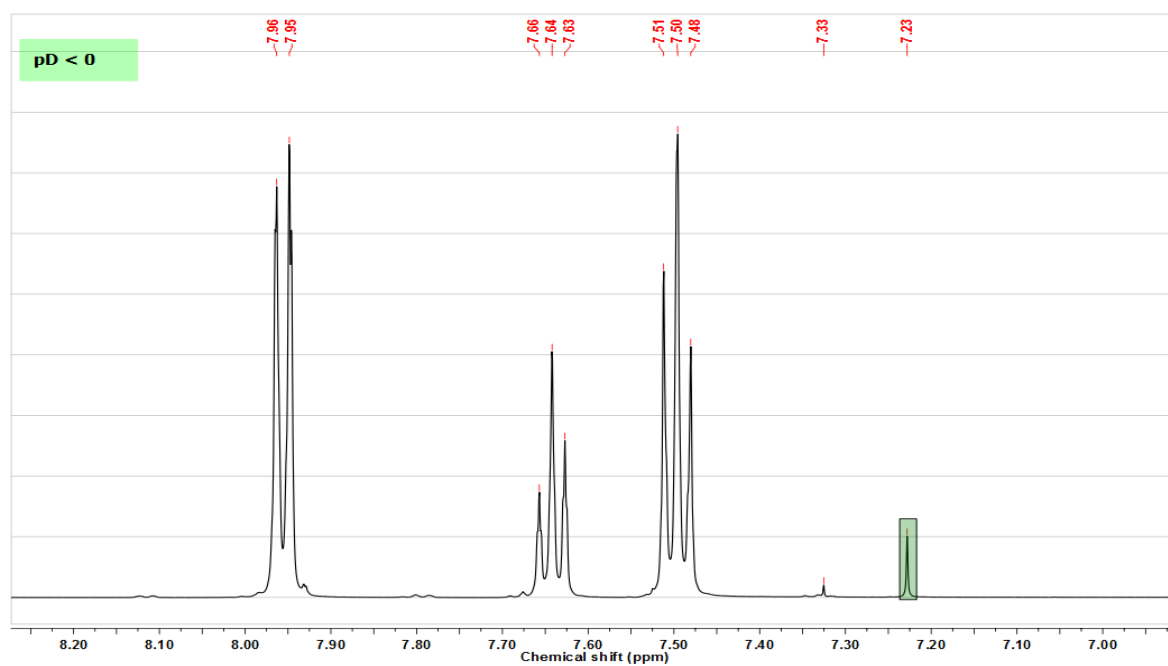


Fig. 1 ^1H NMR spectrum of 4PDA in CF_3COOD ($t = 25 \pm 1$ °C), where compound exists solely as a H_2A form. The signal of enolic, H_γ atom is highlighted (green box)

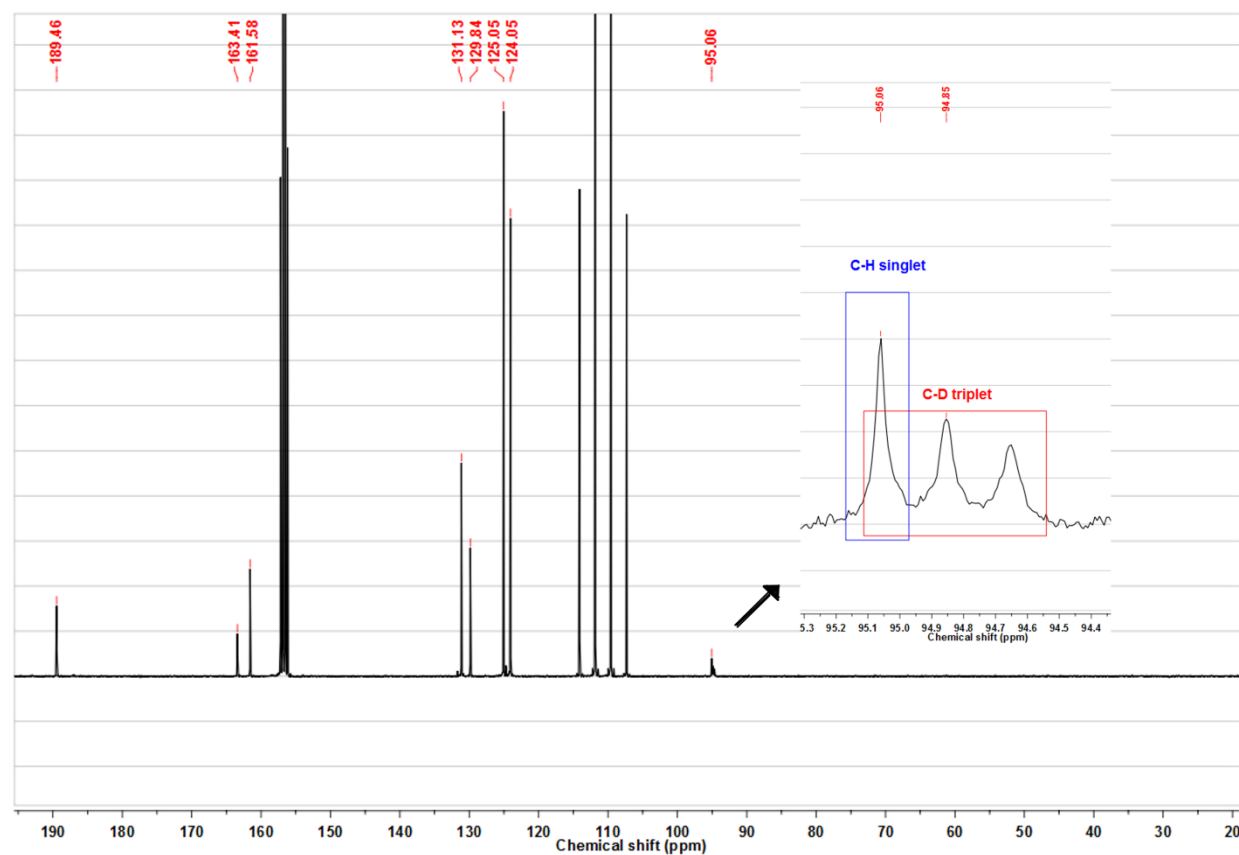


Fig. 2 ^{13}C NMR spectrum of 4PDA in CF_3COOD ; ($t = 25 \pm 1$ °C). The inset shows a magnified area around 95 ppm.

