UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Theoretical studies of nitrilotriacetic acid and nitrilotripropionic acid geometries for estimation of the stability of metal complexes by Density Functional Theory

by

Krishna Kuben Govender

Supervisor: Professor Ignacy Cukrowski

Department of Chemistry

University of Pretoria, Pretoria

Co-Supervisor: Professor Helder M. Marques

School of Chemistry

University of Witwatersrand, Johannesburg

Submitted in partial fulfilment of the requirements for the degree

Master of Science

in the Faculty of Natural and Agricultural Sciences

University of Pretoria

Pretoria

June, 2009

Declaration

I hereby declare that the work contained within this Dissertation is my own. It is being submitted for the Degree of Master of Science in the University of Pretoria, Pretoria. It has not been submitted before for any degree or examination in any other University.

_____

(Signature of candidate)

_____ day of _____ 2009.

Outputs from this work

Conference presentations:

Krishna K. Govender and Ignacy Cukrowski. Theoretical prediction of protonation constants in solvent for Nitrilopropionic acid using Density Functional Theory, *Proceedings of the South African Chemical Institute Young Chemists' Symposium*, University of Witwatersrand, October 2008. (Oral)

Krishna K. Govender, Ignacy Cukrowski and Helder M. Marques. Theoretical prediction of protonation constants for NTA and NTPA, *Proceedings of the 39th National Convention of the South African Chemical Institute*, University of Stellenbosch, December 2008. (Poster)

Publications:

Krishna K. Govender and Ignacy Cukrowski. Density functional theory in prediction of four stepwise protonation constants for nitrilotripropanoic acid (NTPA), *J. Phys. Chem. A*, **2009,** *113*, 3639-3647.

Krishna K. Govender and Ignacy Cukrowski. 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, *J. Phys. Chem. A*, Submitted for publication.

Abstract

Nitrilotriacetic Acid (NTA) is an organic ligand which has been extensively studied due to its biological significance and excellent chelating properties. Nitrilotripropionic Acid (NTPA) is a ligand that is believed to possess similar properties to NTA, but has not been as extensively studied. It has been experimentally determined that metal complexes of NTA are orders of magnitude stronger than those formed with NTPA. This is surprising, especially considering that the ligands do not differ that much from each other. NTPA contains an additional $–CH_2–$ group in each of the acid containing arms as compared to NTA. The aim of these studies were to explain, theoretically, why this is the case. Analyses were conducted with a number of software programs including, Gaussian 03, Schrödinger Maestro and AIM 2000. All Density Functional Theory (DFT) studies were conducted in solvent at the RB3LYP/6-311+G(d,p) level of theory in conjunction with a number of different solvation models. En route to explaining why the complexes differ in stability a new methodology was utilized (isodesmic reactions) in which the four stepwise protonation constants of both NTA and NTPA were successfully predicted; in fact these were the most accurate values predicted to date by DFT methods. The final step of these studies focused on predicting stability constants of metal (Zn^{2+} and Ni^{2+}) complexes of NTA and NTPA. These predictions were not as accurate as those achieved for the prediction of protonation constants; however, success was achieved in predicting the trend – complexes with NTA are orders of magnitude stronger than complexes formed with NTPA. The most important observation revealed that H–clashes and C–H···O hydrogen bonds present in M(NTPA) complexes, which are not present in M(NTA) complexes, result in the formation of additional rings which contributes to the formation of a cage. It was discovered that the H-clashes present in the M(NTPA) complexes were contributing to the overall stability of the molecule. This is completely contradictory to a previous explanation in which H-clashes, being a result of steric crowding, resulted in destabilization of a complex. If the H-clashes were not present in the M(NTPA) complexes there would not be enough stabilizing factors present in the molecule which will inevitably result in the non-existence of M(NTPA) complexes.

Keywords: NTA, NTPA, computational chemistry, protonation constants, complex stability.

Acknowledgements

Firstly, my deepest gratitude goes to my parents and siblings, especially my sister, as well as my roommate and brother-in-law for their moral and financial support, offered to me wholeheartedly, during the course of my postgraduate studies. I also wish to extend my most sincere gratitude to my research supervisor Professor Ignacy Cukrowski. It was really an enjoyable experience working with him. I thank him for all of his academic guidance as well as the moral and financial support which he provided throughout the duration of my research project. My appreciation also goes to my fellow colleagues from the Electrochemistry Research Group (at the University of Pretoria) who have all contributed, in one way or the other, to the success of this project. Finally, I wish to bestow my gratitude to the University of Pretoria and the National Research Foundation for all of the financial support which they provided.

Table of contents

List of Figures Page

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Tables Page

Chapter 2

Chapter 3

List of Abbreviations

MNDO – Modified Neglect of Differential Overlap
AM1 – Austin Model 1
PM3 – Parameterized Model number 3
NDDO – Neglect of diatomic differential overlap
SAM1 – Semi-Ab-initio Model 1
HF – Hartree-Fock
MO – Molecular Orbital
DFT – Density Functional Theory
B3LYP – Becke's three parameter hybrid exchange potential combined with Lee-Yang-Parr correlational functional
X3LYP – Extended three parameter hybrid exchange potential combined with Lee-Yang-Parr correlational functional
STO – Slater-type orbitals
SCF – Self-Consistent Field
PCM – Polarized Continuum Model
IPCM – Isodensity Polarized Continuum Model
SCIPCM – Self-Consistent Polarized Continuum Model
IEFPCM – Integral Equation Formalism Polarized Continuum Model
CPCM – Conductor-like Polarized Continuum Model
COSMO – Conductor-like Screening Solvation Model
UA0 – United Atom Topological Model
UAHF – United Atom Hartree-Fock
UAKS – United Atom Kohn-Sham
IR – Infrared
AIM – Atoms in Molecules
NBO – Natural Bond Orbital
NAO – Natural Atomic Orbital
NHO – Natural Hybrid Orbital
NLMO – Natural localized Molecular Orbital
LMO – Localized Molecular Orbital

NTA – Nitrilotriacetic Acid

NTPA – Nitrilotripropionic Acid

TC – Thermodynamic Cycle

IRn – Isodesmic Reaction

ZPVE – Zero point vibrational energy

SPC – Single Point Calculation

IDA – Iminodiacetic Acid

MIDA – N-Methyliminodiacetic Acid

EIDA – N-Ethyliminodiacetic Acid

PIDA – N-Propyliminodiacetic Acid

HIDA – N-(2-Hydroxyethyl)iminodiacetic Acid

CSD – Cambridge Structural Database

MM/MD – Molecular Mechanics/Molecular Dynamics

$E^{(2)}$ – Second order perturbation energy

QTAIM – Quantum Theory of Atoms in Molecules

BCPs – Bond Critical Points

CCPs – Cage Critical Points

RCPs – Ring Critical Points

ρ – Electron charge density

$\nabla^2\rho$ – Laplacian of the electron charge density

ADF – Amsterdam Density Functional

1. Introduction

From the time computers were developed they had the potential to perform tremendous tasks. Over the years computers continued getting faster and more advanced. We are now at a point in time where one can easily obtain a computer whose computational speed was unimaginable a few years ago. With this increase in speed it is now possible to perform computational calculations on molecules (also known as computational chemistry) within a reasonable time. This has brought about an era where calculations that once took months to complete can now be done in a few days, or better yet in a few hours (depending on the number of processors and amount of Random Access Memory (RAM) available). Computational chemistry, however, does not only involve the use of computers to model or simulate molecular behaviour. This is a field, which deals with how the fundamental equations, used to perform the simulations, are derived from the Schrödinger equation or from classical mechanics, amongst other things.[1] Computational chemistry is focused on obtaining results, which are relevant to specific chemical problems and not directly developing new theoretical methods.[2] It must be kept in mind that the ability to perform a calculation is no guarantee that the results obtained can be trusted.[2] Thorough analysis of results must be done in order to make sure that they are reliable.

It is important to mention at this point that the field of computational chemistry involves the use of a computer as an “experimental” tool, much like, for example, Infrared and Raman spectrometers. Being a subfield of physical chemistry, it is not surprising then that computational chemistry focuses on the analysis of physical properties of molecules. The studies conducted in this work focus on analysis of physical aspects at the atomic level in order to explain, in particular, why some metal-ligand complexes are more stable than others. According to previous research[3] so-called hydrogen clashes were responsible for the decrease in stability. There was unfortunately no physical (computational) data available to support this hypothesis. According to this work it was discovered that hydrogen clashes were in fact not responsible for the difference in stability and as one will discover, from the chapters that follow, there has been a sufficient amount of physical data obtained to support this finding. The aspects considered were metal-ligand bond lengths, thermochemical data such as Gibbs free energy, Natural Bond Orbital (NBO) and Bader charges and electron density, all of which were determined by making use of computational software packages discussed in detail in the chapters that follow.

Computation of inorganic molecules is far more intensive than that of organic molecules. This is due to the fact that organic molecules are mostly tetrahedral in shape (around a carbon or nitrogen atom, for example); therefore computation in terms of geometrical shape is predominantly tetrahedral. Inorganic molecules, on the other hand, are more complex as a 4-coordinated metal, for example, may either be tetrahedral or square planar, and a 5-coordinated metal may have square pyramidal or trigonal bipyrimidal structure.[2] This variation for metal coordinated molecules results in large increases in processing time. It is therefore advisable that initial computations always be done on free ligand species (i.e. molecules not containing a coordinated metal ion), before performing calculations on molecules with coordinated metal ions.

