Revised, Nov 2009

DFT and isodesmic reaction based prediction of four stepwise protonation constants, as log $K_H^{(n)}$, for Nitrilotriacetic Acid (NTA). The importance of a kind and protonated form of a reference molecule used.

Krishna K. Govender and Ignacy Cukrowski[*]

Department of Chemistry, Faculty of Natural Sciences, University of Pretoria, Lynnwood Road, Hillcrest, Pretoria 0002, South Africa

Email: ignacy.cukrowski@up.ac.za

Keyword: DFT; Isodesmic reaction; Thermodynamic cycle; PCM, Protonation constant; Dissociation constant; NTA, IDA, MIDA, EIDA, PIDA, HIDA.

Abstract

An explicit application of isodesmic reaction (a proton exchange between the studied and similar in structure reference molecule), where the free energy change of the protonation reaction in water was obtained using the free energies in solution from a single continuum model, was used to predict stepwise protonation constants of nitrilotriacetic acid. Calculations were performed at the RB3LYP/6-311+G(d,p) level of theory in conjunction with PCM-UA0 solvation model. Five reference molecules were investigated. It has been established that one must pay a special attention to structural similarities between the studied and reference molecules and selection of a protonated form of the reference molecule. The protonation reactions in which the studied and reference molecule are involved in must be (if possible) of the same order; e.g. first (or generally *n*th) protonation reaction of the reference molecule must be used to compute the first (or *n*th) protonation constant of studied molecule. The lowest energy conformer must always be used. The first, second, third and fourth computed protonation constants differed, on average, from experimental values by 3.3, 0.8, 0.2 and 0.2 log units, respectively. It appears that the charge on the reference molecule has more decisive influence on accuracy of computed protonation constants than its structural differences when compared with the studied molecule. Results reported can be used as guide in constructing isodesmic reactions useful for the theoretical prediction of protonation constants by use of methodology described in this work.

1. Introduction

Knowledge of protonation, K_H, and dissociation, K_a, constants is of special interest to many chemists and life scientists[1] as they constitute important thermodynamic property of a compound that might be of either biological, medicinal, or industrial (just to mention few) importance. Although a number of experimental techniques has been developed to measure protonation/dissociation constants under various experimental conditions, many of the chemical species are not easily amenable to a full experimental characterization.[2] A number of papers has been reported[1-54] on theoretical prediction of dissociation constants. Most of them employed thermodynamic cycles (TC) to compute the free energies of dissociation reaction. Often, high-level theories were used in the gas-phase calculations where they are known to be accurate. The solution-phase calculations were used to provide the solvation energies (G_{sol}); usually low-level continuum models were employed for the purpose. When the above protocol is used, the absolute pK_a value is obtained. To avoid uncertainties related to the solvation energies of either H^+ or H_3O^+ ions, isodesmic reaction (IRn) was incorporated within TC;[1,3–7] this protocol of calculation results in relative[3] p*K*a values. Results reported to date predominantly describe the calculations of singly charged molecules, either anions[1-29] (a study of doubly charged anions is very rare), or cations.[30-35] This is most likely due to the fact that (i) it is very difficult for DFT methods to properly describe anions (with multiple negative charges) in gas phase because in absence of an external stabilization of the charge (e.g. solvent) DFT methods have a bias towards "over-delocalization" of the charge (one might observe bonds that are longer than expected and significant reduction on the HOMO-LUMO gap),[55] and (ii) inaccurate computational evaluations of ionic solvation free energies for highly charged anions; these energies are highly dependent on the solvation model used due to different models chosen to generate the 'best' electrostatic cavity.[56] Accuracies achieved thus far for computed dissociation constants (for a singly dissociable organic acids) are often within ±1.0 log unit, on average, when compared with experimentally available values, but differences of several log units are not uncommon.[3–5,56]

Recently we reported the DFT-predicted four stepwise protonation constants, expressed as log $K_H^{(n)}$, for a highly charged molecule nitrilotripropanoic acid (NTPA).[54] An explicit application of an isodesmic reaction involving two and similar in structure ligands, where the free energy change of the protonation reaction in solution was obtained using the free energies in solution from a single continuum model, resulted in the average difference between predicted and experimental stepwise protonation constants being ± 0.5 log unit.

This suggested that in principle, even though serious concerns were expressed,[55,56] accurate determination of stepwise protonation constants for highly negatively charged molecules is possible.

In this paper our focus is on parameters that influence accuracy in predicting four consecutive protonation constants when IRn-based procedure, as reported by us recently,[54] is employed. It is important to investigate a wide range of poly-charged compounds (with negative and positive charges) in order to establish (i) whether implementation of the protocol can indeed produce consistently good predictions, (ii) how significant the selection of a reference molecule is from the point of view of its structural similarity to the studied compound, (iii) to what degree the selection of a conformer impacts on accuracy in computed protonation constants, (iv) does the selection of a different protonated form of the reference molecule have an influence on accuracy of computed protonation constants, (v) which one, structural similarity, or the charge on a protonated form of a reference molecule, plays a more important role when IRn-based methodology described in this work is employed, (vi) is it possible to predict protonation constants in correct order, as determined from an experiment, (vii) to what extend the selection of level of theory and a basis set has an influence on accuracy in computed protonation constants, and (viii) how significant is the kind of solvation model employed. These are just a few important questions and this work will address only some of them.

We have chosen the ligand nitrilotriacetic acid (NTA) because it is an important derivative of glycine that is widely studied due to its excellent chelating abilities.[57] This is a ligand that has enjoyed numerous applications in medicine,[58,59] biochemistry[58,59] and industry.[60–62] In medicinal and biological studies it was shown that aliphatic amine salts of NTA inhibit the growth of bacteria and fungi and have herbicidal activity.[63] NTA has also been used as a transient phytoextraction agent that combines high biodegradability and low phytotoxicity with chelating strength.[64]

The application of IRn requires a reference molecule and its selection appears to be crucial. Molecules shown in Figure 1 (iminodiacetic acid (IDA), methyliminodiacetic acid (MIDA), ethyliminodiacetic acid (EIDA), propyliminodiacetic acid (PIDA), and hydroxyethyliminodiacetic acid (HIDA)) were used here as they have many structural similarities with the compound of interest (NTA) and their experimental stepwise protonation constants are well known.[57]

2. Computational Details

All calculations were performed using GAUSSIAN 03, revision D.01,[65] on a 64-bit Linux workstation in parallel environment (Opensuse 10.3). Molecular visualizations were accomplished with the aid of GaussView 4.[66] Since it is of paramount importance to include diffuse functions for anions,[19] both gas-phase and solvent (water, ε = 78.39) optimizations were performed at the RB3LYP level of theory[67] in conjunction with a 6-311+G(d,p) basis set. Full solvent-optimization was performed with the default solvation model provided by Gaussian, i.e. Tomasi's Polarized Continuum Model (PCM),[68–70] and UA0 radii (United Atom Topological Model). We have chosen this model because it generated acceptable results in the prediction of log $K_H^{(n)}$ values for NTPA.[54] Single point calculations (SPCs) were carried out in solvent at the RB3LYP and HF levels of theory in conjunction with the 6-311+G(d,p) basis set on the gas-phase structures as well as structures fully optimized in solvent, using the PCM-UAHF and CPCM-UAHF[71,72] (polarizable conductor model in combination with the United Atom for Hartree-Fock radii) solvation models. With CPCM, the solute cavities are modeled on the optimized molecular shape, and include both electrostatic and non-electrostatic contributions to energies.[10] The HF level of theory also was used for single point calculations since the UAHF radii were optimized for HF.[19]

Full geometry optimization of all protonated forms of the reference molecules, seen in Figure 1, was carried out in solvent using the same procedure as for NTA; there was no need to perform single point calculations on these molecules. Frequency calculations were also performed, along with the geometry optimization, to ensure that each of the optimized structures did not lie at a saddle point (no imaginary frequencies were present in all structures reported here).

3. Results and Discussion

Structural considerations. In the Cambridge Structural Database (CSD)[73] there are only two crystallographic structures of NTA[74,75] and they both are of the H_3L form of the ligand; all crystallographic structures are marked with asterisks. Even though H_3L^* has no overall charge, it has two charged centers with opposite polarities, the positive one on the protonated N-atom and negative one on the de-protonated $–COO^–$ group – see Figure 2. The remaining available reported crystal structures, that of the H_2L^* form of IDA and MIDA, as well as H_3L^* form of HIDA, are shown in Figure S1, Supporting Information. Unfortunately, no crystallographic data was found for EIDA and PIDA. From visual analysis of available structures one might conclude that there are two distinctive types of

arrangements of arms containing the –COOH and $–COO^-$ groups, namely (i) all arms are bent towards the central and protonated N-atom resulting in the formation of H-bonds between carboxylic oxygen and H-atom on nitrogen (the H_2L^* form of MIDA and H_3L^* form of HIDA), and (ii) all arms are placed as described above except one that is bent outwards the central and protonated N-atom (the H_3L^* form of NTA and H_2L^* form of IDA); because of that this arm is not involved in the formation of the intramolecular H-bond. The analysis of packing in the available crystal structures provided appropriate explanation for that, somewhat unexpected, placement of one of the arms – see Figure 3 as an example. It is clear that one of the NTA arms is involved in very strong intermolecular H-bond characterized by the shortest bond distance (1.603 Å) among all inter- and intramolecular H-bonds observed. In the case of HIDA (see Figure S2, Supporting Information) all the intramolecular H-bonds are preserved even though O-atoms are involved in extremely strong interactions with other molecules; this is most likely due to specific, in this case, crystal packing. We came to the conclusion that in solvent the attractive interactions between H-atom on the central N-atom and carboxylic O-atoms must prevail and have constructed an additional H_3L form of the NTA molecule with all three arms bent towards the central N-atom. To maintain structural (conformational) consistency within a full set of necessary protonated forms of the ligand NTA, the remaining 4 possible forms, namely H_2L^-, HL^{2-}, L^{3-} (fully deprotonated ligand), and H_4L^+ (fully protonated form of the ligand) were constructed as follows. We started with the energy-optimized solvent structure of H_3L. The H_2L^- form was generated by removing a dissociable proton from a –COOH group and H_4L^+ was generated by adding a proton to the remaining $–COO^-$ group. A similar procedure was followed to generate HL^{2-} and L^{3-}, where proton was removed from energy-optimized solvent structure to generate the product of stepwise dissociation reaction. The same procedure was applied to generate all the protonated forms of the reference molecules seen in Figure 1.