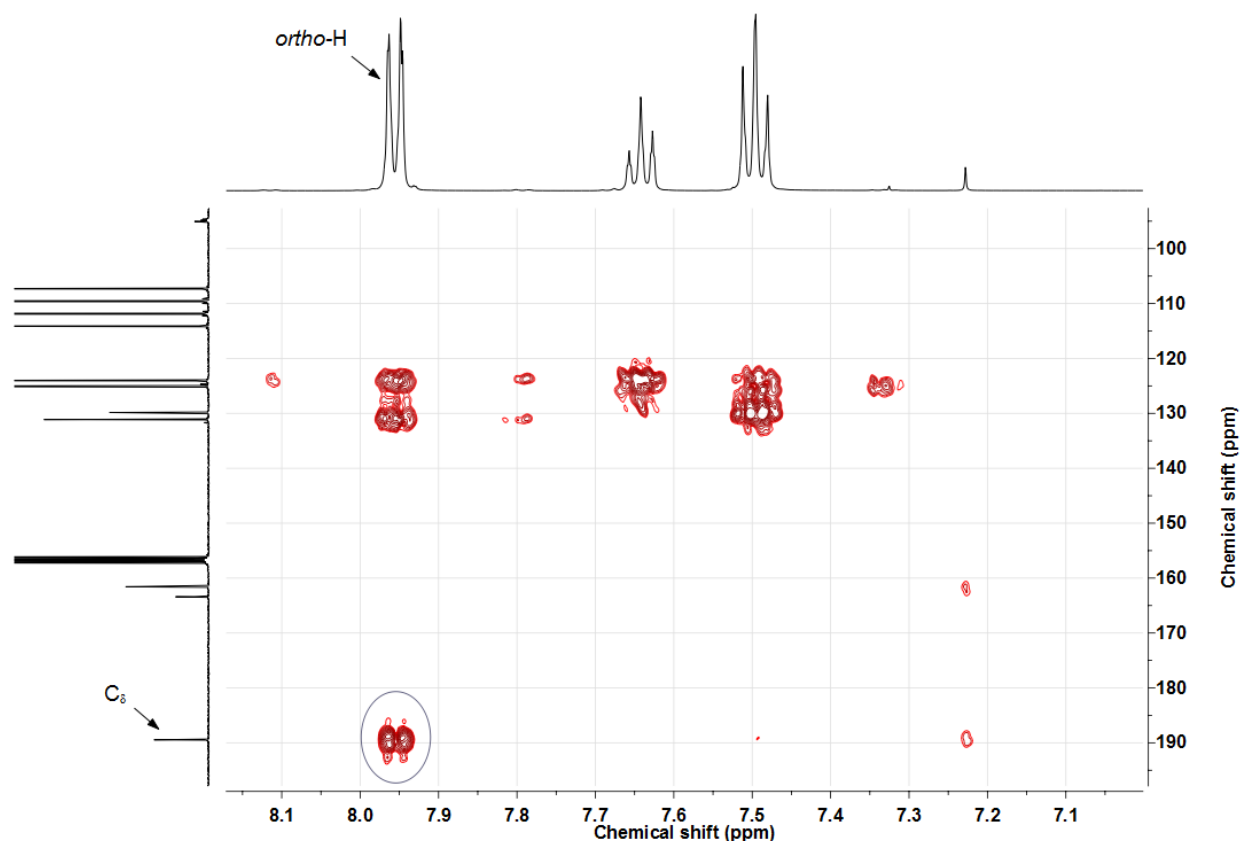


Fig. 3 HMBC NMR spectrum of 4PDA in CF_3COOD ; $t = 25 \pm 1$ °C; $I = 0.1$ M (NaNO_3). The encircled cross-peak depicts the long range correlation between *ortho*-phenyl hydrogens and C_δ atom

HMBC Spectrum (**Fig. 3**) provides important data about the major tautomeric form of 4PDA in CF_3COOD solution. Strong coupling between *ortho*-phenyl hydrogens and C_δ keto group carbon atom (circled signal) indicate that enol **I** is the major form of 4PDA in solution. If an enol form **III** was the predominant one, no coupling between *ortho*-hydrogens and C_β would be visible in HMBC spectrum. This confirms the predominance of enol **I** form of 4PDA in highly acidic solution.

Calculated ^1H NMR spectrum for diketo tautomer **II** has signals at 3.89 and 5.14 ppm, assigned to H atoms from $>\text{CH}_2$ group; corresponding C_γ atom has signal at 36.38 ppm in ^{13}C spectrum. The asymmetric structure of the most stable geometry of 4PDA in a diketo form **II** results in the different chemical environment of two H atoms of a methylene group. As a consequence, two signals appeared in calculated ^1H NMR spectrum.

It was mentioned that signals characteristic for the diketo tautomer were not observed in experimentally obtained NMR spectra of 4PDA in CF_3COOD . During routine characterization of some ADKs in aprotic solvents (CDCl_3 or $\text{DMSO}-d_6$), signals characteristic for diketo tautomer were present in ^1H and ^{13}C NMR spectra. H atoms from $>\text{CH}_2$ group give one broad signal at 4.2

to 4.5 ppm, depending on substitution pattern on the phenyl ring, while in ^{13}C NMR spectrum, the signal at ~ 50 ppm could be found [18-20].