When performing computational calculations (geometry optimization) there are a number of aspects that have to be considered before submitting a computational job, these include computational methods, basis sets, solvation models, and frequencies. All of these aspects will be addressed in detail in the sections that follow.

1.1. Computational Method

There are a number of different methods that one can choose from, but only most commonly used methods are briefly discussed here.

Semi-Empirical Method. These methods solve an approximate form of the Schrödinger equation that depends on having appropriate parameters available for the type of chemical system under investigation.[4] There are a number of different semi-empirical methods, with some of the best known being MNDO, AM1 and PM3.[4-7] MNDO (Modified Neglect of Differential Overlap) is a method used for the quantum calculation of molecular electronic structure,[8, 9] AM1 (Austin Model 1) is a method that reduces the repulsion of atoms at close separation distances,[10] and PM3 (Parameterized Model number 3) uses the same formalism and equations as AM1, except (i) PM3 uses two Gaussian functions for the core repulsion function, instead of the variable number used by AM1; (ii) numerical values of parameters used in PM3 are different to those used in AM1.[11, 12] MNDO can cover a wide variety of elements, such as aluminium, silicon, germanium, tin, bromine and lead. A serious limitation of MNDO is its inability to accurately model intermolecular systems involving hydrogen bonds (for example, the heat of formation of the water dimmer is very low in MNDO).[13] AM1 and PM3 differ from MNDO only in the choice of empirical core repulsion function and they can be viewed as

attempts to explore the limits of the MNDO electronic structure model through extensive parameterization.[5] These three methods (MNDO, AM1, PM3) are all based on the NDDO (Neglect of diatomic differential overlap) integral approximation, which means that they were all designed to neglect all three-center and four-center two-electron integrals.[5] NDDO approximations are employed to integrals that involve Coulomb interactions. One of the more recent semi-empirical methods is SAM1 (Semi-Ab-initio Model 1)[14] which uses a standard STO-3G Gaussian basis set (discussed later in this chapter) to evaluate the electron repulsion integrals which were then scaled, partly to enable some of the effects of electron correlation to be included and partly to compensate for the use of a minimal basis set.[13] The number of parameters in SAM1 is not greater than in AM1 and fewer than in PM3, but it does take longer to run due to an explicit consideration of only the valence electrons of a system; the core electrons are subsumed into the nuclear core.[13] Semi-empirical methods are considerably computationally inexpensive as compared to *ab initio* methods. It should, however, be kept in mind that semi-empirical methods are not very accurate (they are poorly parameterized for many atoms and cannot represent interactions, such as hydrogen bonding, accurately) and in most cases should only be used as an initial first step method to obtain a starting structure for subsequent Hartree-Fock or Density Functional Theory optimizations.[4] By performing calculations on a molecule with the aid of semi-empirical methods, more often than not one will find that the output obtained will be very far from that of the optimum structure, i.e. the bond distances and angles will differ largely from that of a crystal structure for example, and it is also possible that the molecular geometry of the molecule completely changes.

Ab initio methods. *Ab initio* (Latin for 'from the beginning') can be considered as the direct opposite of semi-empirical methods. *Ab initio* methods are focused less on making short-term solutions, and more on long-term development of a rigorous methodology;[6] they are being developed continually. Their computations are based solely on the laws of quantum mechanics and on the values of physical constants such as speed of light, masses and charges of electrons and nuclei and Planck's constant.[4] At present *ab initio* methods are the most accurate at solving the Schrödinger equation and it is for this reason that calculations based on these methods are much more computationally intensive than their Hartree-Fock or Density Functional Theory counterparts. *Ab initio* methods are, however, not the 'ultimate' theory, they are just a stepping stone on the way to more sophisticated

theories (i.e. theories that come closer to the accurate solution of the Schrödinger equation).[6]

Hartree-Fock (HF) Theory. It is the basis of molecular orbital (MO) theory, which stipulates that each electron's motion can be described by a single-point function (orbital) which does not depend explicitly on the instantaneous motions of the other electrons.[15] The HF approximation is not only the cornerstone of almost all conventional, i.e. wavefunction based quantum chemical methods, it is also of great conceptual importance.[16] It can be described as an approximate theory/method used for the determination of the ground-state wave function and ground-state energy of a quantum many-body system.[5] HF theory is typically applied to the solution of the electronic Schrödinger equation of atoms, molecules and solids. It is very useful in providing initial, first-level predictions for many systems and is reasonably good at computing the structures and vibrational frequencies of stable molecules and some transition states. As such, it is considered a good base-level theory.[4] However, since this is a theory that neglects electron correlation it is insufficient for accurate modelling of the energetics of reactions and bond dissociation.[4] HF overemphasizes bonding, as a result all force constants are too large, and thus so are all frequencies.[6] Application of a constant scaling factor to the HF frequencies improves accuracy enormously,[7] but this is still not accurate enough as compared to results obtained with Density Functional Theory.

Density Functional Theory (DFT). This is an approach to the electronic structure of atoms and molecules which has enjoyed a dramatic surge of interest since the late 1980s and 1990s.[13,17] Its popularity stems from the fact that it is the most cost-effective method used to achieve a given level of accuracy at the lowest cost,[6] and unlike HF theory, DFT does provide electron correlation.[4] The central idea underpinning DFT is that there is a relationship between the total electronic energy and the overall electronic density. Hohenberg and Kohn[18] developed a theorem which showed that the ground-state energy and other properties of a system were uniquely defined by the electron density. The disadvantage of this theorem, however, was that there is no simplification over MO theory, since the final step is still solution of the Schrödinger equation and this is prohibitively difficult in most instances.[6] Kohn and Sham[19] later suggested a practical way to solve the Hohnberg-Kohn theorem for a set of interacting electrons. In their scheme Kohn and Sham apply an initial guess for the density from which a set of orbitals

can be derived, leading to an improved value for the density, which is then used in the second iteration, and so on until convergence is achieved.[13] DFT computes electron correlation via functionals of the electron density, where the functionals merely partition the electronic energy into several components which are computed separately: the kinetic energy, the electron-nuclear interaction, the Coulomb repulsion and an electron-correlation term accounting for the remainder of the electron-electron interaction.[4] There are a number of functionals available for DFT calculations and these include:

a) B1LYP,[20] B1PW91,[20] B1B95,[21] mPW1PW91[22] and PBE1PBE.[23] These functionals are described as being one parameter models, which combine elements of HF theory with local density-functional approximations for dynamical correlation,[6] and this is discussed thoroughly by Becke.[24] The functionals have, however, attracted less attention than their three-parameter counterparts (discussed below), because they depend explicitly on the kinetic energy density in addition to the density and its gradient, which complicates the implementation into standard molecular structure computer programs.[16]

b) B3LYP (Becke three parameter hybrid exchange potential combined with Lee-Yang-Parr correlation functional)[25] and B3PW91 (Becke three parameter hybrid exchange potential combined with the Perdew-Wang correlation functional).[6] These functionals are known as hybrid functionals, which define the exchange functional (B3) as a linear combination of HF, local and gradient-corrected exchange terms; this exchange functional is then combined with a local and/or gradient-corrected correlation functional (LYP or PW91). To date B3LYP is one the most widely used functionals,[5] since its overall performance and accuracy is remarkably good. This shall be illustrated in the following chapters of this dissertation as B3LYP was utilized for all of the computational calculations performed in this work.

c) The extended three parameter hybrid exchange potential combined with Lee-Yang-Parr correlation functional (X3LYP) has been introduced only recently.[26-29] The functional predicts quantities, such as heat of formation, ionization potential, electron affinities, and total atomic energies with higher level of accuracies,[26] better than those predicted by B3LYP functionals.[29] It has also been shown that it outperforms the B3LYP functional when it comes to describing van der Waals and hydrogen bonded interactions.[28] Due to the fact that this is still a relatively new functional it was decided that for the purposes of these studies the better known

B3LYP functional would be used.

DFT, using gradient-corrected functionals, can give results for a wide variety of properties that are competitive with, and in some cases superior to, *ab initio* calculations.[13] Gradient-corrected functionals are required for the calculation of relative conformational energies and the study of intermolecular systems, particularly those involving hydrogen bonding.[30]

1.2. Basis Set

A basis set is the mathematical description of the orbitals within a system used to perform the theoretical calculation. The larger the basis set, the fewer the number of restrictions on the locations of the electrons in space, and the more accurately the orbitals of atoms are approximated.[4] There are a number of pre-defined basis sets which can be chosen from; they are classified by the number and types of basis functions that they contain. Basis functions are assigned to each atom within a molecule to approximate its orbitals, and they are themselves composed of a linear combination of Gaussian functions.[4]

Minimal Basis Sets. These use fixed-size atomic-type orbitals. They normally include all the atomic orbitals in each atom. Thus, for hydrogen and helium a single s-type function would be required; for the elements from lithium to neon the 1s, 2s and 2p functions are used, and so on.[13] STO-3G (STO – Slater-type orbitals)[31] is an example of a minimal basis set, which uses three gaussian primitives per basis function and it approximates Slater orbitals with gaussian functions.[4] The STO-3G basis set does perform remarkably well in predicting molecular geometries, though this is due in part to a fortuitous cancellation of errors.[13] Minimal basis sets are well known to have several deficiencies, one of them being that compounds containing atoms at the end of a period, such as oxygen and fluorine are described using the same number of basis functions as the atoms at the beginning of the period, despite the fact that they have more electrons.