The computed structural matrix of the solvent-optimized H_3L^* and H_3L of NTA, together with the data available from the CSD,[73] is given in Table S1, Supporting Information; the relevant data obtained for IDA, MIDA, and HIDA are provided in Tables S2-S4 (Supporting Information), where numbering of atoms is the same as seen in Figure S1, Supporting Information.

All solvent-optimized protonated forms of NTA, including the crystallographic structure H_3L^*, are shown in Figure 4 (those of the reference molecules are provided in Figures S3-S7, Supporting Information). All protonated forms of the ligand NTA have

considerably strong intramolecular H-bonds between oxygen on the $–COO^-$ or –COOH groups and a proton on the N-atom and they vary in length between about 2.0 and 2.35 Å. Interestingly, the shortest H-bonds, of 1.985 and 1.960 Å, were found in H_3L and H_3L^*, respectively. As discussed above, the self-constructed and energy-minimized H_3L molecule differs significantly from H_3L^* and it was of importance to find out which one will generate better estimates in the computed protonation constants.

Preliminary Investigations. Initially, we have used TCs to compute absolute protonation constants. A detailed description of TCs used and results obtained are provided in Supporting Information (see Tables S5-S7). Unfortunately, the methodology based on TC principles could not be fully applied in our studies due to significant structural differences of molecules in gas phase and solvent. Migration of proton from the central N-atom to the $–COO^-$ group took place when optimization was performed in the gas phase; similar observation was also reported elsewhere.[13,54]

We have also tested whether higher level of theory could result in preserving the solvent structure. The H_3L form of NTA was subjected to the full gas-phase optimization at the RMP2/6-311+G(d,p) level of theory. Unfortunately, the proton migrated again from N-atom to carboxylic group. Clearly, even with a higher level of theory that generates accurate gas-phase free energies, we were unable to apply commonly used TCs in the study of molecules considered in this work.

Isodesmic Reaction. To date, isodesmic reaction principles (where two structurally similar compounds were used, investigated and reference molecule) have been extensively utilized in the prediction of enthalpies of formation.[76-84] In some cases a reference molecule has been incorporated within TCs[1,3-7] to eliminate uncertainties related to either H^+ or H_3O^+ ions. Dissociation constants of a number of compounds with a single dissociable proton have also been computed directly in solvent from ΔG_{diss} for the dissociation reaction $HL + H_2O = L^- + H_3O^+$.[23]

The implementation of IRn has an advantage, when the total free energies in solution from a single continuum model calculation are used, because it should minimize (or systematically eliminate) errors related to the solvation models used provided that the same level of theory, basis set and solvation model are used for each component involved in the reaction of interest.[56] The main challenge associated with the use of IRn, however, is the selection of appropriate reference molecule.[56,80] The ligand NTA can be seen as a set of 5 molecules that differ in (i) a number of protons (from 0 to 4), and (ii) charges on the molecule (from +1 to –3). Clearly, a careful selection must be made to find the most

appropriate protonated form of the reference molecule that has to be included in each of IRn needed to compute four protonation constants of NTA. By considering the structural properties of NTA (called further $L_{(1)}$) we opted for IDA, MIDA, EIDA, PIDA and HIDA as reference compounds (called further $L_{(2)}$) because each of them has two acetate groups (there are three in NTA) and the same kind of electron donor atoms ($–COO^-$ and R_3N:) that can be protonated in a solution. In addition, protonation constants for all of the chosen reference molecules are well known as these ligands are widely-studied.[57]

IRn employed here can be seen as a competition reaction between two ligands for a proton (proton transfer reaction) and for the first protonation constant of NTA it can be written as

$$L_{(1)}(aq) + HL_{(2)}(aq) = HL_{(1)}(aq) + L_{(2)}(aq) \qquad \Delta G_{IRn}(aq) \qquad (1)$$

In order to investigate the impact the kind and protonated form of the reference molecule has on the theoretically generated protonation constants of NTA, a large number of isodesmic reactions was tested. Here, each reference molecule has three protonation constants (NTA has four), hence for each pair of ligands ($L_{(1)}$ and $L_{(2)}$) 12 isodesmic reactions (such as eq 1, but involving different protonated forms of the ligands) had to be considered. For simplicity, only first protonation reaction (PRn) in which each of the two ligands (NTA and a reference molecule) is involved in is shown as eqs 2 and 3

$$H^+ + L_{(1)} \leftrightarrow HL_{(1)} \qquad \Delta G_{PRn}(aq)_{L(1)} \qquad (2)$$

$$H^+ + L_{(2)} \leftrightarrow HL_{(2)} \qquad \Delta G_{PRn}(aq)_{L(2)} \qquad (3)$$

A protonation reaction is the reverse of a weak acid dissociation reaction (DRn) and in the case of stepwise reactions the following holds

$$\log K_H^{(k)} = \log \frac{1}{K_a^{(n)}} = pK_a^{(n)} \qquad (4)$$

where $k = 1 + m – n$, m and n represent the highest dissociation constant (here $m = 4$) and an nth consecutive dissociation constant ($1 \le n \le m$), respectively, and k applies to a kth consecutive protonation constant, $1 \le k \le m$. Note that the ligand NTA has three acidic groups and only three dissociation constants would be reported when, for example, the TC-

based methodology was employed. However, due to the protonation/de-protonation of N-atom in NTA, it is of paramount importance to consider also the first protonation constant, log $K_H^{(1)}$. From this it follows that the fourth dissociation constant of NTA is linked through eq 4 with the first protonation constant of this ligand. In case of the reference molecule, the third dissociation constant is linked through eq 4 with the first protonation constant of this reference molecule. (The above is provided here for convenience and to assure clarity in nomenclature used in this work). The relationships between the change in Gibbs energies for protonation and dissociation reactions applicable to eqs 2 and 3 can be written as

$$\Delta G_{PRn}(aq)^{(1)}_{L(1)} = -\Delta G_{DRn}(aq)^{(4)}_{L(1)} \quad (5)$$

$$\Delta G_{PRn}(aq)^{(1)}_{L(2)} = -\Delta G_{DRn}(aq)^{(3)}_{L(2)} \quad (6)$$

where $\Delta G_{PRn}(aq)^{(k)}_{L(1)}$ and $\Delta G_{DRn}(aq)^{(n)}_{L(1)}$ refers to the kth (here k = 1) stepwise protonation reaction and relevant nth (here n = 4 for NTA) stepwise dissociation reaction, respectively (the same applies to the reference molecule $L_{(2)}$, but n = 3).

The change in Gibbs energies for each PRn, eqs 2 and 3, can be written, respectively, as

$$\Delta G_{\mathrm{PRn}}(\mathrm{aq})_{\mathrm{L}(1)} = G_{\mathrm{aq}}(\mathrm{HL}_{(1)}) - G_{\mathrm{aq}}(\mathrm{H}^+) - G_{\mathrm{aq}}(\mathrm{L}_{(1)}) \quad (7)$$

$$\Delta G_{\mathrm{PRn}}(\mathrm{aq})_{\mathrm{L}(2)} = G_{\mathrm{aq}}(\mathrm{HL}_{(2)}) - G_{\mathrm{aq}}(\mathrm{H}^+) - G_{\mathrm{aq}}(\mathrm{L}_{(2)}) \quad (8)$$

The isodesmic reaction of interest (eq 1) can be obtained from subtracting eq 3 from 2, and hence from subtracting eq 8 from 7 one obtains expressions for the change in Gibbs energy applicable to this isodesmic reaction

$$\Delta G_{\mathrm{IRn}}(\mathrm{aq}) = \Delta G_{\mathrm{PRn}}(\mathrm{aq})_{\mathrm{L}(1)} - \Delta G_{\mathrm{PRn}}(\mathrm{aq})_{\mathrm{L}(2)} \quad (9)$$

$$\Delta G_{\mathrm{IRn}}(\mathrm{aq}) = G_{\mathrm{aq}}(\mathrm{HL}_{(1)}) - G_{\mathrm{aq}}(\mathrm{L}_{(1)}) - G_{\mathrm{aq}}(\mathrm{HL}_{(2)}) + G_{\mathrm{aq}}(\mathrm{L}_{(2)}) \quad (10)$$

where the uncertainty related to $G_{\mathrm{aq}}(\mathrm{H}^+)$ is no longer applicable as this term cancels of (this eliminates any error that might have been introduced by the use of an experimental value for this quantity).