Calculated NMR shifts of enolic hydrogens (H_7) are different for two enol tautomers (**Table 1**). Thus, the low intensity signal at 7.33 ppm in experimentally obtained ^1H NMR spectrum (**Fig. 1**) could be ascribed to the second, less abundant, enol tautomer **III**.

^1H NMR spectra of 4PDA were also acquired in solutions with different acidity, within pD range of 1–10. As the acidity decreases, ^1H NMR spectrum becomes more complicated because carboxylic group dissociates, leading to coexistence of H_2A and HA^- forms in solution. NMR spectrum of 4PDA in $\text{D}_2\text{O}/\text{CD}_3\text{COOD}$ mixture (pD = 2.49) is shown in **Fig. 4**.

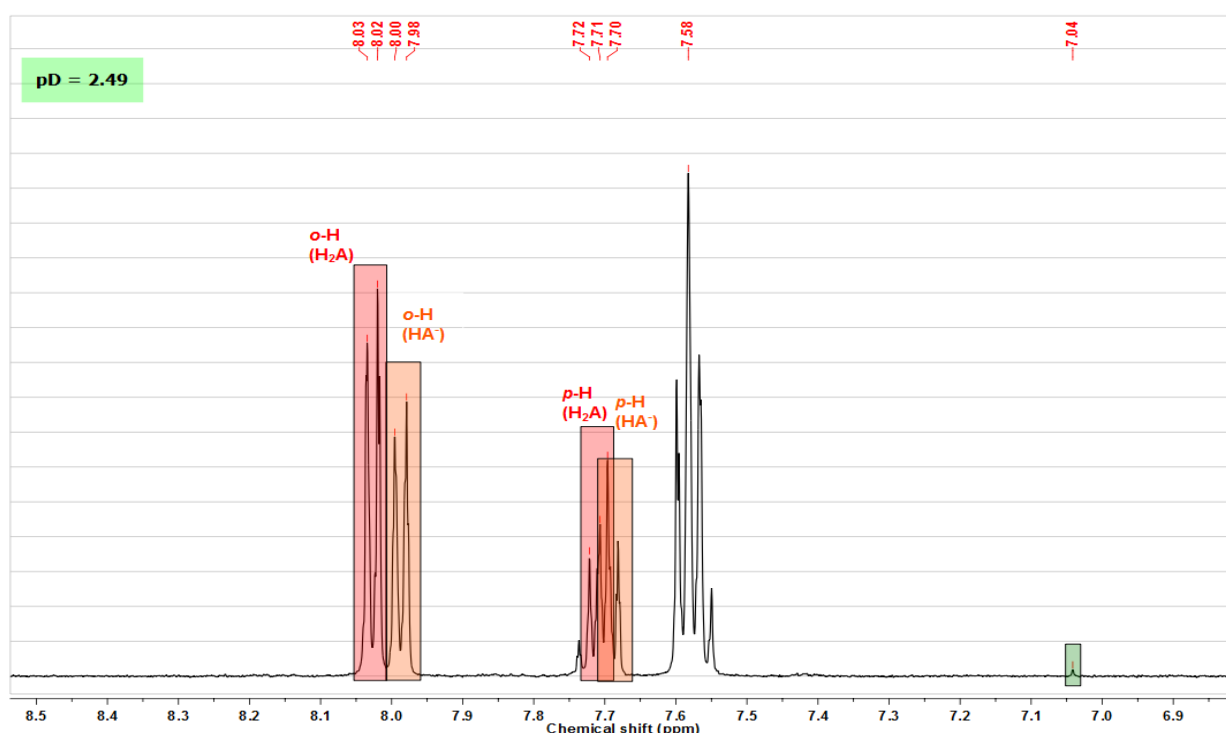


Fig. 4 A characteristic part of ^1H NMR spectrum of 4PDA in the $\text{D}_2\text{O}/\text{CD}_3\text{COOD}$ mixture, pD = 2.49; $t = 25 \pm 1$ °C; $I = 0.1$ M (NaNO_3). Highlighted are signals of *ortho*- and *para*-hydrogens of molecular (red) and monoanionic (orange) form

Two forms of 4PDA (H_2A and HA^-) were present in this solution as is expected according to experimentally obtained $\text{p}K_{\text{a}1}$ value (2.06) [20]. Separated doublets at 7.99 (H_2A) and 8.02 ppm (HA^-) correspond to the *ortho*-H atoms. The molar ratio of H_2A to HA^- in a solution ($n_{\text{H}_2\text{A}}/n_{\text{HA}^-} = 0.78$) was calculated from corresponding peak areas, and was in good agreement with the ratio ($n_{\text{H}_2\text{A}}/n_{\text{HA}^-} = 0.93$) calculated using Henderson-Hasselbalch equation. Two overlapped triplets at ~ 7.7 ppm correspond to *para*-H atoms of two ionization forms, while *meta*-H atoms (at ~ 7.6 ppm) are overlapped and could not be distinguished. The signal of H_7 atom was

observed at 7.04 ppm, but the intensity was even smaller than in CF₃COOD, as the exchange rate between H_γ and D atom in D₂O/CF₃COOD is faster than in pure CF₃COOD. Lack of corresponding signals of enol **III** and diketo form **II** confirmed that the enol **I** is the dominant form of 4PDA in a solution.