Split Valence Basis Sets. These have two sizes, or more, of basis function for each valence orbital. For example, N-atom is represented as 1s,2s,2s´,$2p_x$,$2p_y$,$2p_z$,$2p_x$´,$2p_y$´,$2p_z$´ where the primed and unprimed orbitals differ in size. Examples of such basis sets would be 3-21G,[32] 6-31G[33] and 6-311G (which is called a triple split valence basis set).[4]

Polarized Basis Sets. The electron cloud in an isolated hydrogen atom is symmetrical, but when the hydrogen atom is present in a molecule the electrons are attracted towards the other nuclei. The distortion can be considered to correspond to mixing p-type character into the 1s orbital of the isolated atom to give a form of sp hybrid. In a similar manner, the unoccupied d orbitals introduce asymmetry into p orbitals.[13] The best solution to this problem is the introduction of polarization functions into the basis set, hence the existence of polarized basis sets. The polarisation functions have a higher angular quantum number and so correspond to p orbitals for hydrogen and d orbitals for the first- and second-row elements.[4, 13] Examples of polarized basis sets are 6-31G(d)[32] and 6-31G(d,p), also known as 6-31G* and 6-31G** respectively. These are usually used for calculations involving medium-sized systems.[4]

Diffuse Functions. The deficiency of the above mentioned basis sets described so far is their inability to deal with species, such as anions and molecules containing lone pairs, which have a significant amount of electron density away from the nuclear centres. To remedy this deficiency highly diffuse functions are added to the basis set.[13] Diffuse functions are large size versions of s- and p-type functions, allowing orbitals to occupy a larger region of space.[4] Examples of these basis sets would be 6-31+G(d) (which is a 6-31G(d) basis set with diffuse functions added to heavy atoms) and 6-31++G(d) (which adds diffuse functions to hydrogen atoms as well).[4]

High Angular Momentum Basis Sets. These can be considered as being large basis sets. An example would be 6-311++G(3df,2pd),[34] which contains three sets of valence region functions, diffuse functions on both heavy atoms and hydrogens, and multiple polarization functions, i.e. three sets of d functions and one set of f functions for first-row atoms and two sets of p functions and one set of d functions for hydrogen. These basis sets are typically useful for describing the interactions between electrons in electron correlation methods.[4]

1.3. Solvation models

Computational calculations are usually performed under vacuum, but experimentally not all reactions take place in gas phase, especially those that involve biological molecules. In addition, vacuum calculations can lead to significant problems since a vacuum boundary tends to minimise the surface area and so may distort the shape of the

system if it is non-spherical.[13] It is for these reasons that solvent effects on molecular energies, structures, and properties must be accounted for in any theoretical treatment of solvent influenced molecules (such as biological molecules) and in order to do so a number of solvation models can be chosen from. The models belong to a family known as the Self-Consistent Field (SCF) method, which considers the solvent as a uniform polarisable medium with a dielectric constant ε, with the solute placed in a suitably shaped hole in the medium.[2]

Polarized Continuum Model (PCM). This model, which was first formalized by Miertus et al,[35] defines the cavity as a series of interlocking atomic spheres having radii 20% larger than their tabulated Van der Waals radii.[6] PCM permits the computation of the effect of solvation on the energy,[36] nuclear gradients,[37] and even frequency-dependent polarizibilities.[38] There are two slight complications with the PCM approach. The first of these arises as a consequence of representing a continuous charge distribution over the cavity surface as a set of single point charges. When calculating the electrostatic potential due to the charges on the surface elements one must exclude the charge for the current surface element. To include it would cause the charges to diverge rather than converge. The second complication arises because the wavefunction of the solute extends beyond the cavity. Thus the sum of the charges on the surface is not equal and opposite to the charge of the solute.[13] The first complication is overcome by determining the charge on the surface element using the Gauss theorem,[13] whereas in the case of the second complication the charge distribution on the surface needs to be scaled so that it is equal and opposite to the charge of the solute.

Isodensity Polarized Continuum Model (IPCM). The cavity is defined by a calculation of the gas phase isodensity surface.[39] An isodensity surface is a very natural, intuitive shape for the cavity since it corresponds to the reactive shape of the molecule to as great a degree as possible.[4] This model tends to be considerably less stable in implementation than the original PCM process, and can be subject to erratic behaviour in charged systems.[6]

Self-Consistent Isodensity Polarized Continuum Model (SCIPCM). This defines the cavity as an isosurface coupled to electron density. Therefore the effects of solvation are folded into the iterative SCF computation rather than comprising an extra step

afterwards.[4] The model thus accounts for the full coupling between the cavity and the electron density and includes coupling terms that IPCM neglects.[4] However, as is the case with IPCM, this model can be subject to erratic behaviour in charged systems.[6]

Integral Equation Formalism Polarized Continuum Model (IEFPCM). This is an example of a formalism which treats different solvation systems within a common integral equation-like approach.[40] IEFPCM treats solute-solvent systems very differently as compared to other quantum mechanical (QM) continuum methods with almost the same computational expense.[40]

Conductor-like Polarized Continuum Model (CPCM). With this model the solute cavities are modelled on the optimized molecular shape, and include both electrostatic and nonelectrostatic contributions to energies.[41]

Conductor-like Screening Solvation Model (COSMO). This is an interesting variant on the PCM method.[42-44] The cavity is considered to be embedded in a conductor with an infinite dielectric constant. The advantage of this is that screening effects in an infinitely strong dielectric, i.e. a conductor, are much easier to handle.[13] A small correction to the results for this conductor can provide the appropriate value for water, which has a high dielectric constant.

Once the solvation model has been chosen, it is still necessary to choose the atomic radii that should be used to define the solute cavity.

United Atom Topological Model (UA0). This is the default set of radii, which is automatically used in Gaussian with the chosen solvation model, if not specified otherwise.[45] Within this algorithm, the effective atomic radius for an element is determined by the product of α and its corresponding Van der Waals' (VdW) radius. α is defined as the electrostatic scaling factor by which the sphere radius is multiplied, where the default specification is $\alpha = 1.0$.[45]

United Atom Hartree Focke (UAHF). With this model the standard characteristics of the PCM, as described by Barone et. al.,[46] remain unchanged, with only the VdW radii for atoms and atomic groups being optimized. The parameters of this model were imposed by

two initial conditions; the hydrogens do not have individual spheres, but they are included in the spheres of heavy atoms to which they are bonded, and elements of each periodic table row have the same "basic" radius, modified by the molecular environment.[46]

United Atom Kohn-Sham (UAKS). With this model the cavity is defined exactly the same as for UAHF. The difference between the two radii, i.e. UAHF and UAKS, lies in the fact that UAHF radii are optimized for the HF/6-31G(d) level of theory and UAKS radii are optimized for the PBE0/6-31G(d) level of theory.[46]

There are other radii options available, but the three listed above are considered as the major models, and these were the models that were utilized for the purposes of these studies.

Solvation models are, however, not the only way to account for solvent effects. Another mean to account for solvent effects is to add explicit water molecules to the calculation.[47] With this approach water molecules are placed at different positions relative to the studied molecule where it is expected that interaction might take place. This method of accounting for solvent effects is considerably disadvantageous as it is never known exactly were the water molecules need to be placed. Another disadvantage is that numerous water molecules may have to be included in the computation in order to account for the solvent effects and the more water molecules that are added to the computation the more intensive the computation becomes. With this in mind, it is advisable that, if a molecule in solvent is to be studied, one of the implicit solvent models provided above should be used for this purpose, as this would save time and yield results that are often comparable to those obtained experimentally.

1.4. Frequencies

Frequency calculations account for the vibrations in molecular systems; in so doing they can be used to compute the vibrational spectra of molecules in their ground state.[4] These calculations can serve a number of different purposes:

a) Prediction of Infrared (IR) and Raman spectra of molecules (both the frequencies and intensities).
b) Computation of force constants for a geometry optimization.
c) Compute zero-point vibration and thermal energy corrections to total energies.

d) Compute thermodynamic quantities such as Gibbs free energy, Enthalpy and Entropy.

e) Predict if the system studied lies at a minimum on the potential energy surface (negative frequencies are generated if the system is not at a minimum). However, even if the system lies at a minimum, it does not mean that it is at a global minimum. In most cases, especially with systems that have numerous conformations, the system will only be at a local energy minimum on the potential energy surface, unless of course one performs a thorough conformational analysis prior to optimization.

Frequency calculations are only valid at stationary points on the potential energy surface, which implies that frequency calculations should only be performed on optimized structures.[4] It is therefore necessary to run geometry optimization prior to performing a frequency calculation. The best way to do this is to perform both optimization and frequency calculations on all systems studied; this is exactly what was done for all of the systems that were studied.