Eq 10 was used to calculate $\Delta G_{\mathrm{IRn}}(\mathrm{aq})$ of IRn (eq 1) from appropriate Gibbs energies obtained for relevant and fully solvent-optimized structures of the ligand NTA and

reference molecule $L_{(2)}$. Table S8 in Supporting Information provides the ZPVE-corrected minimum energies, E_{min}, as well as the Gibbs free energies of NTA and all of the reference molecules studied here. The value for $\Delta G_{PRn}(aq)_{L(2)}$ was obtained from well-known relationship

$$\Delta G(aq) = -RT \ln K \qquad (11)$$

using the reported protonation constants[57] (at 20 and 25 ^{o}C, μ = 0.0 and 0.1 M) of the reference molecule $L_{(2)}$. Once $\Delta G_{IRn}(aq)$ and $\Delta G_{PRn}(aq)_{L(2)}$ have been calculated, the value of $\Delta G_{PRn}(aq)_{L(1)}$, which is needed to calculate the protonation constants of NTA from eq 11, was obtained from eq 9.

Table 1 provides values for the functions required to calculate protonation constants, calculated and experimental protonation constants of NTA, along with differences between calculated and experimental protonation constants (δ). Values for $\Delta G_{PRn}(aq)^{(k)}_{L(2)}$ were calculated from the experimentally available stepwise protonation constants,[57] which have been reproduced in Table S9, Supporting Information. Only isodesmic reactions that produced the best results are shown in Table 1; the remaining results are provided in Table S10, Supporting Information. There are several interesting observations one can make from the analysis of data seen in Table 1:

(i) All isodesmic reactions seen in Table 1 predicted four protonation constants in correct order, $\log K_H^{(1)} > \log K_H^{(2)} > \log K_H^{(3)} > \log K_H^{(4)}$, as observed from an experiment. One must realize that the experimental values of the second, third, and forth protonation constants of NTA differ between each other only by one log unit (or less) and this is a typical error reported in theoretically predicted values of dissociation constants reported to date for organic acids containing only one carboxylic group. Clearly, results obtained here should be seen as satisfactory and encouraging further studies aimed at improvements in accuracy of theoretical predictions.

(ii) As one goes from IDA to HIDA, it is seen that the prediction of protonation constants becomes more accurate, with HIDA yielding the most accurate estimates. The observed trend can be linked with an increase in structural similarity between the reference molecule and NTA.

(iii) Regardless of the reference molecule studied here, the use of HL^{-} and L^{2-} of $L_{(2)}$ resulted in the best estimates of the first protonation constant of NTA that involves L^{3-} and

HL^{2-} forms of NTA. This can be generalized: computing the first protonation constant of studied molecule requires the first protonation constant and components involved in the first protonation reaction of a reference molecule.

(iv) The best estimates in the second and third theoretical protonation constants of NTA were always obtained (regardless of the reference molecule involved in IRn) when the second protonation constant of the reference molecule was used.

(v) The best prediction of the forth (highest) protonation constant of NTA always required the highest (third) protonation constant of the reference molecule. Here again one might generalize that to calculate the highest protonation constant of the studied molecule one must involve either the same or as high as possible protonation constant of the reference molecule.

(vi) The least accurate computed log K_H, regardless of the reference molecule used, was always obtained for the first protonation reaction of NTA that involved the most negatively charged forms, that of the studied ligand NTA, $L_{(1)}^{3-}$, and the reference molecule $L_{(2)}^{2-}$.

(vii) Interestingly, for all reference molecules studied, the smallest errors in the predicted log K_H values were obtained for the second and third protonation constants of NTA and all of them might be regarded as acceptable estimates at μ = 0.0 M. One might rationalize this observation in terms of a charge placed on the studied and reference molecules. In these isodesmic reactions, the HL^{2-}, H_2L^-, and H_3L forms of NTA as well as HL^- and H_2L forms of the reference molecules were used where charge varied from –2 to zero. Since good predictions in log $K_H^{(2)}$ and log $K_H^{(3)}$ of NTA were obtained, regardless of the reference molecule studied here, one might conclude that (a) the more similar charges on the studied and reference molecule are, the better prediction is achieved, even though significant differences in structures of the reference molecules are present, and (b) the charge on a reference molecule has more significant impact than its structural similarity to NTA, when accuracy in predicted protonation constants is considered.

(viii) Unexpectedly, somewhat worse estimates in log $K_H^{(4)}$ values were obtained when molecules involved in the isodesmic reactions had 0 and +1 charge; this applies to all systems studied here. This might be a significant observation because many important ligands (among them macrocyclic ligands) do not have carboxylic groups (they are neutral in their fully deprotonated form) and when protonated they have multiple and positive charges on them. Clearly, this requires a dedicated investigation in order to establish whether IRn-based protocol described here can be applied successfully for positively charged molecules.

If one considers the structural features of each of the reference molecule and that of NTA, it is possible to conclude that the molecule which is structurally most similar to NTA is HIDA. Both NTA and HIDA have one N-donor atom (R_3-N:) and three O-donor atoms with one of them being part of the –OH group (in HIDA) instead of –COOH group (in NTA). All the other reference molecules have only two O-atoms. The structural similarity of HIDA to NTA correlates well with results seen in Table 1.

It also appears that the cavity of the reference molecule, when full energy optimization is performed in solvent, plays a significant role. The values of δ for the first protonation constant of NTA (at μ = 0.0 M) were 4.81, 3.59, 3.05, and 3.08 log unit when IDA, MIDA, EIDA, and PIDA were used as the reference molecule, respectively. The same trend, the decrease in error with an increase in the cavity of the reference molecule, is seen for the forth protonation constant, namely –2.19, –1.63, –0.99, and –0.70 log unit, respectively, for the same reference molecules. Also, careful attention needs to be paid to the positioning and presence of atoms, especially heteroatom, as the additional –OH group present in HIDA, which is not present in IDA, MIDA, EIDA or PIDA, seems to make a huge difference, as far as prediction of protonation constants is concerned.

Additional test. All four theoretically predicted protonation constants of NTA seen in Table 1 differ from experimental values by less than a single log unit when HIDA was used; results of this accuracy are often referred to as excellent in the literature. On the other hand, the absolute protonation constants obtained from TCs carried significantly larger errors (see Table S7, Supporting Information). One might argue that accuracy obtained from straight continuum model free energies in solution calculation is not directly comparable with that obtained from TC as totally different chemical reactions are involved.

Since some solvent structures were not preserved in the gas phase, we performed a single point frequency calculation in gas phase on the solvent-optimized structures. We selected IDA for the test as inclusion of this reference molecule in the proton competition reaction (IRn) resulted in the worst results, hence one might expect significant improvement in protonation constants calculations. The test was performed for the reaction

$$H_3L_{(1)} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L_{(2)} \qquad (12)$$

where $L_{(1)}$ and $L_{(2)}$ are NTA and the reference molecule IDA, respectively. Our aim here was to find out whether improvement of the G_{gas} component to a better (higher) level of theory improves the computed protonation constant when TC-based protocol is

implemented. The following protocol was implemented. Let G_{sol} stand for the Gibbs free energy in solution obtained from a single continuum model calculation at B3LYP/6-311+G(d,p) in conjunction with PCM/UA0 solvation model. For each component involved in the above IRn (eq 12) one can compute the free energy of solvation from $\Delta G_{sol} = G_{sol} - G_{gas}^{(1)}$ where $G_{gas}^{(1)}$ is computed as a single point on the solution-phase geometries at the same B3LYP/6-311+G(d,p) level of theory. Let $G_{gas}^{(2)}$ stand for the improved estimate obtained at the higher level of theory, in this case also calculated as a single point on the solution-phase geometry, then G_{sol}(TC), the improved Gibbs free energy obtained from thermodynamic cycle calculations, can be expressed as $G_{sol}(TC) = G_{gas}^{(2)} + \Delta G_{sol}$. The values of $G_{gas}^{(2)}$ obtained at the RMP2/6-311+G(d,p) level of theory along with other necessary components are presented in Table 2. From G_{sol}(TC) values we calculated the change in the Gibbs free energy for the reaction 12, $\Delta G_{sol}(TC) = 6.137$ kcal mol^{-1}, followed by the forth protonation constant of NTA, $\log K_H^{(4)} = -4.50$. This is much worse estimate when compared with the value of –1.19 obtained directly from implementation of IRn (see Table 1).

An attempt was made to obtain $G_{gas}^{(2)}$ values using the G3/6-311+G(d,p) level of theory for the single point frequency calculations in gas phase on the structures optimized at the B3LYP/6-311+G(d,p) level of theory in conjunction with the PCM/UA0 solvation model. Unfortunately, the solvent structures of $H_3L_{(1)}$ and $H_2L_{(2)}$ have not been preserved. The H-atom has moved away from nitrogen and protonated the -COO^- group – see Figure S8, Supporting Information.