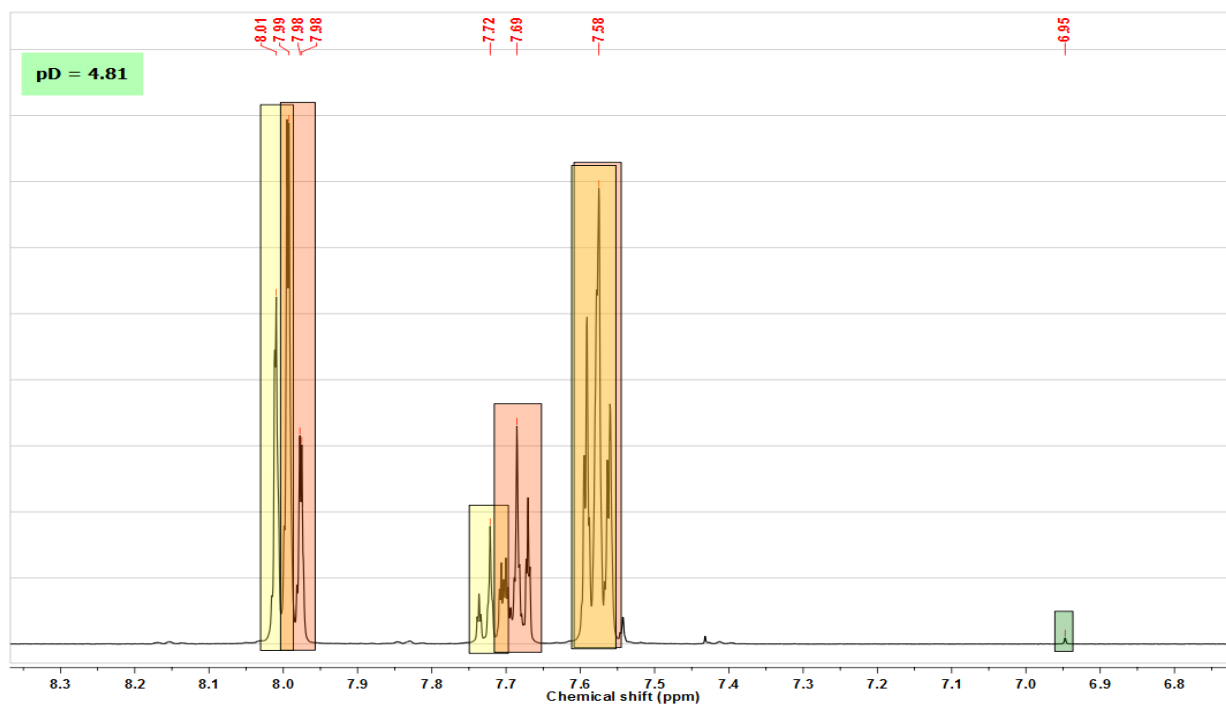


Fig. 5 A characteristic part of ¹H NMR spectrum of 4PDA in a deuterated acetate buffer, pD = 4.81; *t* = 25±1 °C; *I* = 0.1 M (NaNO₃). Highlighted (yellow and orange boxes) are signals of two forms, which could not be ascribed to different ionization forms (see the discussion below)

NMR spectra were also recorded in deuterated acetate buffer, pD 4.81 (**Fig. 5**), where 4PDA, according to p*K*_{a2} value (7.56, [20]), is expected to be present in a pure HA⁻ form. No signals of diketo tautomer were observed. Signals of *ortho*-H (*d*, 7.99 ppm), *para*-H (*t*, 7.69 ppm), and *meta*-H (*t*, 7.58 ppm) are all doubled with *meta*-H being the least resolved, indicating the coexistence of another species in a solution. Although there is a possibility for a free rotation around C_γ-C_δ bond in enol **I**, intramolecular H-bond ‘locks’ the pseudo 6-membered ring, making conformer/rotamer in which enolic oxygen on C_β and keto oxygen on C_δ are proximal to each other, a predominant one. Deprotonation of carboxyl group enables the formation of intramolecular hydrogen bond between carboxylate’s oxygen and enolic –OH group on C_β, forming a pseudo 5-membered ring. This facilitates rotation around C_γ-C_δ bond, so different conformer/rotamer, in which keto oxygen on C_δ and enolic oxygen on C_β are distal from each other, can exist in a solution. Therefore, we assume that the presence of two (stable) rotamers

around C_γ - C_δ bond causes the duplication of signals of aromatic protons in ^1H NMR spectrum. Similar explanation is given in the literature to justify the complex ^1H NMR spectrum of ADK methyl esters sodium salts. The formation and coexistence of *Z,Z*- and *E,Z*- isomers is described, as a consequence of rotation around C_β - C_γ and C_γ - C_δ bonds [33].