1.5. Atoms in Molecules

Atoms are objects in real space. Theory defines them through a partitioning of real space as determined by the topological properties of a molecular charge distribution, that is, by its form in real space.[48] The constancy in the properties of an atom as defined in this theory, including its contribution to the total energy of a system, is observed to be directly determined by a corresponding constancy in its distribution of charge. When the charge distribution over an atom is the same in two different molecules, i.e. when the atom or some functional grouping of atoms is the same in the real space of two systems, then it makes the same contribution to the total energy and other properties in both systems.[49] It is this direct relationship between the spatial form of an atom and its properties that one is able to identify them in different systems. This does not imply that atoms are always transferable. It is simply that, in the limit of an atom being transferable between systems, the relationship between the form and properties of an atom is most evident.[48] The properties of atoms in molecules (AIM) can be experimentally measured and, when the atoms are transferable, these properties can be determined with impressive precision. Bader shows, in his book,[48] that there is an excellent correlation between experimental and theoretically determined properties of AIM. For example, researchers at

the National Bureau of Standards determined that the incremental increase in the standard heat of formation of a normal hydrocarbon per methylene group is –20.6 ± 1.3 kJ/mol.[48] The theory of AIM not only defines a methylene group, whose distribution of charge in real space up to its atomic boundaries is transferable between hydrocarbon molecules, but also defines the energy of this group. This energy is as constant and transferable as the group itself and it accounts for the observed additivity of the energy in this series of molecules.[48]

The theory of AIM provides an alternative way to partition the electrons between the atoms in a molecule. The theory is based upon the concept of a gradient vector path, which is a curve around the molecule such that it is always perpendicular to the electron density contours.[13] Some of these gradient paths terminate at the atomic nuclei, whereas others are attracted to points (called critical points) that are not located at the nuclei; particularly common are the bond critical points, which are located between two bonded atoms linked via a bond path.[13,50] If the bond paths are linked so as to form a ring of bonded atoms then a ring critical point is found in the interior of the ring. A ring is defined as a part of a molecular graph which bounds a ring surface.[48] If the bond paths are so arranged as to enclose the interior of a molecule with ring surfaces, then a cage critical point is found in the interior of the resulting cage.[48] A cage is a part of a molecular graph which contains at least two rings, such that the union of the ring surfaces binds a region which contains a cage critical point.[48]

Wiberg and Rablen[51] found that the charges obtained with the AIM method were relatively invariant to the basis set. The charges were also consistent with the experimentally determined C-H bond dipoles in methane (in which the carbon is positively charged) and ethyne (in which the carbon is negatively charged).

1.6. Natural Bond Orbital

Natural bond orbital (NBO) analysis originated as a technique for studying hybridization and covalency effects in polyatomic wavefunctions.[52] It comprises a sequence of transformations from the input basis set to various localized basis sets [natural atomic orbitals (NAOs), natural hybrid orbitals (NHOs), natural bond orbitals (NBOs) and natural localized molecular orbitals (NLMOs)].[52]

$$\text{Input basis} \rightarrow \text{NAOs} \rightarrow \text{NHOs} \rightarrow \text{NBOs} \rightarrow \text{NLMOs} \qquad (1)$$

A NAO is defined as a valence-shell atomic orbital derived from the diagonalization of the localized block of the full density matrix of a given molecule, associated with basis functions $\chi_1(A)$ on that atom, and fulfilling the simultaneous requirement of orthonormality and maximum occupancy.[53, 54] Although in an isolated atom, NAOs coincide with natural orbitals, in polyatomic molecules the NAOs mostly retain one-centre character.[53, 54] Consequently, NAOs are considered optimal for describing the molecular electron density around one-centre in polyatomic molecules. Also, NHOs[55] result from a symmetrically orthogonalized directed hybrid orbital derived through unitary transformation of NAO centred on a particular atom.[53, 54] According to the simple bond orbital picture,[56] a NBO is defined as an orbital formed from NHOs. Therefore, for a localized σ-bond between atoms A and B, the NBO is defined as:

$$\sigma_{AB} = c_A h_A + c_B h_B \qquad (2)$$

where c_A and c_B are the polarization coefficients of atoms A and B. h_A and h_B are natural hybrids centred on atoms A and B.

So, the NBOs closely correspond to the picture of localized bonds and lone pairs as basic units of molecular structure, and therefore it will be possible to conveniently interpret *ab initio* wavefunctions in terms of the classical Lewis structure concepts by transforming these functions to the NBO forms.[52]

Finally, the natural localized molecular orbitals (NLMOs)[57] procedure is a direct extension of the NAO and NBO procedures, and it builds up the localized molecular orbitals (LMOs) step-by-step from the NAOs and NBOs. This approach not only lends the method great efficiency (it is noniterative, involving only simple matrix operations), but also allows one to gain great direct insight into the electronic origin of the "delocalization tails"[58] of LMOs (through the examination of the NBO composition of each NLMO).[52]

All the information provided up to this point serves as a brief guideline to individuals who wish to perform computational analysis, especially those who have never done computation before but wish to apply it to systems that they are studying. It is for this reason that a thorough mathematical explanatory approach was not adopted. With the remainder of this chapter a brief description of the molecules studied as well as the objectives which were considered, will be provided.

1.7. Molecules studied

Studies were focused on nitrilotriacetic acid (NTA) and nitrilotripropionic acid (NTPA). Diagrams depicting the deprotonated forms of ligands are given in Figures 1.1 and 1.2, respectively.

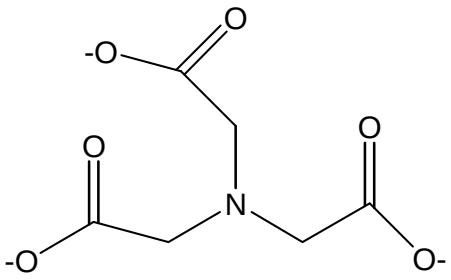


Figure 1.1: Free ligand NTA^{3-}

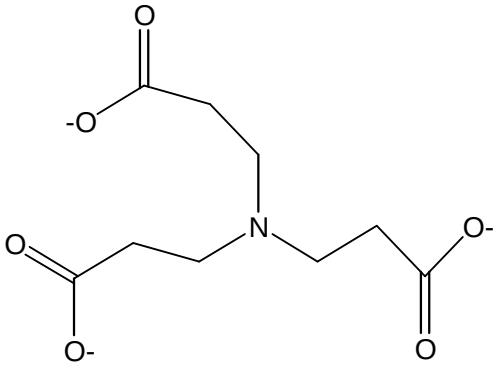


Figure 1.2: Free ligand $NTPA^{3-}$

NTA is known to possess excellent chelating properties.[59-61] It is commonly found as a multi-dentate ligand in many metal-chelate compounds.[62-72] As can be seen from Figure 1.1, NTA can chelate to metal ions through the three carboxylate oxygen atoms as well as the nitrogen atom, thereby making it a tetradentate ligand. It has, however, been recorded that NTA can also function as a tridentate ligand with one of the oxygens being protonated, while the remaining two oxygen atoms and the nitrogen atom coordinate to the metal ion[69,70] resulting in formation of the M(HL) complex. Chelating agents are known to be very useful when it comes to biological applications,[3] and NTA is no exception to this. It is commonly used in biochemistry and medicine[64,73] and is gaining increased use in biotechnology, particularly in the protein purification technique known as immobilized metal ion.[74] Since it is a multi-dentate ligand, it can complex to most metal ions, including those that are possibly harmful to the biological system, and by doing so the harmful contaminates can then be removed from the system.

Aside from having medicinal properties, NTA has also been used in industry for boiler water treatments,[59] cleaning instruments containing toxic metals and radionuclide species,[60] various redox desulfurization processes[75] and removal of lead from contaminated soil.[76] It is well known that NTA chelates to many metal ions including Al(III), Be(II), Bi(III), Ca(II), Co(II), Cr(II), Cu(II), Fe(II), Nd(III), Ni(II), Pb(II), Ti(III), W(VI), Zn(II), and Zr(IV)[77,78] forming ML, ML_2, M(HL), MLOH and M_2L complexes. NTPA, on the other hand, has only been reported to complex to Be(II), Co(II), Ni(II), Cu(II), Zn(II), Cd(II)[78] and Pb(II)[79] forming predominately ML complexes.

Experimental protonation constants have been reported as, $pK_H^{(1)}$ = 8.94, $pK_H^{(2)}$ = 2.31, $pK_H^{(3)}$ = 1.78, $pK_H^{(4)}$ = 1.00 for NTA and $pK_H^{(1)}$ = 9.59, $pK_H^{(2)}$ = 4.28, $pK_H^{(3)}$ = 3.77, $pK_H^{(4)}$ = 2.71 for NTPA.[78] These values were recorded at 25 ^oC and ionic strength = 0.5. With the chapters that follow the theoretical estimates of protonation constants for both NTA and NTPA is considered and compared with those obtained experimentally.

For cases where the metal ion is common for both NTA and NTPA, most complexes formed with NTA have stability constants that are several orders of magnitude larger than complexes formed with NTPA.[78] It is only in the case of Be(II) that the stability constant of NTPA is larger than that of NTA. This is due to the fact that beryllium is a small metal ion, which provides additional space for the six-membered rings of NTPA without any hydrogen clashes being present. Figure 1.3 provides a graphical representation of the difference in stability constants for different metal complexes of NTA and NTPA.

The following objectives were considered with respect to these studies:

a) Theoretical prediction of protonation constants for NTA and NTPA.
b) Theoretical prediction of stability constants for complexes of NTA and NTPA, with focus being on complexes of Zn(II) and Ni(II), respectively. Looking at the relationship in Figure 1.3 it is seen that Zn(II) and Ni(II) have stabilities, $\Delta\log K = \log K_{ML}(\text{NTPA}) - \log K_{ML}(\text{NTA})$, that are slightly smaller when compared with $\Delta\log$ K values for Cd(II) and Pb(II), so the information and conclusions obtained for the Zn(II) and Ni(II) complexes of NTA and NTPA might also be applicable to complexes of Cd(II) and Pb(II). In addition Zn(II) and Ni(II) are well parameterized in Gaussian, whereas parameters for Pb(II) are still under development. It is because of these reasons that Zn(II) and Ni(II) were chosen for these studies.
c) Theoretical explanation of why complexes formed with NTA are orders of magnitude larger than complexes formed with NTPA.