Conformational considerations. All protonated forms of NTA and the reference molecules were subjected to the Schrödinger's Maestro[85] conformational analysis in solvent. Generated, based on Molecular Mechanics/Molecular Dynamics (MM/MD) principles, lowest energy conformers C-1 are shown in Figures S9-S14, Supporting Information. Table 3(a) provides energies $E_{C\text{-}1}$ to $E_{C\text{-}5}$ (in kJ/mol) for five lowest in energy conformers of all the protonated forms of NTA. MM/MD-based SPC in solvent was also performed on the NTA DFT-structures shown in Figure 4; results are shown as E_{SPC} in Table 3a. It is seen that all E_{SPC} are larger than energies of the MM/MD-generated conformers. Initially, all of the C-1 conformers of NTA were fully DFT-optimized in solvent – results obtained are shown in Table 3(b). It was gratifying to note that the differences δG in Table 3(b) between the relevant structures became almost negligibly small. For all protonated forms of NTA, except H_4L^+, the value of δG is about ± 0.1 kcal mol^{-1}. This is equivalent to about ±0.07 log unit of the computed protonation constant, a

typical experimental uncertainty. Similar procedure was applied to all the protonated forms of IDA, MIDA, EIDA, PIDA and HIDA (Table S11, Supporting Information). It is seen in Table 3(b) that only two out of six of the DFT-optimized C-1 conformers have lower energies (HL^{2-} and H_2L^-) when compared with the energies of the original structures seen in Figure 4. Similar observation applies to energies of the DFT-optimized C-1 conformers of the reference molecules. The lowest in value Gibbs free energies G_{aq} were used to calculate protonation constants - results are provided in Table S12, Supporting Information.

Because the use of C-1 conformers has not resulted in any significant change in the computed protonation constants and since there was no direct correlation between energies of the DFT-optimized C-1 conformers and original structures, we have decided to use six lowest in energy MM/MD-generated conformers of NTA and HIDA for further studies. HIDA was selected because this reference molecule, when used in IRns seen in Table 1 generated best estimates in protonation constants. Energy-minimized in solvent conformers C-1 to C-6 of all the protonated forms of NTA and HIDA are shown in Figures S15-S23, Supporting Information. Their ZPVE-corrected energies (E_{min}) and Gibbs free energies (G_{aq}) along with the values of the self-constructed (S-c) structures are shown in Table 4; for each protonated form of $L_{(1)}$ and $L_{(2)}$, the lowest in value E_{min} and G_{aq} are printed in bold. Implementation of an extended conformational analysis resulted in three new lower in energy conformers of NTA (H_2L^-, H_3L, and H_4L^+) as well as HIDA (L^{2-}, H_2L, and H_3L^+).

All lowest DFT-computed G_{aq} values for NTA and HIDA were used to calculate protonation constants – see first row in Table 5. One can note from Table 4 that in two cases involving HIDA (L^{2-} and H_2L) the conformers with lowest energy, E_{min}, do not have the lowest G_{aq} value. It has been decided to include these G_{aq} values (printed in Italic in Table 4) in computing protonation constants to estimate to what degree they influence the final result – see second row in Table 5. When one assumes a study of a new ligand, conformational search most likely would be the first analysis performed to find lowest in energy conformer. To simulate this approach, the self-constructed molecules were excluded in search for lowest energy conformers and results obtained are included in rows 3 and 4 in Table 5; for convenience, protonation constants seen in Table 1 (part involving HIDA) were also included as last row.

Analysis of data seen in Table 5 leads to several interesting observations and conclusions:

(i) It is very important to use lowest energy conformers as it leads to more accurate predictions in protonation constants. It is seen that significant improvement was achieved

for the third and fourth protonation constant where new and lower energy conformers were found. These two theoretically predicted values of protonation constants, which differ from experimental values only by about 0.2 log unit, can be seen as of analytical quality.

(ii) A good prediction was also achieved for the second protonation constant. When the self-constructed structure is excluded (it was not involved in conformational analysis) the average theoretical value tends to be on average about 0.8 log unit larger than experimental value.

(iii) The worst prediction was consistently obtained for the first protonation constant (molecules involved in IRn have highest and negative charges). This is, as it was mentioned above, most likely due to computational evaluations of ionic solvation free energies for highly charged anions; these energies are highly dependent on the solvation model used.

(iv) Consistency in the first protonation constant obtained from conformational analysis and significantly different value obtained for self-constructed molecules suggests that the latter one is most likely an artifact that was obtained due to elimination of errors in energies of molecules used to calculate this protonation constant.

Very important conclusions one can arrive at from the analysis of structures of all the conformers seen in Figures S15-S23 and their relevant energies seen in Table 4:

(i) The lowest energy free ligand L^{3-} of NTA has three carboxylic groups symmetrically distributed around the de-protonated N-atom. The highest energy conformer C-5 has one carboxylic group bent down (see Figure S15, Supporting Information)..

(ii) The lowest energy HL^{2-} form of NTA (C-1) has all three O-atoms from carboxylic groups symmetrically distributed around the N-atom and H-bonded to the hydrogen bonded to nitrogen. Here again, when one of the carboxylic groups is bent down (and hence it is not involved in H-bonding) it results in highest energy conformer C-5 (Figure S16, Supporting Information).

(iii) Another interesting pattern immerges when carboxylic groups start to be protonated. It appears that, as above, the lowest energy structures have symmetrically distributed carboxylic groups around the central N-atom but H-bonded to H-atom on nitrogen via O-atom in carboxylic group that is not protonated. This appears to be a general trend that applies to any number of protonated carboxylic groups - see the C-3 (H_2L^-), C-5 (H_3L), and C-5 (H_4L^+) conformers of NTA in Figures S17-S19, Supporting Information.

(iv) Interestingly, none of the lowest energy H_3L structures has one carboxylic group bent down, as it is present in solid state (see Figure 3). This confirms our earlier

supposition that the strong intermolecular H-bonding in solid state must not be present in a solution and all the carboxylic groups preferentially be involved in intramolecular H-bonding with H-atom on nitrogen. The same applies to all the protonated forms of NTA; in all cases with carboxylic group bent down the structures belong to the higher energy conformers.

(v) Structural similarities of HIDA to NTA (this pair of ligands produced best predictions in protonation constants) can now be appreciated from analysis of structures seen in Figures S7, S20-S23, Supporting Information. As an example, the lowest in energy HL^- structure of HIDA resembles the lowest energy C-1 conformer of NTA. Similar applies to C-4 conformer of the H_3L^+ form of HIDA when compared with the H_4L^+ form of NTA. Interestingly, only one carboxylic group in the H_3L^+ form of HIDA is H-bonded to H-atom on nitrogen via O-atom that is not protonated, as it is in NTA. It is possible that our conformational search has not found the structure with both carboxylic groups placed as is in NTA, or such a conformer is not stable – additional study would have to be carried out to clarify that.

(vi) It is seen in Table 4 that the self-constructed structure of the L^{2-} form of HIDA is significantly different from the lowest energy found. This might suggest that a very good prediction in the first protonation constant of NTA obtained with involvement of only self-constructed structures could indeed be an artifact. In order to improve the first protonation constant (highly and negatively charged molecules are involved) one might consider making use of higher level of theory, more sophisticated solvation models, or incorporation of water molecules in the primary solvation layer. On the other hand, it appears that the low level of theory and solvation model employed here were sufficient for the remaining three protonation constant.

All the above observations lead to the conclusion that MM/MD-based conformational search might be a useful tool in a quick and preliminary search of conformers that have to be properly studied using the QM methods. In addition, structure-energy correlation analysis, as performed above, provides plenty of important hints. Having a substantial bank of information of this kind (many more systems must be studied) would be invaluable as it should allow a quick selection of most likely conformers from the large set of structures generated by MM/MD-based conformational analysis.

4. Conclusions

It has been shown that prediction of several stepwise protonation constants for poly-charged molecules, such as NTA (its charge varied between –3, through neutral, to +1) is not only possible but also accuracy achieved preserved their experimental order ($\log K_H^{(1)} > \log K_H^{(2)} > \log K_H^{(3)} > \log K_H^{(4)}$) even though differences between some of them is within one log unit. This has been achieved using methodology based on an explicit application of isodesmic reaction principles (where the free energy change of the protonation reaction in solution was obtained using the free energies in solution from a single continuum model) involving two similar in structure molecules, the studied one and reference molecule, and the lowest in energy conformers. Results obtained strongly suggest that to obtain computed values to within ± 1 log unit (or better) of the experimental protonation constants, the structure of (the cavity it occupies in solvation models utilized), kind and a number of donor atoms in, as well as charges on the reference molecule must be similar to the molecule of interest; the charge and its distribution within the reference molecule appear to be of utmost importance. A selection of an appropriate protonated form of the reference molecule has a decisive impact on accuracy in predicted protonation constants. The protonation reactions in which the studied and reference molecule are involved in, when incorporated in isodesmic reaction, must be (if possible) of the same order; first (or generally *n*th) protonation reaction of the reference molecule must be used to compute the first (or *n*th) protonation constant of studied molecule. The third and fourth protonation constants obtained here were of analytical quality. However, in case of the first protonation constant, that involves highly charged anions (–3 and –2 for the studied and the reference molecule, respectively), a significantly larger than experimental value (by 3 log units) was obtained. Clearly, more work has to be done to improve this value. Nevertheless, we are convinced that results reported here can be used as a guide in constructing isodesmic reactions that are useful for the theoretical prediction of protonation/dissociation constants, particularly for compounds for which solvent and gas structures differ significantly.