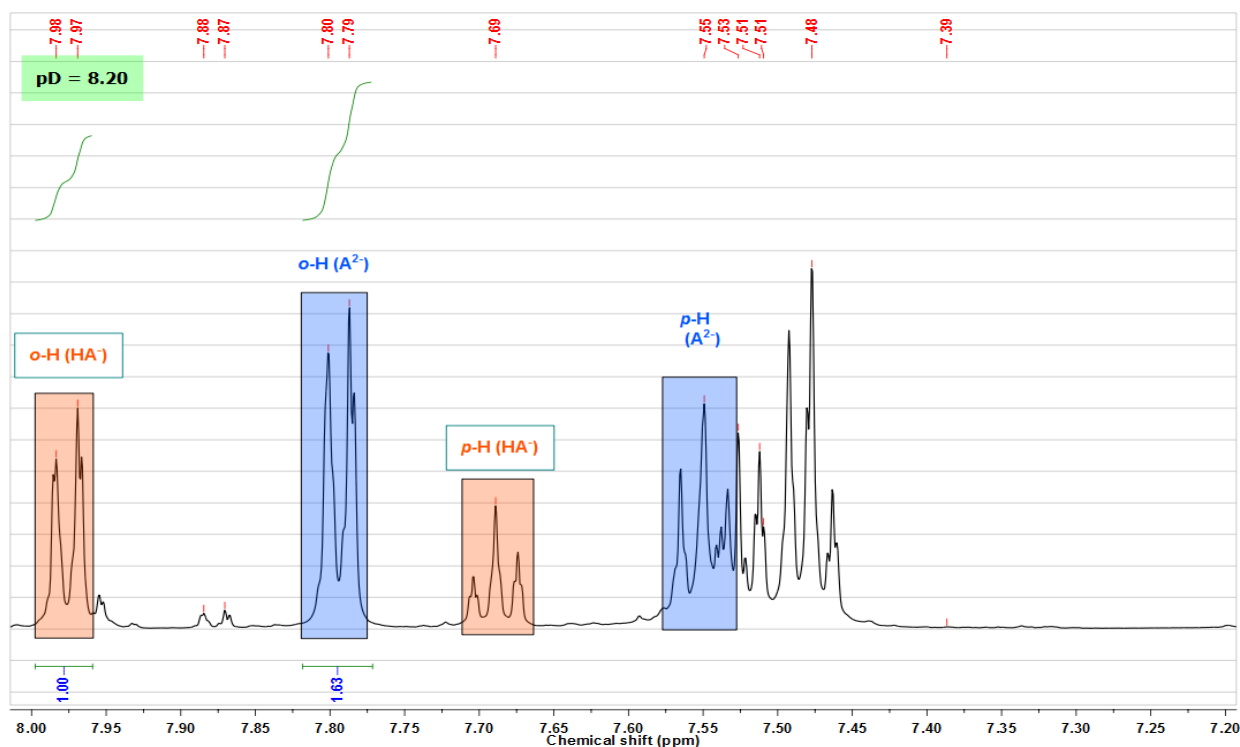


Fig. 6 A characteristic part of ^1H NMR spectrum of 4PDA in a carbonate buffer, pD = 8.20; $t = 25 \pm 1$ °C; $I = 0.1$ M (NaNO_3). Highlighted are signals of *ortho*- and *para*- hydrogens of monoanionic (orange) and dianionic (blue) form

At pD 8.20, two forms of 4PDA (HA^- and A^{2-}) are observed in a solution (**Fig. 6**), and the ratio of two forms was again in a good agreement with the one predicted from $\text{p}K_a$ values.

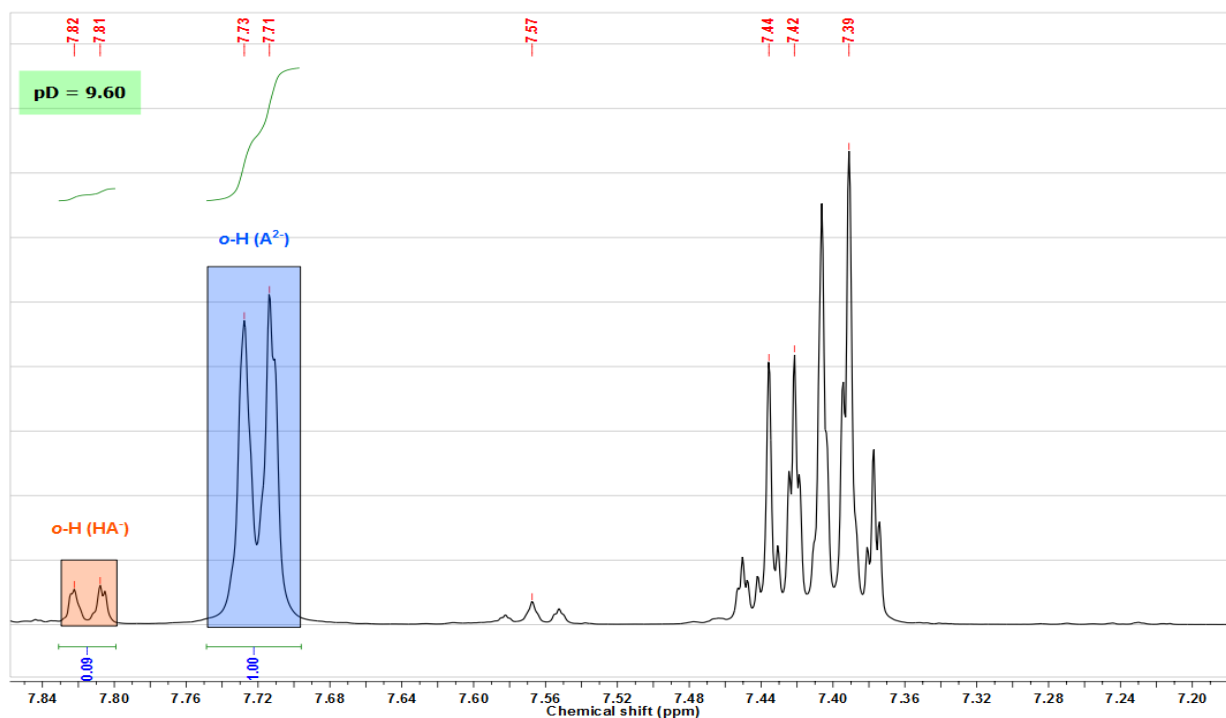


Fig. 7 A characteristic part of ^1H NMR spectrum of 4PDA in carbonate buffer, $\text{pD} = 9.60$; $t = 25 \pm 1$ °C; $I = 0.1$ M (NaNO_3). Highlighted are signals of *ortho*-hydrogens of monoanionic (orange) and dianionic (blue) form

In carbonate buffer at $\text{pD} = 9.60$ dianionic (A^{2-}) form should be the dominant species (**Fig. 7**). *Ortho*-H atoms appeared as a doublet at 7.72 ppm, and *meta*- and *para*-H atoms as overlapped triplets around 7.40 ppm. As π -electrons of dianion are delocalized over an entire keto-enol moiety, the spectral distinction between tautomers is not possible. Thus, two weak, but visible signals (*t*, $\delta = 7.57$ ppm and *d*, $\delta = 7.82$ ppm) indicated the presence of HA^- with the abundance $< 10\%$, which is in accordance with measured pD and $\text{p}K_{\text{a}2}$ value.

A summary of ^1H NMR shifts at all studied pD values, and calculated vs. predicted ratio of different protonation forms of 4PDA is given in **Table 2**.