These objectives will be achieved by focusing on analysis of physical properties of the molecules, as well as trends that can help correlate the computationally generated values with those determined experimentally.

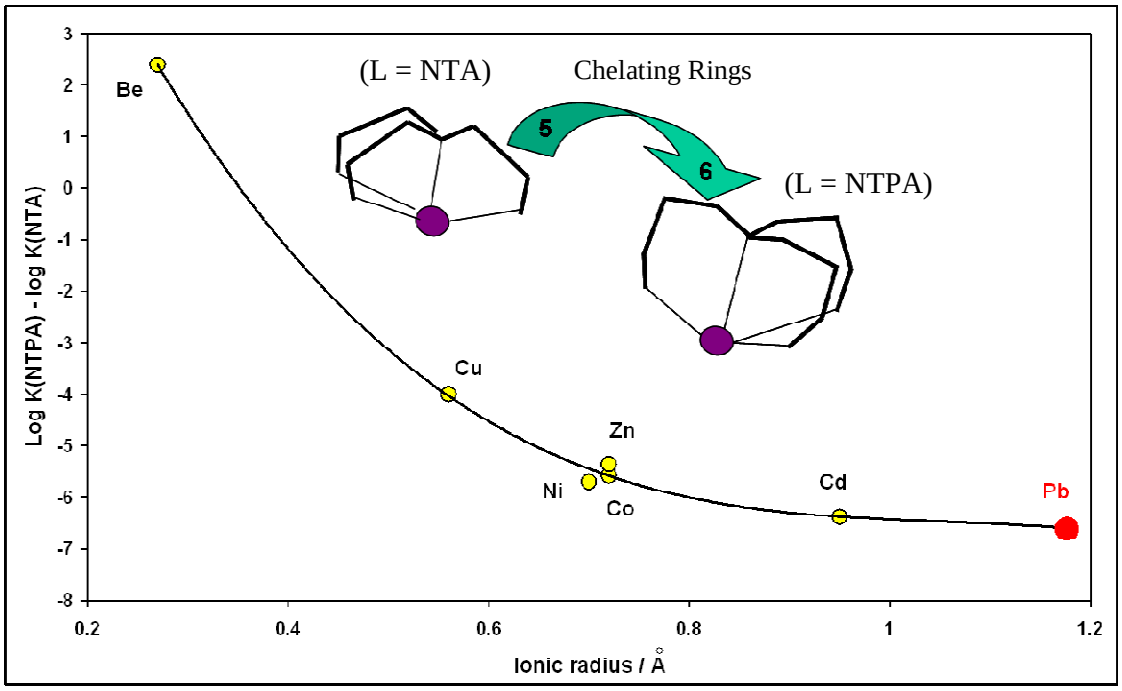


Figure 1.3: Illustration of the difference in stability of complexes formed with NTA and NTPA. All stability constants, except for Pb,[79] are from Martell and Smith compilation.[78] The value of log K_1 for Pb was obtained from Cukrowski (and it was established from voltammetric experiments)[79]

1.6. References

1. J. Simons, 'An Introduction to Theoretical Chemistry', Cambridge University Press, 2003.

2. F. Jensen, 'Introduction to Computational Chemistry', John Wiley and Sons Ltd., 1999.

3. A. E. Martell, R. D. Hancock, 'Metal Complexes in Aqueous Solutions', Plenum Press, New York, 1996.

4. J. B. Foresman, A. Frisch, 'Exploring Chemistry with Electronic Structure Methods', Gaussian, Inc., Pittsburgh, PA, 1996.

5. C. E. Dykstra, G. Frenking, K. S. Kim, G. E. Scuseria, 'Theory and Applications of Computational Chemistry: The First Forty Years', Elsevier B. V., 2005.

6. C. J. Cramer, 'Essentials of Computational Chemistry', John Wiley & Sons Ltd, 2002.

7. J. A. Pople, A. P. Scott, M. W. Wong, L. Radom, *Isr. J. Chem.* **1993**, *33*, 345.

8. M. J. S. Dewar, W. Thiel, *J. Am. Chem. Soc.* **1977**, *99*, 4899.

9. M. J. S. Dewar, W. Thiel, *J. Am. Chem. Soc.* **1977**, *99*, 4907.

10. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, *J. Am. Chem. Soc.* **1985**, *107*, 3902.

11. J. J. P. Stewart, *J. Comput. Chem.* **1989**, *10*, 209.

12. J. J. P. Stewart, *J. Comput. Chem.* **1989**, *10*, 221.

13. A. R. Leach, 'Molecular Modelling: Principles and Applications', Pearson Education Limited, 2001.

14. M. J. S. Dewar, C. Jie, J. Yu, *Tetrahedron* **1993**, *49*, 5003.

15. C. D. Sherrill, 'An introduction to Hartree-Fock Molecular Orbital Theory', School of chemistry and biochemistry, Geogia Institute of Thechnology, 2000, http://vergil.chemistry.gatech.edu/notes/hf-intro/hf-intro.pdf.

16. W. Koch, M. C. Holthausen, 'A Chemist's Guide to Density Functional Theory', Wiley-VCH, 2002.

17. R. G. Parr, *Annual Rev. Phys. Chem.* **1983**, *34*, 631.

18. P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, *B136*, 864.

19. W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *A140*, 1133.

20. C. Adamo, V. Barone, *Chem. Phys. Lett.* **1997**, *274*, 242.

21. A. D. Becke, *J. Chem. Phys.* **1996**, *104*, 1040.

22. C. Adamo, V. Barone, *J. Chem. Phys.* **1998**, *108*, 664.

23. K. Burke, M. Ernzerhof, J. P. Perdew, *Chem. Phys. Lett.* **1997**, *265*, 115.

24. A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 1372.

25. P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623.

26. J. Cerny, P. Hobza, *Phys. Chem. Chem. Phys.* **2005**, *7*, 1624.

27. J. T. Su, X. Xu, W. A. Goddard III, *J. Phys. Chem. A* **2004**, *108*, 10518.

28. X. Xu, W. A. Goddard III, *P. Natl. Acad. Sci. USA* **2004**, *101*, 2673.

29. X. Xu, Q. Zhang, R. P. Muller, W. A. Goddard, *J. Chem. Phys.* **2005**, *122*, 1.

30. F. Sim, A. St-Amant, I. Papai, D. R. Salahub, *J. Am. Chem. Soc.* **1992**, *114*, 4391.

31. A. P. J. Jansen, R. A. van Santen, *J. Phys. Chem.* **1990**, *94*, 6764.

32. M. J. Shephard, M. N. Paddon-Row, *J. Phys. Chem. A* **2000**, *104*, 11628.

33. Q. Zhu, P. L. LeBreton, *J. Am. Chem. Soc.* **2000**, *122*, 12824.

34. M. Lewis, Z. Wu, R. Glaser, *J. Phys. Chem. A* **2000**, *104*, 11355.

35. S. Miertus, E. Scrocco, J. Tomasi, *Chem. Phys.* **1981**, *55*, 117.

36. E. Cances, B. Mennucci, J. Tomasi, *J. Chem. Phys.* **1997**, *107*, 3032.

37. E. Cances, B. Mennucci, *J. Chem. Phys.* **1998**, *109*, 249.

38. B. Mennucci, C. Amovilli, J. Tomasi, *Chem. Phys. Lett.* **1998**, *286*, 221.

39. A. L. Speelman, J. G. Gillmore, *J. Phys. Chem. A* **2008**, *112*, 5684.

40. B. Mennucci, E. Cances, J. Tomasi, *J. Chem. Phys. B* **1997**, *101*, 10506.

41. M. D. Liptak, G. C. Shields, *J. Am. Chem. Soc.* **2001**, *123*, 7314.

42. A. Klamt, G. Schuurmann, *J. Chem. Soc., Perkin Trans.* **1993**, *2*, 799.

43. A. Klamt, *J. Chem. Phys.* **1995**, *99*, 2224.

44. A. Klamt, V. Jonas, T. Burger, J. C. W. Lohrenz, *J. Phys. Chem.* **1998**, *102*, 5074.

45. W. H. Mu, G. A. Chasse, D. C. Fang, *Int. J. Quantum Chem.* **2008**, *108*, 1422.

46. V. Barone, M. Cossi, J. Tomasi, *J. Chem. Phys.* **1997**, *107*, 3210.

47. W. Sang-Aroon, V. Ruangpornvisuti, *Int. J. Quantum Chem.* **2008**, *108*, 1181.

48. R. F. W. Bader, 'Atoms in Molecules A Quantum Theory', Clarendon Press, Oxford, 1990.

49. R. F. W. Bader, P. M. Beddall, *J. Chem. Phys.* **1972**, *56*, 3320.

50. R. F. W. Bader, H. Essen, *J. Chem. Phys.* **1984**, *80*, 1943.

51. K. B. Wiberg, P. R. Rablen, *J. Comput. Chem.* **1993**, *14*, 1504.

52. A. E. Reed, L. A. Curtiss, F. Weinhold, *Chem. Rev.* **1988**, *88*, 899.

53. A. Ebrahimi, F. Deyhimi, H. Roohi, *J. Mol. Struct. (Theochem)* **2003**, *626*, 223.

54. M. S. Sadjadi, B. Sadeghi, K. Zare, *J. Mol. Struct. (Theochem)* **2007**, *817*, 27.

55. J. P. Foster, F. Weinhold, *J. Am. Chem. Soc.* **1980**, *102*, 7211.

56. F. Weinhold, C. R. Landis, 'Valency and Bonding - A Natural Bond Orbital Donor-Acceptor Perspective', Cambrige University Press, 2005.