Acknowledgments

Financial support of the National Research Foundation of South Africa and the University of Pretoria is highly appreciated.

Supporting Information Available

The following has been included in the Supporting Information: reported crystal structures of IDA, MIDA, and HIDA; structural matrix of the solvent-optimized H_3L^* and H_3L forms

of NTA, H_2L form of IDA and MIDA, and H_3L form of HIDA; energy-minimized in solvent (PCM/UA0) at the RB3LYP/6-311+G(d,p) level of theory all self-constructed protonated forms of IDA, MIDA, EIDA, PIDA, and HIDA; results obtained from computations involving thermodynamic cycles; ZPVE-corrected minimum and Gibbs free energies of all protonated forms of NTA and reference molecules (IDA, MIDA, EIDA, PIDA, and HIDA) obtained from full energy optimization at the RB3LYP/6-311+G(d,p) level of theory in conjunction with the PCM/UA0 solvation model; experimental stepwise protonation constants of IDA, MIDA, EIDA, PIDA, and HIDA at μ = 0.0 and 0.1 M and 25 ^{o}C; comparison of experimental and calculated stepwise protonation constants of NTA using protonation constants of the reference molecules IDA, MIDA, EIDA, PIDA and HIDA; output structures (H_3L form of NTA and H_2L form of IDA) obtained from the single point frequency calculations at the G3/6-311+G(d,p) level of theory on the solvent optimized molecules optimized at the RB3LYP/6-311+G(d,p) level of theory in conjunction with the PCM/UA0 solvation model; energy-minimized, at the RB3LYP/6-311+G(d,p) level of theory in conjunction with the PCM/UA0 solvation model, all protonated forms of C-1 conformers of NTA, IDA, MIDA, EIDA, PIDA, HIDA; minimum energies of MM/MD-generated conformers in solvent and energies obtained from MM-based SPC performed on the IDA, MIDA, EIDA, PIDA and HIDA structures; DFT-calculated solvent-optimized energies of all self-constructed protonated forms of IDA, MIDA, EIDA, PIDA and HIDA and lowest energy MM/MD-generated C-1 conformers; comparison of experimental and calculated stepwise protonation constants of NTA obtained from the lowest energy MM/MD conformers of IDA, MIDA, EIDA, PIDA and HIDA; energy-minimized in solvent (PCM/UA0) at the RB3LYP/6-311+G(d,p) level of theory MM/MD-generated conformers of the protonated forms of NTA and HIDA.

Reference

1. Namazian, M.; F. Kalantary-Fotooh; Noorbala, M. R.; Searles, D. J.; Coote, M. C. *J. Mol. Struct. (Theochem)* **2006**, *758*, 275.
2. Jacquemin, D.; Perpete, E. A.; Ciofini, I.; Adamo, C. *J. Phys. Chem. A* **2008**, *112*, 794.
3. Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 5610.
4. Namazian, M.; Heidary, H. *J. Mol. Struct. (Theochem)* **2003**, *620*, 257.
5. Namazian, M.; Halvani, S.; Noorbala, M. R. *J. Mol. Struct. (Theochem)* **2004**, *711*, 13.
6. Namazian, M.; Halvani, S. *J. Chem. Thermodyn.* **2006**, *38*, 1495.
7. Namazian, M.; Zakery, M.; Noorbala, M. R.; Coote, M. L. *Chem. Phys. Lett.* **2008**, *451*, 163.
8. Charif, I. E.; Mekelleche, S. M.; Villemin, D.; Mora-Diez, N. *J. Mol. Struct. (Theochem)* **2007**, *818*, 1.
9. Liptak, M. D.; Shields, G. C. *Int. J. Quantum Chem.* **2001**, *85*, 727.
10. Liptak, M. D.; Shields, G. C. *J. Am. Chem. Soc.* **2001**, *123*, 7314.
11. Liptak, M. D.; Gross, K. C.; Seybold, P. G.; Feldgus, S.; Shields, G. C. *J. Am. Chem. Soc.* **2002**, *124*, 6421.
12. Bensen, M. T.; Moser, M. L.; Peterman, D. R.; Dinescu, A. *J. Mol. Struct. (THEOCHEM)* **2008**, *867*, 71.
13. Sang-Aroon, W.; Ruangpornvisuti, V. *Int. J. Quantum Chem.* **2008**, *108*, 1181.
14. Jang, Y. H.; Goddard III, W. A.; Noyes, K. T.; Sowers, L. C.; Hwang, S.; Chung, D. S. *J. Phys. Chem. B* **2003**, *107*, 344.
15. Jitariu, L. C.; Masters, A. J.; Hiller, I. H. *J. Chem. Phys* **2004**, *121*, 7795.
16. Bryantsev, V. S.; Diallo, M. S.; Goddard III, W. A. *J. Phys. Chem. A* **2007**, *111*, 4422.
17. Gao, D.; Svoronos, P.; Wong, P. K.; Maddalena, D.; Hwang, J.; Walker, H. *J. Phys. Chem. A* **2005**, *109*, 10776.
18. Murlowska, K.; Sadlej-Sosnowska, N. *J. Phys. Chem. A* **2005**, *109*, 5590.
19. Saracino, G. A. A.; Improta, R.; Barone, V. *Chem. Phys. Lett.* **2003**, *373*, 411.
20. da Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *J. Phys. Chem. A* **1999**, *103*, 11194.
21. Li, G.; Cui, Q. *J. Phys. Chem. B* **2003**, *107*, 14521.
22. Dahlke, E. E.; Cramer, C. J. *J. Phys. Org. Chem.* **2003**, *16*, 336.
23. Klamt, A.; Eckert, F.; Diedenhofen, M.; Beck, M. E. *J. Phys. Chem. A* **2003**, *107*, 9380.
24. Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110*, 2493.
25. Pliego Jr., J. R.; Riveros, J. M. *J. Phys. Chem. A* **2002**, *106*, 7434.
26. Klicic, J. J.; Friesner, R. A.; Liu, S.; Guida, W. C. *J. Phys. Chem. A* **2002**, *106*, 1327.
27. Jang, Y. H.; Sowers, L. C.; Cagin, T.; Goddard III, W. A. *J. Phys. Chem. A* **2001**, *105*, 274.
28. Himo, F.; Eriksson, L. A.; Blomberg, M. R. A.; Siegbahn, P. E. M. *Int. J. Quantum Chem.* **2002**, *76*, 714.
29. Himo, F.; Noodleman, L.; Blomberg, M. R. A.; Siegbahn, P. E. M. *J. Phys. Chem. A* **2002**, *106*, 8757.
30. Chipman, D. M. *J. Phys. Chem. A* **2002**, *106*, 7413.
31. Mujika, J. I.; Mercero, J. M.; Lopez, X. *J. Phys. Chem. A* **2003**, *107*, 6099.
32. Scharnagl, C.; R. A. Raupp-Kossmann. *J. Phys. Chem. B* **2004**, *108*, 477.
33. Lee, I.; Kim, C. K.; Lee, I. Y.; Kim, C. K. *J. Phys. Chem. A* **2000**, *104*, 6332.

34. Maksic, Z. B.; Vianello, R. *J. Phys. Chem. A* **2002**, *106*, 419.
35. Magill, A. M.; Cavell, K. J.; Yates, B. F. *J. Am. Chem. Soc.* **2004**, *126*, 8717.
36. Li, H.; Hains, A. W.; Everts, J. E.; Robertson, A. D.; Jensen, J. H. *J. Phys. Chem. B* **2002**, *106*, 3486.
37. da Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *J. Phys. Chem. A* **2000**, *104*, 2402.
38. da Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *Chem. Phys. Lett.* **2003**, *381*, 244.
39. Alvarado-Gonzalez, M.; Orrantia-Borunda, E.; Glossman-Mitnik, D. *J. Mol. Struct. (THEOCHEM)* **2008**, *869*, 105.
40. Sulpizi, M.; Sprik, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 5238.
41. Tao, L.; Han, J.; Tao, F. M. *J. Phys. Chem. A* **2008**, *112*, 775.
42. Kieseritzky, G.; Knapp, E. W. *J. Comp. Chem.* **2008**, *29*, 2575.
43. da Silva, G.; Kennedy, E. M.; Dlugogorski, B. Z. *J. Phys. Chem. A* **2006**, *110*, 11371.
44. Ulander, J.; Broo, A. *Int. J. Quantum Chem.* **2005**, *105*, 866.
45. Gross, K. C.; Seybold, P. G. *Int. J. Quantum Chem.* **2000**, *80*, 1107.
46. Gross, K. C.; Seybold, P. G. *Int. J. Quantum Chem.* **2001**, *85*, 569.
47. Pliego Jr., J. R.; Riveros, J. M. *Chem. Phys. Lett.* **2000**, *332*, 597.
48. Pliego Jr., J. R. *Chem. Phys. Lett.* **2003**, *367*, 145.
49. Adam, K. R. *J. Phys. Chem. A* **2002**, *106*, 11963.
50. Kim, Y. J.; Streitwieser, A. *J. Am. Chem. Soc.* **2002**, *124*, 5757.
51. Schuurmann, G.; Cossi, M.; Barone, V.; Tomasi, J. *J. Phys. Chem. A* **1998**, *102*, 6706.
52. Soriano, E.; Cerdan, S.; Ballesteros, P. *J. Mol. Struct. (Theochem)* **2004**, *684*, 121.
53. Sadlej-Sosnowska, N. *Theor. Chem. Acc.* **2007**, *118*, 281.
54. Govender, K. K.; Cukrowski, I. *J. Phys. Chem. A* **2009**, *113*, 3639.
55. Private communication, *Gaussian Technical Support.*
56. Cramer, C. J. *in Essentials of Computational Chemistry*; John Wiley & Sons Ltd.: New York, 2002, pp. 371-372.
57. NIST Standard Reference Database 46. NIST Critically Selected Stability Constants of Metal Complexes Database; Version 8.0, Data collected and selected by R. M. Smith and A. E. Martell ed.; US Department of Commerce, National Institute of Standards and Technology, 2004.
58. Grigoriev, M. S.; Auwer, C. D.; Meyer, D.; Moisy, P. *Acta Cryst.* **2006**, *C62*, m163.
59. Yu, L. C.; Liu, S. L.; Liang, E. X.; Wen, C. L. *J. Coord. Chem.* **2007**, *60*, 2097.
60. Malevich, D.; Wang, Z.; Tremaine, P. R. *J. Sol. Chem.* **2006**, *35*, 1303.
61. Vohra, M. S.; Davis, A. P. *J. Colloid and Interface Sci.* **1997**, *194*, 59.
62. Karhu, J.; Harju, L.; Ivaska, A. *Analytica Chimica Acta* **1999**, *380*, 105.
63. Souaya, E. R.; Hanna, W. G.; Ismail, E. H.; Milad, N. E. *J. Coord. Chem.* **2004**, *57*, 825.
64. Quartacci, M. F.; Irtelli, B.; Baker, A. J. M; Navari-Izzo, F. *Chemosphere* **2007**, *68*, 1920.
65. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Pettersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V.