Table 2. Summary of NMR shifts of 4PDA in buffered aqueous solutions. ^1H NMR shifts of H_2A , HA^- and A^{2-} forms are shown in red, black and blue, respectively.

pD	δ (ppm)				$\text{H}_2\text{A}/\text{HA}^-$ ratio calculated according to		$\text{HA}^-/\text{A}^{2-}$ ratio calculated according to	
	H_7	H_2	H_3	H_4	NMR spectra	pK_a values	NMR spectra	pK_a values
<0	7.23	7.96	7.50	7.64	pure H_2A	pure H_2A	/	/
2.49	7.04	8.02 (7.99)	7.58	7.72 (7.70)	0.78	0.93	/	/
4.81	6.95	7.99	7.58	7.72 (7.69)	/	pure HA^-	/	pure HA^-
8.20	/	7.98 (7.80)	7.53 (7.48)	7.69 (7.55)	/	/	0.61	0.58
9.60	/	7.82 (7.72)	7.39	7.57 (7.44)	/	/	0.09	0.03

Several observations important for understanding the solution chemistry of 4PDA should be stressed. The area under the ^1H NMR signal of enolic hydrogen atom (H_7) is less than expected since it exchanges with the ^2D from the solvent. With the increase of pD value, the intensity of H_7 decreases. This results in absence of H_7 signal in solutions with high pD values (8.20 and 9.60). As compound's carboxylic group dissociates the electron density of nucleus increases and enolic hydrogen atom becomes more shielded.

At pD 2.49, 8.20 and 9.60, 4PDA was found in solution as a mixture of two ionization forms, whose ratio of NMR peak areas correlates well with the ratio calculated from pK_a values (**Table 2**). When a mixture of two ionization forms is present in a solution, NMR signals of *ortho*-H atoms are resolved much better than the signals of *para*-atoms, while *meta*-H signals of two forms are overlapped. A transfer of negative charge from 4PDA (di)anion to aromatic ring is the most effective in *ortho*-position, because of the influence of both inductive and resonance effect. Since resonance effect is transferred only through *ortho*- and *para*-positions of the phenyl ring, *para*-H atom is more influenced by a negative charge of dioxobutanoic moiety than *meta*-H atoms.

Going from molecular to dianionic form, chemical shifts of all aromatic H atoms are moving upfield. It shows that the electron density of dioxobutanoic moiety is, to some extent, inductively transferred to aromatic ring.

At pD 4.81, we hypothesized the coexistence of two stable rotamers in a solution, and the trend in spectral resolution of aromatic H atoms remained the same as when two ionization forms are present.

3.2. A computational study of the stability of 4PDA tautomeric forms

The stability of different tautomers of 4PDA in molecular form (H_2A) was investigated using quantum chemical calculations employing MP2 Hamiltonian, and 6-31G(d,p) basis set. For all three tautomers, forms with carboxyl hydrogen oriented toward surroundings (*out* orientation) and with carboxyl hydrogen oriented inside (*in* orientation) were considered. The inclusion of implicit solvation model (H_2O) led to the stabilization of all tautomeric forms studied. Such stabilization was larger for diketo than for enol tautomers of 4PDA. Larger dipole moments of diketo comparing to enol tautomers, calculated with implicit solvation model (ESM, **Table S1**), indicate a possibility for more favorable electrostatic interactions between polar solvent and diketo tautomers.

Additional intramolecular hydrogen bond appears in the geometries with carboxyl hydrogen *in* when compared to *out* forms, and hence a possibility for the interaction of carboxyl hydrogen with a polar solvent is reduced. Although an implicit solvation model had larger influence on forms with carboxyl hydrogen oriented toward surroundings (*out* orientation), the intramolecular hydrogen bonding made forms with carboxyl hydrogen oriented *in* more stable, both in calculations with and without implicit solvent model applied. Geometries of the lowest energy forms of 4PDA with carboxyl hydrogen in both orientations are given in **Fig. 8** (a-f). Relative energies are given in respect to the apparently most stable tautomer (enol **III** *in*), for calculations in a vacuum and in implicit solvation model.

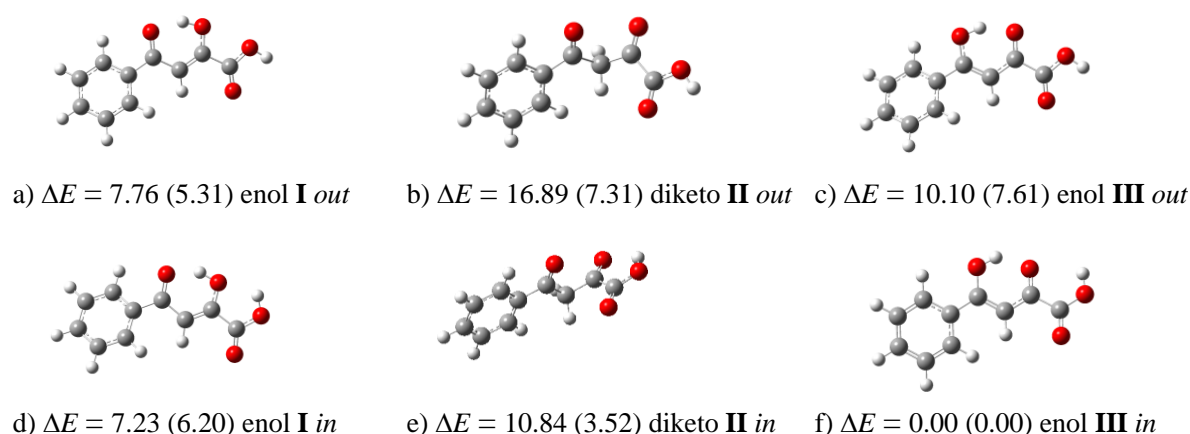


Fig. 8 Optimized structures of three 4PDA tautomers with carboxyl H *out* (a-c) and *in* orientations (d-f); calculated with MP2/6-31G(d,p). Relative energies (ΔE , in kJ/mol) *in vacuo* and in implicit solvation model of water (values in brackets) with respect to enol **III** *in* tautomer (f)

Energy differences between two enol tautomers are 7.23 kJ/mol in vacuum and 6.20 kJ/mol in implicit solvent model, for *in* orientation of carboxyl hydrogen (enol **I** *in* vs. enol **III** *in*). Diketo forms were less stable, but such difference was significantly reduced upon the inclusion of implicit solvent in calculations (from 10.84 to 3.52 kJ/mol for diketo **II** *in*). Comparing three tautomers having carboxyl H *out* orientation, the most stable form was enol **I**. Energy differences between enol **I** and enol **III** are 2.34 kJ/mol *in vacuo* and 2.30 kJ/mol when implicit solvation was accounted. Enol **I** is 9.13 kJ/mol more stable than diketo **II** form *in vacuo*, and only 2.00 kJ/mol in implicit water model.