57. A. E. Reed, F. Weinhold, *J. Chem. Phys.* **1985**, *83*, 1736.

58. M. Levy, *J. Chem. Phys.* **1976**, *65*, 2473.

59. D. Malevich, Z. Wang, P. R. Tremaine, *J. Sol. Chem.* **2006**, *35*, 1303.

60. M. S. Vohra, A. P. Davis, *J. Colloid Interf. Sci.* **1997**, *194*, 59.

61. Y. Tomita, T. Ando, K. Ueno, *J. Phys. Chem.* **1965**, *69*, 404.

62. N. Shan, A. D. Bond, W. Jones, *Acta Cryst.* **2001**, *E57*, o811.

63. M. Kaneyosh, A. Bond, W. Jones, *Acta Cryst.* **1999**, *C55*, 1260.

64. L. C. Yu, S. L. Liu, E. X. Liang, C. L. Wen, *J. Coord. Chem.* **2007**, *60*, 2097.

65. B. Li, Y. Xu, Y. Dong, J. Chen, Z. Xu, *J. Chem. Cryst.* **2001**, *31*, 357.

66. I. N. Polyakova, A. L. Poznyak, V. S. Sergienko, *Crystallogr. Rep.* **2006**, *51*, 609.

67. I. N. Polyakova, V. S. Sergienko, A. L. Poznyak, *Crystallogr. Rep.* **2006**, *51*, 459.

68. V. S. Sergienko, *Crystallogr. Rep.* **2006**, *51*, 236.

69. P. A. Petrenko, M. Gdaniec, Y. A. Simonov, V. G. Stavila, A. P. Gulea, *Russ. J. Coord. Chem.* **2004**, *30*, 813.

70. I. N. Polyakova, A. L. Poznyak, O. A. Egorova, *Russ. J. Coord. Chem.* **2001**, *27*, 852.

71. L. A. Zasurskaya, I. N. Polyakova, T. N. Polynova, A. L. Poznyak, V. S. Sergienko, *Russ. J. Coord. Chem.* **2001**, *27*, 270.

72. M. A. Walters, V. Vapnyar, A. Bolour, C. Incarvito, A. L. Rheingold, *Polyhedron* **2003**, *22*, 941.

73. M. S. Grigoriev, C. D. Auwer, D. Meyer, P. Moisy, *Acta Cryst.* **2006**, *C62*, m163.

74. H. A. Ewais, *Int. J. Chem. Kinet.* **2008**, *40*, 103.

75. S. Loridant, M. Digne, M. Daniel, T. Huard, *J. Raman Spectrosc.* **2007**, *38*, 44.

76. E. R. Souaya, W. G. Hanna, E. H. Ismail, N. E. Milad, *J. Coord. Chem.* **2004**, *57*, 825.

77. E. Skrzypczak-Jankun, D. A. Smith, H. Maluszynska, *Acta Cryst.* **1994**, *C50*, 1097.

78. 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, US Department of Commerce, National Institute of Standards and Technology, 2004.

79. I. Cukrowski, *Unpublished data*.

Chapter 2

This is a typed copy of a paper entitled “Density functional theory in prediction of four stepwise protonation constants for nitrilotripropanoic acid (NTPA)”, *J. Phys. Chem. A*, **2009**, *113*, 3639-3647.

2. Protonation Constants of NTPA

2.1. Introduction

The protonation/dissociation property of a compound is very important in chemistry, biology and material sciences, since the ability of a compound to donate or accept a proton is fundamental to understanding many chemical and biochemical processes.[1, 2] An example in medicinal chemistry is the ability of drugs to pass biological membranes as well as their potential to interact with intracellular receptors, both of which are affected by the readiness of the drug to accept or donate a proton.[3] Availability of dissociation (protonation constants as well as complex formation constants) is of fundamental significance as it allows for modelling of solution composition at required experimental conditions, such as blood plasma, natural waters, industrial effluents, etc. Experimental techniques provide protonation/dissociation constants typically with uncertainty on the second decimal place of the log unit and well-renown compilations, such as by Martell and Smith[4] or IUPAC[5] are readily available.

However, there is an intrinsic interest to develop theoretical procedures that would eventually match experimental accuracy. This is not only because not all chemical species are readily amenable or available (due to small quantity available) to experimental characterization, but also due to fundamental insight gained during theoretical modelling as well as prediction (or evaluation) of thermodynamic parameters related to the compound and reaction(s) in which this compound is involved. There have been a number of studies performed thus far where dissociation constants were predicted theoretically with considerable accuracy.[1-3, 6-50] Molecules studied include, dissociation constants of carboxylic acids,[1-3, 6-21] amides,[22, 23] bicarbonates,[16, 24] and proteins,[25] amongst others. However, it is important to emphasize that the molecules studied thus far were predominantly neutral or singly charged and on average the reported computed dissociation constants are within ±1.0 log unit when compared with experimental values. From the theoretical point of view, this might be regarded as excellent result particularly since the deviation in computed value by a log unit is equivalent to accuracy of energy computed in the range of a single kilocalorie per mol. Unfortunately, large drop in accuracy of computed dissociation constants is observed when multiple and particularly negative charges on a molecule are present,[51] hence there is very little published on stepwise multiple-dissociation constants in solvent (water).[24]

To date all the reported theoretical values are from thermodynamic cycles that typically involve two-step operation, namely (i) full gas-phase energy optimizations of

components involved in dissociation reaction, followed by (ii) a single point calculations in solvent (water) on those structures from which ΔG(aq) is obtained and used to calculate the dissociation constant at room temperature. Four thermodynamic cycles were recently evaluated by Liptak and Shields[11] on several singly-dissociable simple carboxylic acids using complete basis set and Gaussian-n models combined with Barone and Cossi's implementation of CPCM.[52] The high-level ab initio CBS-QB3[53] and CBS-APNO[54, 55] methods (using HF 6-31 G(d) and HF 6-31+ G(d)) with CPCM generated smallest inaccuracy of about 0.5 log unit when commonly used and the simplest thermodynamic cycle (without involving water molecule) was employed.[11] Similar accuracy was reported by Namazian et al[9] recently who used CPCM solvation energies at the B3LYP/6-31+G(d) level in conjunction with CBS-QB3 or G3 gas-phase energies of trifluoroacetic acid and its anion. It appears that there is a strong tendency to avoid water and simple proton in thermodynamic cycle computations and rather use most recent experimental values of –6.28 kcal/mol[12, 56] and –265.9 kcal/mol[16] or –263.98 kcal/mol[16] for the gas-phase Gibbs free energy of H^+ and solvation energy of H^+ in water, respectively.

Somewhat more elaborate approach was also tested[6, 7] where the thermodynamic cycle was combined with an isodesmic reaction involving acetic acid (and its dissociation reaction) as a single reference molecule used to theoretically predict numerous dissociation constants of mono-dissociable organic acids; the differences between experimental and theoretical values were between a fraction of and up to about 5 log units for some molecules. A new model of dissociation constant computation, called S03 and based on thermodynamic cycle, was proposed by Barone et al[2] who used PBE0/6-31+G(d,p) and PBE0/6-311+G(2d,2p) level of theories for full energy optimization and single point calculations in gas phase, respectively, as an initial step from which gas phase basicities of eight investigated organic acids were predicted. This was followed by estimation of solvation Gibbs energies by single point HF/6-31+G(d,p) calculations using geometries optimized in water at the PBE0/6-31+G(d,p) level. Although gas-phase basicities were of analytical accuracy, the dissociation constants in water were of commonly reported differences of ± 0.7 log units from experimental values with aqueous solution geometries; much worse results were obtained with gas-phase geometries used for single point calculations in solvent.

Recently DFT-based theoretical studies of metal complexes[57] were done in order to explore physical and molecular (and structural) properties controlling the strengths of

complexes formed. Interest was focused on understanding why 'small' but apparently significant structural changes in a ligand are causing 'unexpected' large changes in stability constants of metal complexes; it is of great importance and significance to find out about fundamental rules governing the strength of metal-ligand interactions on atomic and molecular level. One such, among many known examples, set of ligands is nitrilotriacetic acid (NTA) and nitrilotripropionic acid (NTPA) whose structural difference is an additional -CH_2- group in each acid-containing arm of the ligand NTPA. Even though the kind and the number of donor atoms, which form bonds with a metal ion, is the same for the both ligands the formation constants of metal complexes with NTA (strong complexing agent) are several log units larger when compared with equivalent complexes involving NTPA[4] (weak complexing agent). To illustrate this point, the complex ML of Cd(II) with NTA is over six orders of magnitude more stable when compared with NTPA (log K_1 = 9.76 and 3.4, respectively). The difference must be even larger for Pb(II) (log K_1 = 11.48 for NTA) as there is no value reported for NTPA,[4] it appears that Pb(NTPA) complex must be very weak and hence difficult to study experimentally.