G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03, Revision D.01; Gaussian, Inc., Wallingford, CT, 2004.
66. GaussView 4.1.2; Gaussian Inc,; Wallingford, CT, 2004.
67. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.
68. Cances, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032.
69. Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *Chem. Phys. Lett.* **1996**, *255*, 327.
70. Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117.
71. Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995.
72. Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669.
73. Allen, F. H. *Acta Cryst.* **2002**, *B58*, 380.
74. Skrzypczak-Jankun, E.; Smith, D. A.; Maluszynska, H. *Acta Cryst.* **1994**, *C50*, 1097.
75. Kaneyoshi, M.; Bond, A.; Jones, W. *Acta Cryst.* **1999**, *C55*, 1260.
76. da Silva, G.; Bozzelli, J. W. *J. Phys. Chem. A* **2006**, *110*, 13058.
77. da Silva, G.; Bozzelli, J. W. *J. Phys. Chem. C* **2007**, *111*, 5760.
78. Guo, Y.; Gao, H.; Twamley, B.; Shreeve, J. M. *Advanced Materials* **2007**, *19*, 2884.
79. Espinosa-Garcia, J.; Garcia-Bernaldez, J. C. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4096.
80. Krossing, I.; Raabe, I. *Chem. Eur. J.* **2004**, *10*, 5017.
81. Kutt, A.; Movchun, V.; Rodima, T.; Dansauer, T.; Rusanov, E. B.; Leito, I.; Kaljurand, I.; Koppel, J.; Pihl, V.; Koppel, I.; Ovsjannikov, G.; Toom, L.; Mishima, M.; Medebielle, M.; Lork, E.; Roschenthaler, G. V.; Koppel, I. A.; Kolomeitsev, A. A. *J. Org. Chem.* **2008**, *73*, 2607.
82. Nolan, E. M.; Linck, R. G. *J. Am. Chem. Soc.* **2000**, *122*, 11497.
83. Peeters, D.; Leroy, G.; Wilante, C. *J. Mol. Struct.* **1997**, *416*, 21.
84. Wang, L.; Heard, D. E.; Pilling, M. J.; Seakins, P. *J. Phys. Chem. A* **2008**, *112*, 1832.
85. Schrodinger, L. L. C. *Schrodinger Maestro*; 32nd Floor, Tower 45, 120 West Forty-Fifth Street, New York, 10036: New York, 2003.

Table 1. Comparison of experimental[57] (Exp) and calculated (from isodesmic reactions) stepwise protonation constants of NTA ($L_{(1)}$), as log K_H, using protonation constants of the reference molecules at ionic strength μ = 0.0 or 0.1 M and 25 °C. All energies are reported in kcal mol^{-1}.

Reaction	ΔG_{IRn}(aq)	ΔG_{PRn}(aq) for $L_{(1)}$	log K_H	Exp[a]	δ	ΔG_{PRn}(aq) for $L_{(1)}$	log K_H	Exp[b]	δ
$L_{(2)}$ = IDA									
$L_{(1)}^{3-} + HL_{(2)}^{-} = HL_{(1)}^{2-} + L_{(2)}^{2-}$	–7.298	–20.654	15.14	10.334	4.81	–20.040	14.69	9.66	5.03
$HL_{(1)}^{2-} + H_2L_{(2)} = H_2L_{(1)}^{-} + HL_{(2)}^{-}$	0.052	–3.822	2.80	2.94	–0.14	–3.522	2.58	2.52	0.06
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)} + HL_{(2)}^{-}$	2.504	–1.371	1.00	2.00[c]	–1.00	–1.071	0.78	1.81	–1.03
$H_3L_{(1)} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L_{(2)}$	4.143	1.620	–1.19	1.00[b]	–2.19	1.729	–1.27	1.00	–2.27
$L_{(2)}$ = MIDA									
$L_{(1)}^{3-} + HL_{(2)}^{-} = HL_{(1)}^{2-} + L_{(2)}^{2-}$	–5.342	–18.998	13.93	10.334	3.59	–18.425	13.51	9.66	3.85
$HL_{(1)}^{2-} + H_2L_{(2)} = H_2L_{(1)}^{-} + HL_{(2)}^{-}$	–0.171	–3.705	2.72	2.94	–0.22	–3.336	2.45	2.52	–0.07
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)} + HL_{(2)}^{-}$	2.280	–1.253	0.92	2.00[c]	–1.08	–0.885	0.65	1.81	–1.16
$H_3L_{(1)} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L_{(2)}$	3.458	0.866	–0.63	1.00[b]	–1.63	0.866	–0.63	1.00	–1.63
$L_{(2)}$ = EIDA									
$L_{(1)}^{3-} + HL_{(2)}^{-} = HL_{(1)}^{2-} + L_{(2)}^{2-}$	–4.435	–18.255	13.38	10.334	3.05	–18.010	13.20	9.66	3.54
$HL_{(1)}^{2-} + H_2L_{(2)} = H_2L_{(1)}^{-} + HL_{(2)}^{-}$	–1.605	–5.288	3.88	2.94	0.94	–4.633	3.40	2.52	0.88
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)} + HL_{(2)}^{-}$	0.847	–2.836	2.08	2.00[c]	0.08	–2.182	1.60	1.81	–0.21
$H_3L_{(1)} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L_{(2)}$	2.175	–0.008	0.01	1.00[b]	–0.99	–0.008	0.01	1.00	–0.99
$L_{(2)}$ = PIDA									
$L_{(1)}^{3-} + HL_{(2)}^{-} = HL_{(1)}^{2-} + L_{(2)}^{2-}$	–4.074	–18.303	13.42	10.334	3.08	–17.785	13.04	9.66	3.38
$HL_{(1)}^{2-} + H_2L_{(2)} = H_2L_{(1)}^{-} + HL_{(2)}^{-}$	–2.081	–5.478	4.02	2.94	1.08	–5.137	3.77	2.52	1.25
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)} + HL_{(2)}^{-}$	0.370	–3.027	2.22	2.00[c]	0.22	–2.686	1.97	1.81	0.16
$H_3L_{(1)} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L_{(2)}$	1.089	–0.412	0.30	1.00[b]	–0.70	–1.326	0.97	1.00	–0.03
$L_{(2)}$ = HIDA[b]									
$L_{(1)}^{3-} + HL_{(2)}^{-} = HL_{(1)}^{2-} + L_{(2)}^{2-}$	–3.535	–15.377	11.27	10.334	0.94	–15.377	11.27	9.66	1.61
$HL_{(1)}^{2-} + H_2L_{(2)} = H_2L_{(1)}^{-} + HL_{(2)}^{-}$	–1.398	–4.399	3.22	2.94	0.28	–4.399	3.22	2.52	0.70
$H_2L_{(1)}^{-} + H_2L_{(2)}$	1.054	–1.948	1.43	2.00[c]	–0.57	–1.948	1.43	1.81	–0.38

$= H_3L_{(1)} + HL_{(2)}^-$									
$H_3L_{(1)} + H_3L_{(2)}^+$ $= H_4L_{(1)}^+ + H_2L_{(2)}$	1.601	–0.581	0.43	1.00[b]	–0.57	–0.581	0.43	1.00	–0.57

[a]) μ = 0.0 M and 20 °C; [b]) μ = 0.1 M and 25 °C, [c]) μ = 0.0 M and 25 °C.

Table 2. Thermochemical data used to calculate forth protonation constant of NTA involving IRn in eq 12. $G_{gas}^{(1)}$ values were obtained at the B3LYP/6-311+G(d,p) level of theory. $G_{gas}^{(2)}$ values were obtained at the RMP2/6-311+G(d,p) level of theory. G_{sol} values were obtained at the B3LYP/6-311+G(d,p) level of theory in solvent (PCM/UA0). $L_{(1)}$ = NTA, $L_{(2)}$ = IDA. For details see the text.