Even though calculations suggested that forms with *in* orientation of carboxylic H were more stable than corresponding *out* forms, it was reasonable to assume the existence of stable *out* forms in a real situation, *i.e.* in a protic solvent like water. It is well known that PCM model is not good for simulating solvents with hydrogen-bonding properties. The existence of tautomer **III** *in* which was predicted as the most stable form could be questioned, since no experimental data point to the existence of tautomer **III** in detectable amount for ADKs, or structurally similar compounds. Literature data on quantum chemical calculations of L-731,988, aryl-diketo acid derivative having *N*-benzyl substituted pyrrole as aroyl moiety, describe also an analog of enol **III** as the most stable one when solvation effects were accounted using PCM model of water [17].

As implicit solvation model accounts for dipolar interactions, influencing mainly C-heteroatom and heteroatom-H bonds polarization, we added one explicit water molecule in the proximity of carboxyl group, and optimized geometry of such system for all six forms. Results are shown in **Fig. 9**.

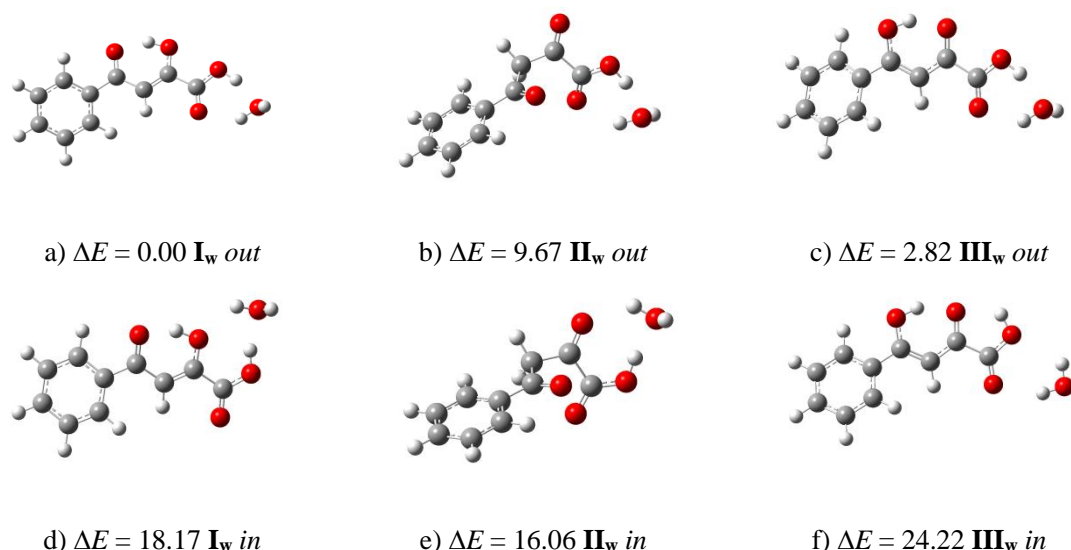


Fig. 9. Optimized structures of all tautomers of the 4PDA+water couple, calculated with MP2/6-31G(d,p). Relative energies (ΔE , in kJ/mol) in respect to the most stable form (enol **I_w out**)

The enol **I** with carboxyl H in *out* orientation + water couple (enol **I_w**) appeared as the most stable of all six forms, followed by enol **III_w** ($\Delta E = 2.82$ kJ/mol) and diketo **II_w** ($\Delta E = 9.67$ kJ/mol) forms. All tautomers with carboxyl H *in* orientation (**I_w in**, **III_w in**, and **II_w in**) appeared significantly less stable compared to *out* counterparts. Thus, an inclusion of one explicit water molecule in the model, which allowed carboxyl H to establish its (intermolecular) hydrogen-bond-donating ability, provided results in much better agreement with experiments.

Conclusions

Aryldiketo acids complexation ability with divalent metal ions, that could be responsible for their mode of biological action, and the hydrolytic C–C bond cleavage by β -ketolases depend on predominant tautomeric form. To the best of our knowledge, no experimental data on keto-enol tautomerism of ADKs in aqueous solution with $\text{pH} \leq 5.5$ was published so far. In a highly acidic medium, 4-phenyl-2,4-dioxobutanoic acid (4PDA) predominantly exists as an enolic form with keto group closer to the phenyl ring. At higher pH values, where 4PDA exists as a mixture of two species with different ionization states, NMR pattern becomes more complex. The ratio of two forms is in a good agreement with the ratio predicted from Henderson-Hasselbalch equation. An exception was found in ^1H NMR spectrum of 4PDA monoanion where signals are doubled, and the coexistence of two stable rotamers is hypothesized. *Ab initio* MP2/6-31G(d,p) method accurately predicted the relative stability of tautomers only when explicit water molecule was included in calculations.

Our combined experimental and computational evidences strongly suggest that enol I form is most prominent form in biologically relevant environments that these molecules can be exposed to. These observations could be taken into account in future efforts to develop drug-like molecules based on this class of compounds. Furthermore, a similar approach can be used to assess the prevalence of tautomers for other classes of molecules with diketo moiety. Findings from this study may be important for further development of this type of compounds as drugs, since absorption, distribution and other pharmacokinetic properties depend on a tautomeric form of a compound.