The first obvious step necessary to achieve this ultimate goal is to determine theoretically the stepwise protonation constants of the ligands of interest. This must be seen as a challenge on its own as (i) there were no successful reports yet where ligands with three negative charges were investigated, and (ii) results reported for less negatively charged ligands were significantly different when compared with experimental data. In this chapter the applicability of thermodynamic cycles and carefully designed isodesmic reactions for theoretical prediction of four stepwise protonation constants of the ligand NTPA are examined. This chapter demonstrates that, at least in case of negatively and multiply-charged ligands, the best option in theoretical prediction of protonation (dissociation) constants is the isodesmic reaction; theoretically predicted protonation constants for NTPA ligand reported in this work compare well with experimental values.

2.2. Computational details

All computational calculations were performed with the aid of Gaussian 03 software package.[58] Gas-phase and solvent (water, ε = 78.39) geometry optimization of protonated NTPA forms was performed at the RB3LYP/6-311+G(d,p) level of theory.[59] As it was pointed out previously,[24] it is essential and of paramount importance to include diffuse functions for anions. The full optimization in solvent involved the default solvation

model, i.e. Tomasi's Polarized Continuum Model (PCM),[60-62] and UA0 radii (United Atom Topological Model). With this model the cavity is defined as the union of a series of interlocking atomic spheres. The effect of polarization of the solvent continuum is represented numerically and it is computed by numerical integration.[63] Single point calculations (SPCs) in solvent were carried out at the same level of theory (i) on the gas-optimized structures using PCM/UA0 model, and (ii) on solvent optimized structures using the polarisable conductor model (CPCM)[52, 64] and UAHF radii (United Atom for Hartree-Fock). With this model the solute cavities are modelled on the optimized molecular shape and include both electrostatic and non-electrostatic contributions to energies.[12]

Geometry optimization of all NTA and iminodiacetic acid (IDA) protonated forms (used as reference molecules in isodesmic reactions) was carried out in solvent using the same procedure as for NTPA; there was no need to perform a single point calculation. Frequency calculations were also performed, along with the geometry optimization, to ensure that each of the optimized molecules were in fact at minimum energies (for all structures considered in this chapter the imaginary frequencies were not present).

2.3. Results and Discussion

2.3.1. *Level of Theory*

From literature reports it follows that there is no strong evidence in support of using high level theories instead of commonly applied cheaper (time and hardware) B3LYP, for instance. The approach implemented here, to validate computational methods, was to select a combination of level of theory and basis set in such a way that it would reproduce a crystallographic structure with commonly acceptable accuracy in the field. Fortunately, there are available two crystallographic H_3L structures of the ligand NTPA[65, 66] – fully labelled reported crystal structure is shown in Figure 2.1. It is important to note that the N-atom is protonated ($N–H^+$) in the solid state form of H_3L ligand (NTPA) leaving one carboxylic group being de-protonated ($–COO^–$), hence this molecule is overall neutral, but with two local positive and negative charges. The energy-optimized in solvent crystallographic structure is marked further as H_3L^*. H_3L^* was used to generate input structures of the remaining 4 possible protonated forms of NTPA, namely $H_2L^–$, $HL^{2–}$, $L^{3–}$ (fully deprotonated ligand), and H_4L^+ (fully protonated form of the ligand). Starting from H_3L^*, input structures for full energy optimization of $H_2L^–$ (by removing a dissociable proton from the –COOH group) and H_4L^+ (by simply adding a proton to O-atom) were

generated. Similar procedure was followed to computationally generate HL^{2-} and L^{3-}, where a proton was removed from energy optimized structure to generate the product of stepwise dissociation reaction.

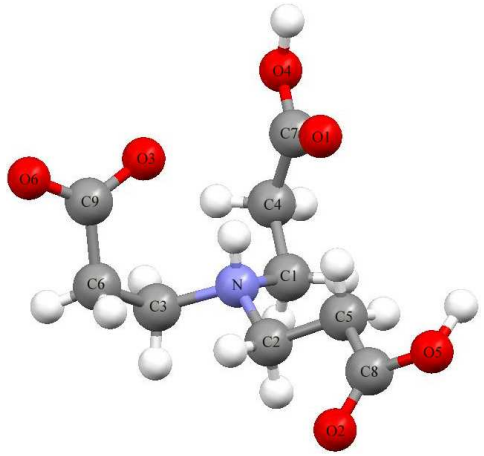
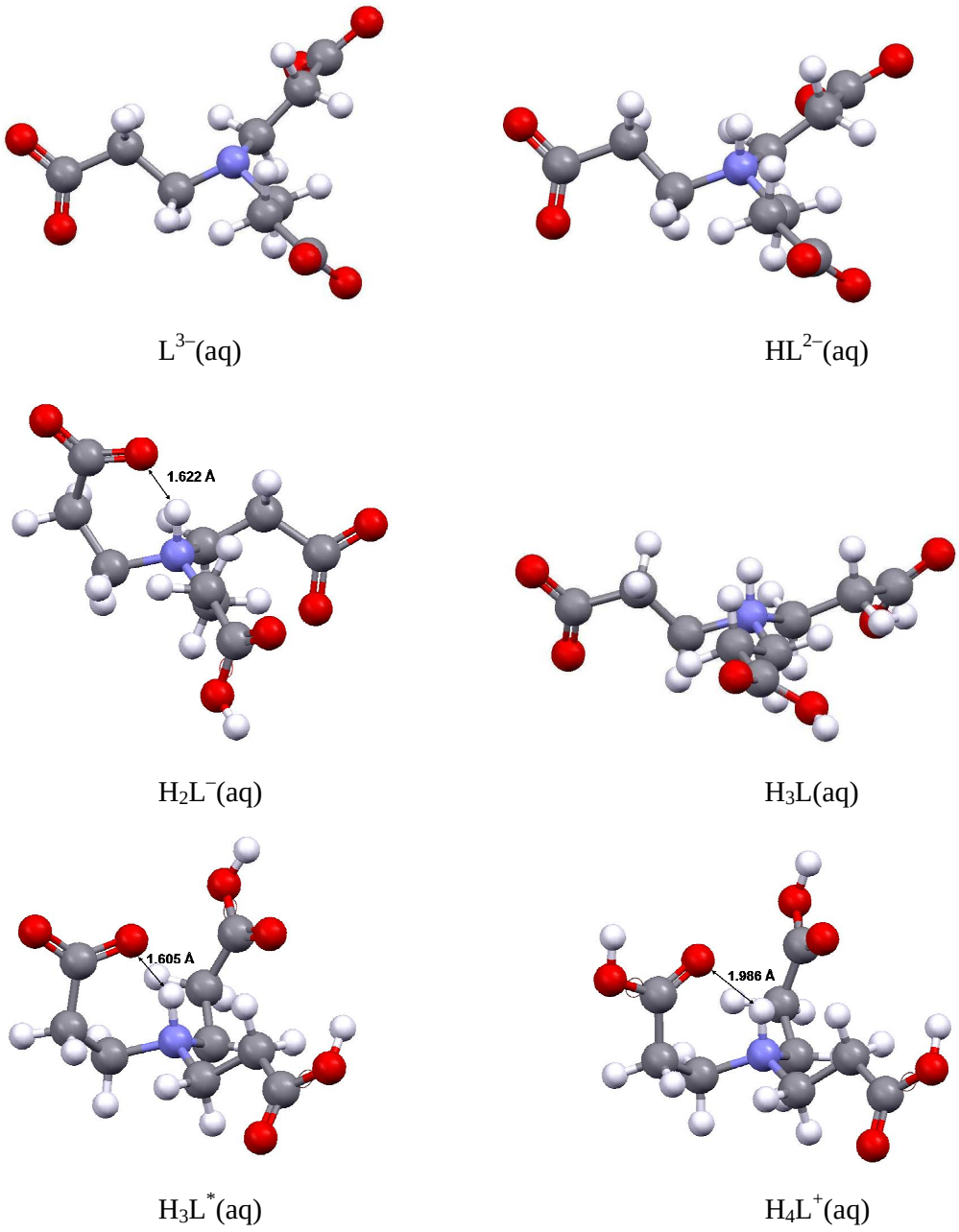


Figure 2.1: Fully labelled reported crystal structure of the H_3L form of NTPA.

Due to the reasons discussed in details further in the text, an H_3L was also constructed that was structurally different when compared with H_3L^*. The computed structural matrix of solvent-optimized H_3L^* and H_3L together with the data available from the Cambridge Structural Database (CSD)[66] is given in Table A1 (Appendix A). It is seen that the energy-optimized solvent-structure (H_3L^* in Figure 2.2), when the input was that of the reported H_3L crystallographic structure, can be regarded as fully satisfactory (Table A1; dihedral angles were not compared because they must vary in different protonated forms of NTPA). The bond lengths and angels were reproduced to within –0.010 ± 0.013 Å and –0.94 ± 2.25°, respectively; the difference (Δ) means experimental minus computed value. On average, the computed values are slightly overestimated and this is most likely due to energy optimization of a single molecule with ignoring lattice effects. More rewarding is the fact that the energy minimized H_3L molecule has marginally smaller differences when compared with crystallographic data; the bond lengths and angels were reproduced to within –0.012 ± 0.012 Å and –0.62 ± 1.92°, respectively, and the differences between the two optimized structures (δ = H_3L^* – H_3L) was considered as negligible, –0.002 ± 0.010 Å (bonds) and –0.32 ± 1.29° (angles). This provided confidence that all constructed structures should be seen as sufficiently reliable for further theoretical considerations. Also, data seen in Table A1 show that the RB3LYP/6-311+G(d,p) level of theory in conjunction with PCM/UA0 solvation model can be regarded as sufficient for the purpose of these studies.