Molecule	G_{sol} / au	$G_{gas}^{(1)}$ / au	ΔG_{sol} / au	$G_{gas}^{(2)}$ / au	G_{sol}(TC) / au
$H_3L_{(1)}$	-740.332226	-740.252149	-0.080077	-738.373664	-738.453741
$H_3L_{(2)}^+$	-512.862287	-512.727315	-0.134972	-511.414007	-511.548979
$H_4L_{(1)}^+$	-740.764545	-740.625849	-0.138696	-738.746987	-738.885683
$H_2L_{(2)}$	-512.423365	-512.349858	-0.073507	-511.03375	-511.107257

Table 3. Part (a). Minimum energies (kJ mol^{-1}) of MM/MD-generated conformers, C-1 to C-5, in solvent ($E_{C\text{-}1}$ to $E_{C\text{-}5}$) and energies obtained from MM-based SPC (E_{SPC}; kJ mol^{-1}) performed on the NTA structures seen in Figure 4. Part (b). DFT-calculated solvent-optimized energies (E_{min} = ZPVE-corrected energy) of all protonated forms of the ligand NTA seen in Figure 4 and lowest energy C-1 conformers.

(a)

L = NTA	E_{SPC}	$E_{C\text{-}1}$	δE[a] kJ/mol	δE[a] kcal/mol	$E_{C\text{-}2}$	$E_{C\text{-}3}$	$E_{C\text{-}4}$	$E_{C\text{-}5}$
L^{3-}	–770.81	–799.18	28.37	6.78	–799.17	–798.69	–798.69	–796.37
HL^{2-}	–1190.19	–1216.05	25.86	6.18	–1216.05	–1201.93	–1201.93	–1201.22
H_2L^{-}	–991.30	–1027.95	36.65	8.76	–1027.95	–1027.67	–1027.67	–1025.89
H_3L	–758.87	–803.94	45.07	10.77	–803.94	–802.89	–802.89	–802.02
H_3L^{*}	–750.22	–803.95	54.38	13.00	–803.94	–802.89	–802.02	–802.02
H_4L^{+}	–489.11	–543.49	53.73	12.84	–543.49	–539.35	–538.46	–538.46

[a]) $\delta E = E_{SPC} - E_{C\text{-}1}$

(b)

	Structures seen in Fig. 4		C-1 structures			
L = NTA	E_{min} (Hartree)	G_{aq} (Hartree)	E_{min} (Hartree)	G_{aq} (Hartree)	$\delta G(aq)$[b] Hartree	$\delta G(aq)$[b] kcal/mol
L^{3-}	–738.936972	*–738.978069*	–738.936883	–738.977928	–0.000141	–0.09
HL^{2-}	–739.410772	–739.451765	–739.410827	*–739.451947*	0.000182	0.11
H_2L^{-}	–739.852748	–739.893949	–739.852711	*–739.894128*	0.000179	0.11
H_3L	–740.290659	*–740.332226*	–740.290552	–740.332033	–0.000193	–0.12
H_3L^{*}	–740.293694	*–740.335141*	–740.290657	–740.332261	–0.000182	–0.11
H_4L^{+}	–740.722681	*–740.764545*	–740.722433	–740.764363	–0.002880	–1.81

[b]) $\delta G_{aq} = G_{aq}$(self-constructed structure) – G_{aq}(C-1)

Table 4. Energies E_{min} (E_{min} = ZPVE-corrected energy) and G_{aq} (both in au) obtained for all protonated forms of the self-constructed structures and conformers C-1 to C-6 that were fully energy-optimized in Gaussian at the RB3LYP/6-311+G(d,p) level of theory in solvent (PCM-UA0). (a) NTA, (b) HIDA.

(a)

	L = NTA									
	L^{3-}		**HL^{2-}**		**H_2L^-**		**H_3L**		**H_4L^+**	
	E **min**	***G*** **aq**	***E*** **min**	***G*** **aq**	***E*** **min**	***G*** **aq**	***E*** **min**	***G*** **aq**	***E*** **min**	***G*** **aq**
S-c	–**738.936972**	**–738.978069**	–739.410772	–739.451765	–739.852748	–739.893949	–740.290659	–740.332226	–740.722681	–740.764545
C1	–738.936883	–738.977928	–**739.410827**	**–739.451947**	–739.852711	–739.894128	–740.290552	–740.332033	–740.722433	–740.764363
C2	–738.936941	–738.977965	–739.410753	–739.451937	–739.852772	–739.894052	–740.290599	–740.332198	–740.722408	–740.764258
C3	–738.934628	–738.976228	–739.407682	–739.448582	–**739.854620**	**–739.895947**	–740.292742	–740.334330	–740.722671	–740.764351
C4	–738.934684	–738.976365	–739.410791	–739.451885	–739.852733	–739.893980	–740.290717	–740.332398	–740.727700	–740.769451
C5	–738.931462	–738.973284	–739.402567	–739.447349	–739.846053	–739.888621	–**740.294578**	**–740.336185**	–**740.729966**	**–740.771858**
C6	–738.934569	–738.976095	—	—	–739.852702	–739.894193	–740.292501	–740.334200	–740.723142	–740.765024

(b)

	L = HIDA							
	L^{2-}		**HL^-**		**H_2L**		**H_3L^+**	
	E **min** **(Hartree)**	***G*** **aq** **(Hartree)**	***E*** **min** **(Hartree)**	***G*** **aq** **(Hartree)**	***E*** **min** **(Hartree)**	***G*** **aq** **(Hartree)**	***E*** **min** **(Hartree)**	***G*** **aq** **(Hartree)**
S-c	–665.299655	–665.340516	–**665.767828**	**–665.808578**	–666.207584	–666.248534	–666.642422	–666.683405
C1	–**665.305277**	*–665.344936*	–665.766303	–665.807676	–666.205011	–666.246706	–666.642406	–666.683411
C2	–665.305221	**–665.344973**	–665.765839	–665.807813	–666.204984	–666.246766	–666.642341	–666.683250
C3	–665.300319	–665.339783	–665.765002	–665.805790	–**666.208044**	*–666.248784*	–666.642415	–666.683437
C4	–665.298207	–665.340665	–665.765071	–665.805962	–666.206880	**–666.248997**	–**666.644447**	**–666.685416**
C5	–665.301309	–665.341919	–665.765305	–665.806013	–666.205287	–666.246723	–666.641862	–666.682718
C6	–665.298902	–665.340283	–665.765156	–665.805658	–666.204193	–666.245406	–666.643550	–666.683807

Table 5. Computed stepwise protonation constants (as log $K_H^{(n)}$) of NTA obtained from isodesmic reactions involving HIDA – see Table 1. δ is the difference between the computed and experimental value. S-c stands for self-constructed structures. C-1 to C-6 represents MM/MD-generated conformers that were fully energy optimized in Gaussian at the RB3LYP/6-311+G(d,p) level of theory in solvent (PCM-UA0).

Structures considered	$K_H^{(1)}$	δ	$K_H^{(2)}$	δ	$K_H^{(3)}$	δ	$K_H^{(4)}$	δ
C-1 to C-6 and S-c[a]	13.40	3.07	3.85	0.91	2.12	0.12	1.26	0.26
C-1 to C-6 and S-c[b]	13.39	3.06	3.94	1.00	2.21	0.21	1.16	0.16
C-1 to C-6[a]	13.80	3.47	3.50	0.56	1.76	–0.24	1.26	0.26
C-1 to C-6[b]	13.85	3.52	3.53	0.59	1.80	–0.20	1.16	0.16
S-c[a]	11.27	0.94	3.22	0.28	1.43	–0.57	0.43	–0.57

[a]Minimum G_{aq} were used; [b]G_{aq} for E_{min} were used

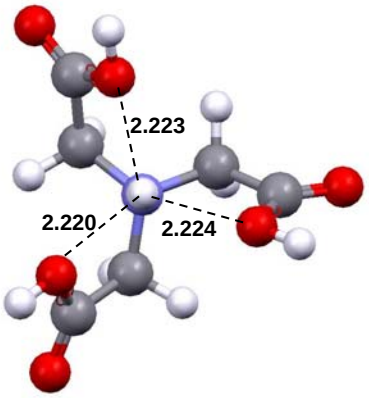
Captures for Figures

Figure 1. Top view of the ligands (in fully protonated forms) discussed in this work.

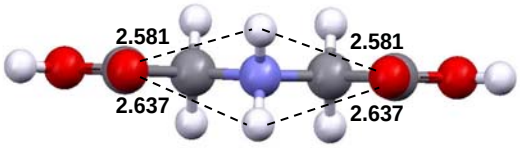
Figure 2: Fully labeled reported crystal structure[74] of the H_3L^* form of NTA.

Figure 3. Crystallographic structure[74] of NTA (molecules within a unit cell) with selected intra- and intermolecular non-bonding interactions marked by dashed lines and distances in Å.