Conflict of Interest: The authors declare that they have no conflict of interest.

References:

1. Wai JS, Egbertson MS, Payne LS, Fisher TE, Embrey MW, Tran LO, Melamed JY, Langford HM, Guare JP Jr, Zhuang L, Grey VE, Vacca JP, Holloway MK, Naylor-Olsen AM, Hazuda DJ, Felock PJ, Wolfe AL, Stillmock KA, Schleif WA, Gabryelski LJ, Young SD (2000) *J Med Chem* 43:4923-4926
2. Hazuda DJ, Felock P, Witmer M, Wolfe A, Stillmock K, Grobler JA, Espeseth A, Gabryelski L, Schleif W, Blau C, Miller MD (2000) *Science* 287:646-650
3. Tan JJ, Liu C, Sun HX, Cong JX, Hu ML, Wang XC, Liang JX (2012) *Mini Rev Med Chem* 12:875-889
4. Cuzzucoli Crucitti G, Métifiot M, Pescatori L, Messori A, Madia VN, Pupo G, Saccoliti F, Scipione L, Tortorella S, Esposito F, Corona A, Cadeddu M, Marchand C, Pommier Y, Tramontano E, Costi R, Di Santo R (2015) *J Med Chem* 58:1915-1928
5. Tomassini J, Selnick H, Davies ME, Armstrong ME, Baldwin J, Bourgeois M, Hastings J, Hazuda D, Lewis J, McClements W (1994) *Antimicrob Agents Chemother* 38:2827-2837
6. Drakulić BJ, Stavri M, Gibbons S, Žižak ŽS, Verbić TŽ, Juranić IO, Zloh M (2009) *ChemMedChem* 4:1971-1975
7. Milletti F, Vulpetti A (2010) *J Chem Inf Model* 50:1062-1074
8. Sechi M, Bacchi A, Carcelli M, Compari C, Duce E, Fisicaro E, Rogolino D, Gates P, Derudas M, Al-Mawsawi LQ, Neamati N (2006) *J Med Chem* 49:4248-4260
9. Billamboz M, Bailly F, Barreca ML, De Luca L, Mouscadet JF, Calmels C, Andréola ML,

- Witvrouw M, Christ F, Debyser Z, Cotelle P (2008) *J Med Chem* 51:7717-7730
10. Koz'minykh VO, Koz'minykh EN (2004) *Pharm Chem J* 38:67-77
 11. Reguera J, Weber F, Cusack S (2010) *PLoS Pathog* 6:e1001101
 12. DuBois RM, Slavish PJ, Baughman BM, Yun M-K, Bao J, Webby RJ, Webb TR, White SW (2012) *PLoS Pathog* 8:e1002830
 13. Krieger IV, Freundlich JS, Gawandi VB, Roberts JP, Gawandi VB, Sun Q, Owen JL, Fraile MT, Huss SI, Lavandera JL, Ioerger TR, Sacchettini JC (2012) *Chem Biol* 19:1556-1567
 14. Zhu W, Zhang Y, Sinko W, Hensler ME, Olson J, Molohon KJ, Lindert S, Cao R, Li K, Wang K, Wang Y, Liu YL, Sankovsky A, de Oliveira CA, Mitchell DA, Nizet V, McCammon JA, Oldfield E (2013) *Proc Natl Acad Sci USA* 110:123-128
 15. Brecker L, Pogorevc M, Griengl H, Steiner W, Kappe T, Ribbons DW (1999) *New J Chem* 23:437-446
 16. Sechi M, Derudas M, Dallochio R, Dessì A, Bacchi A, Sannia L, Carta F, Palomba M, Ragab O, Chan C, Shoemaker R, Sei S, Dayam R, Neamati N (2004) *J Med Chem* 47:5298-5310
 17. Huang M, Richards WG, Grant GH (2005) *J Phys Chem A* 109:5198-5202
 18. Cvijetić IN, Verbić TŽ, Drakulić BJ, Stanković DM, Juranić IO, Manojlović DD, Zloh M (2017) *J Serb Chem Soc* 82:303-316
 19. Verbić TŽ, Drakulić BJ, Zloh M, Juranić IO (2008) *Lett Org Chem* 5:692-699
 20. Verbić TŽ, Drakulić BJ, Zloh MF, Pecelj JR, Popović GV, Juranić IO (2007) *J Serb Chem Soc* 72:1201-1216
 21. Kojić M, Petković M, Etinski M (2016) *Phys Chem Chem Phys* 18:22168-22178
 22. Kojić M, Petković M, Etinski M (2016) *J Serbian Chem Soc* 81:1393-1406
 23. Belova NV, Oberhammer H, Girichev GV (2004) *J Phys Chem A* 108:3593-3597
 24. Moriyasu M, Kato A, Hashimoto Y (1986) *J Chem Soc Perkin Trans 2* 515-520
 25. Yamabe S, Tsuchida N, Miyajima K (2004) *J Phys Chem A* 108:2750-2757
 26. Sigalov MV (2015) *J Phys Chem A* 119:1404-1414
 27. Glasoe PK, Long FA (1960) *J Phys Chem* 64:188-189

28. Wolinski K, Hinton JF, Pulay P (1990) *J Am Chem Soc* 112:8251-8260
29. MOPAC2016, Stewart JJP, Stewart Computational Chemistry, Colorado Springs, CO, USA, [HTTP://OpenMOPAC.net](http://OpenMOPAC.net) (2016).
30. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann, RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas Ö, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2009) *Gaussian 09*, Revis. B.01, Gaussian, Inc., Wallingford CT
31. Guthrie JP (1972) *J Am Chem Soc* 94:7020-7024
32. Maurin C, Bailly F, Buisine E, Vezin H, Mbemba G, Mouscadet JF, Cotelle P (2004) *J Med Chem* 47:5583-5586
33. Maurin C, Bailly F, Cotelle P (2004) *Tetrahedron* 60:6479-6486