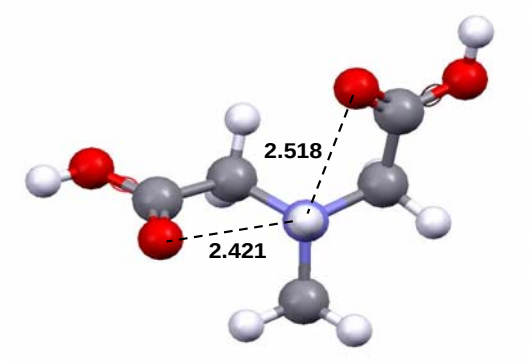
Figure 4: Self-constructed protonated forms of NTA and a crystal structure H_3L^* fully optimized at the RB3LYP/6-311+G(d,p) level of theory in solvent (PCM/UA0).



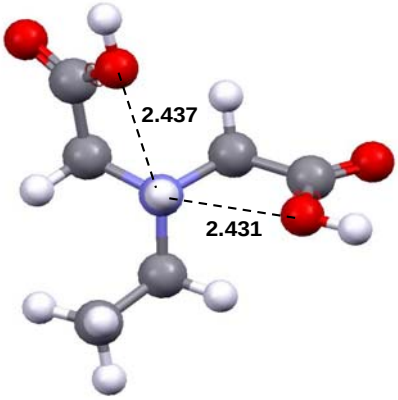
NTA (H_4L^+)



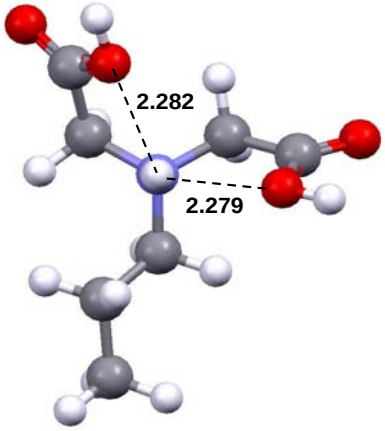
IDA (H_3L^+)



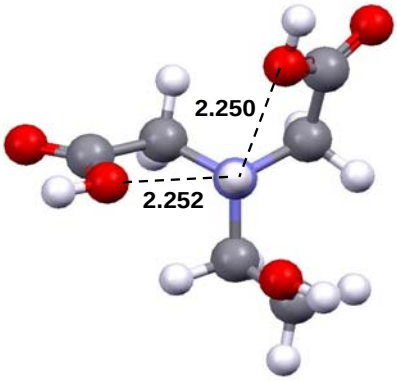
MIDA (H_3L^+)



EIDA (H_3L^+)

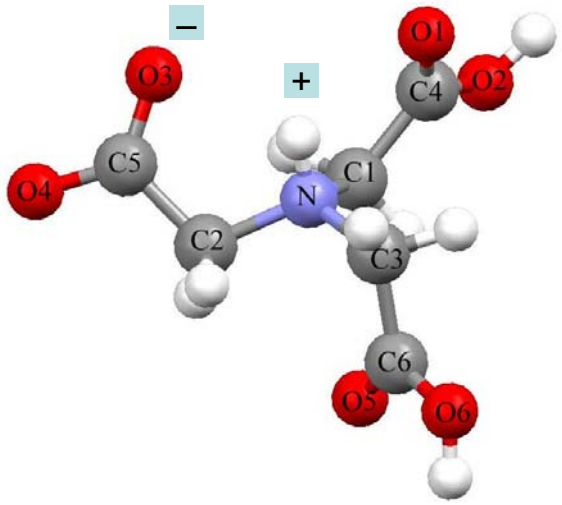


PIDA (H_3L^+)



HIDA (H_3L^+)

Figure 1. Top view of the ligands (in fully protonated forms) discussed in this work.



NTA (H_3L^*)

Figure 2: Fully labeled reported crystal structures[74] of the H_3L^* form of NTA.

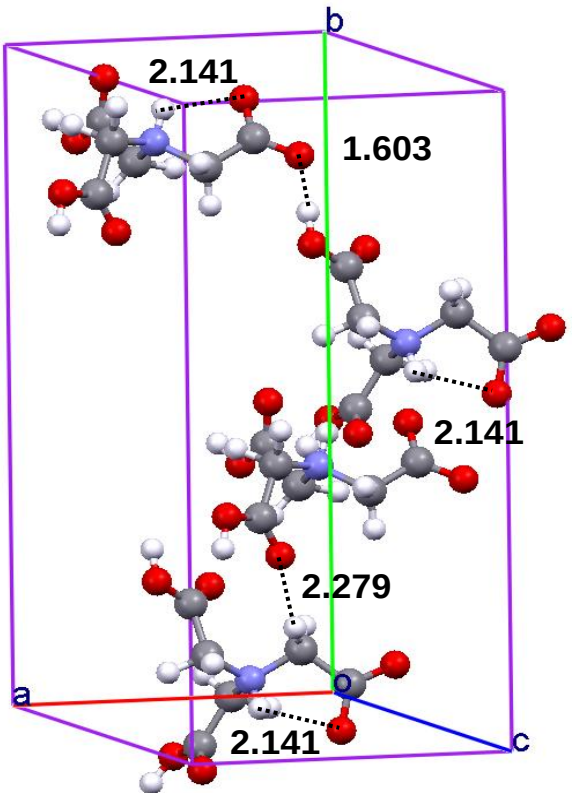
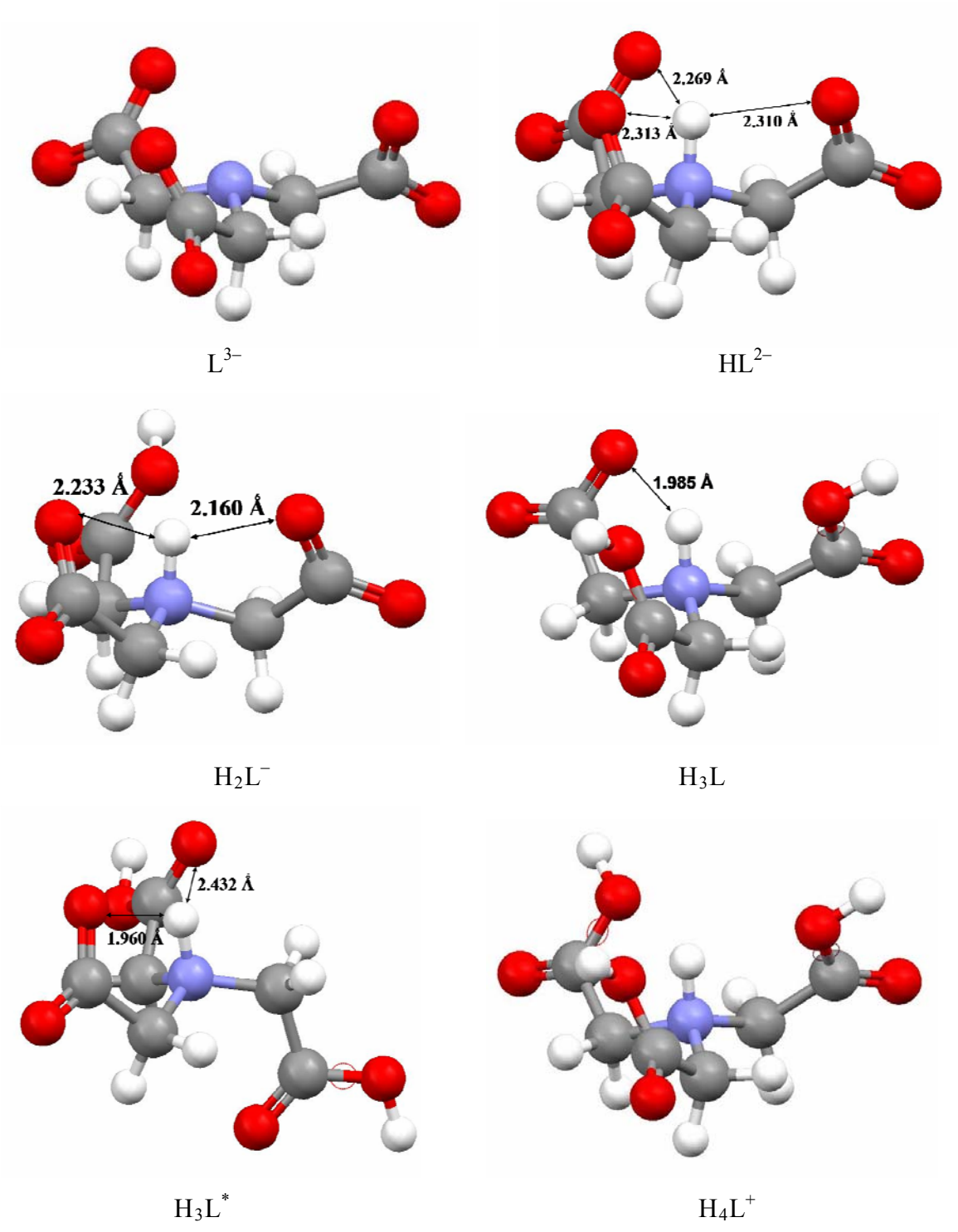


Figure 3. Crystallographic structure[74] of NTA (molecules within a unit cell) with selected intra- and intermolecular non-bonding interactions marked by dashed lines and distances in Å.



L^{3-} HL^{2-}

H_2L^- H_3L

H_3L^* H_4L^+

Figure 4: Self-constructed protonated forms of NTA and a crystal structure H_3L^* fully optimized at the RB3LYP/6-311+G(d,p) level of theory in solvent (PCM/UA0).