$L^{3-}(aq)$ $HL^{2-}(aq)$

$H_2L^{-}(aq)$ $H_3L(aq)$

$H_3L^{*}(aq)$ $H_4L^{+}(aq)$

Figure 2.2: Structures of all protonated forms of the NTPA ligand fully optimized at the RB3LYP/6-311+G(d,p) level of theory in solvent (PCM/UA0).

All protonated forms of the NTPA ligand that were energy-optimized in solvent, including the crystal structure H_3L^{*}, are shown in Figure 2.2. It is seen that H_2L^{-}, H_3L^{*}, and H_4L^{+} have strong H-bonds, C–O^{-}···H–N, and they lengths are 1.622, 1.605, and 1.986 Å, respectively, whereas H_3L is a symmetrical molecule (without any evidence of intra-molecular bonding) that resembles to a large extent structures of L^{3-} and HL^{2-}.

2.3.2. *Thermodynamic cycle*

In the field of metal-ligand equilibria studies the complex formation constants and protonation (K_H) instead of dissociation (K_a) constants are used in solving mass-balance equation needed to develop most likely metal-ligand model (complexes formed) and refine stability constants of all metal-containing species. There are several important compilations of ligand protonation and complex formation constants, among them by Martell and Smith[4] and very recent one by IUPAC.[5] NTPA has four protonation constants

$$L^{3-} + H^{+} \leftrightarrow HL^{2-} \qquad K_H^{(1)} = \frac{[HL^{2-}]}{[L^{3-}][H^{+}]} \tag{1}$$

$$\vdots \qquad\qquad \vdots$$

$$H_3L + H^{+} \leftrightarrow H_4L^{+} \qquad K_H^{(4)} = \frac{[H_4L^{+}]}{[H_3L][H^{+}]} \tag{2}$$

and the values at several ionic strengths (μ) and temperatures are known from experiment.[4, 67] The thermodynamic cycle (TC) was utilized in the literature mainly to theoretically estimate dissociation constants that in the case of NTPA can be written as

$$H_4L^{+} \leftrightarrow H_3L + H^{+} \qquad K_a^{(1)} = \frac{[H_3L][H^{+}]}{[H_4L^{+}]} \tag{3}$$

$$\vdots \qquad\qquad \vdots$$

$$HL^{2-} \leftrightarrow L^{3-} + H^{+} \qquad K_a^{(4)} = \frac{[L^{3-}][H^{+}]}{[HL^{2-}]}. \tag{4}$$

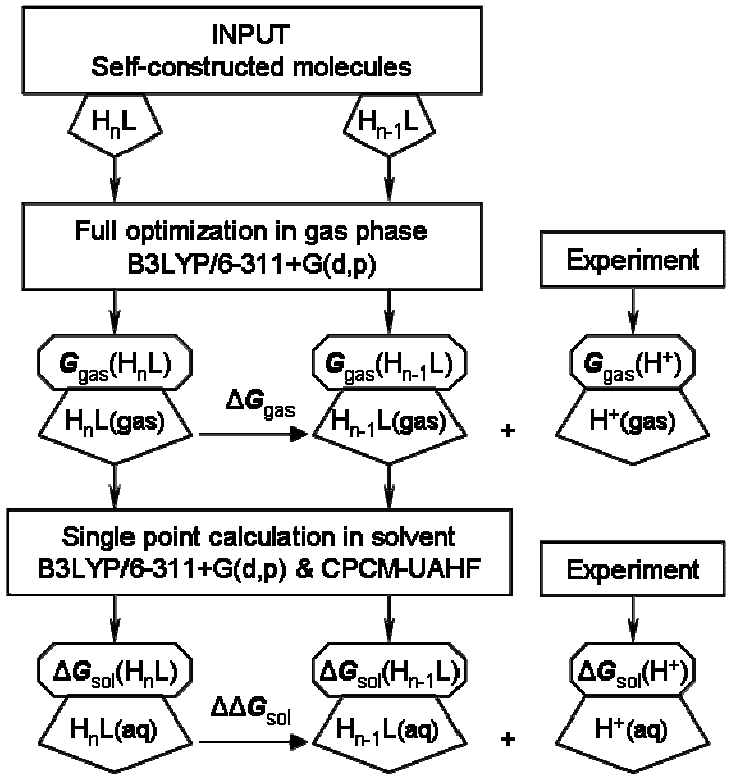
In eqs. 1-4 concentrations are used instead of activities because most of experiments are performed at selected ionic strength and temperature, according to envisaged practical application, such as, for instance blood plasma modelling that would require data at μ = 0.15 mol L^{-1} (Na^{+}(aq),Cl^{-}(aq)) and 37 °C. The computed value of ΔG(aq) can be used to calculate *n*th stepwise dissociation constant at 25 °C that, for convenience, is commonly reported as a pK_a value. The protonation reaction is reverse to weak acid dissociation reaction and in case of stepwise reactions the following equation holds where $k = 1 + m - n$, *m* stands for the highest dissociation constant (here $m = 4$). Note that the ligand NTPA

has three acidic groups and only three dissociation constants would be reported and hence dissociation reaction (eq 3) most likely would not be considered as the first dissociation reaction, log $K_a^{(1)}$, in theoretical prediction of pK_a values employing TC-based methodology. However, due to the protonation/de-protonation of N-atom in NTPA one must consider it as the fourth protonation constant, log $K_H^{(4)}$, as described by protonation reaction (eq 2).

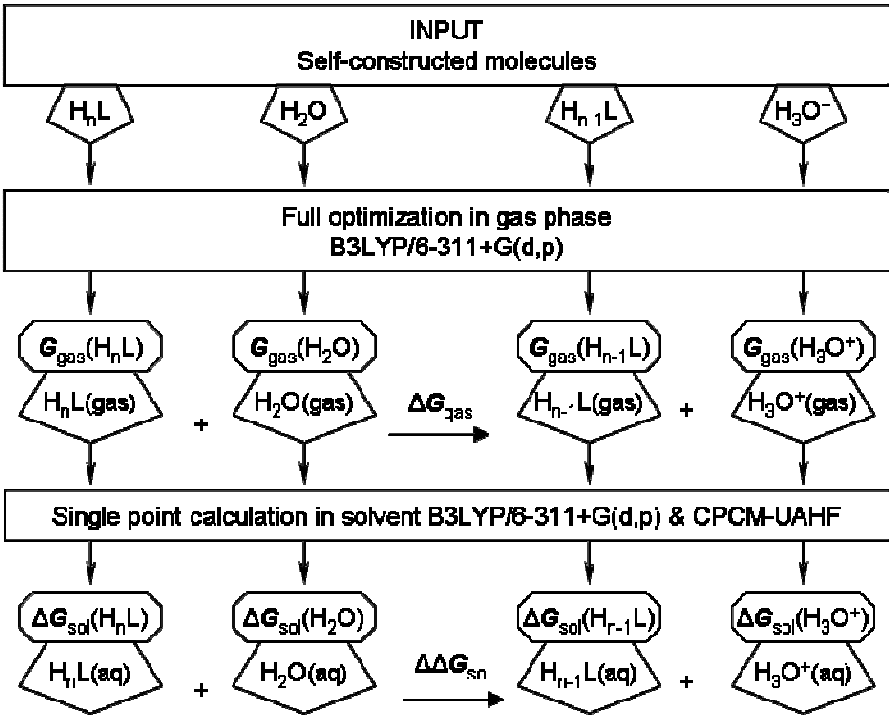
$$\log K_H^{(k)} = \log \frac{1}{K_a^{(n)}} = pK_a^{(n)} \qquad (5)$$

From this it follows that the 1st dissociation constant of NTPA is linked through eq 5 with the 4th protonation constant of this ligand. *n* indicates an *n*th consecutive dissociation constant, $1 \le n \le m$, and *k* applies to a *k*th consecutive protonation constant, $1 \le k \le m$. The above is well known in the field of metal-ligand equilibria studies but it is provided here for convenience and to assure clarity in nomenclature used.

Two TCs were considered in this work and they are shown in self-explanatory fashion in Figure 2.3 as Scheme 1 and 2 (Charges on the ligand species are omitted for simplicity throughout the text). In order to apply TC one needs to optimize each of the protonated form of the ligand NTPA in gas phase first. It is important to stress here that the moment solid NTPA is placed in water (the H_3L reagent has three protons present on carboxylic groups), the N-atom is protonated instantly and at least one proton on the propionic acid arm dissociates fully. The problem experienced here during energy optimization in gas phase of the crystallographic H_3L input structure that contains protonated nitrogen atom (as it is present in a solution) was that this proton shifted to the $–COO^-$ group to form -COOH. Numerous input structures were tested, but each time the presence of hydrogen on N-donor atom was not preserved. This phenomenon was also reported in the literature for the ligand aspartic acid.[21, 68] In attempt to preserve H-atom on nitrogen, another conformer of H_3L was built with all carboxylic groups placed as far as possible from the central N-atom. Unfortunately, even in this case energy optimization in gas phase has also resulted in de-protonation of the N-atom and formation of –COOH with H-atom involved in a hydrogen bond O–H···N. Clearly, in gas phase the molecule H_3L does not exist in zwitterionic form. Because of that TC could not be applied to the H_3L and H_4L forms of NTPA and theoretical studies had to be restricted only to the first two protonation reactions from which HL and H_2L are formed.



SCHEME 1



SCHEME 2

Figure 2.3: Schematic representation of TCs employed in this chapter.

The values of G_{gas} and ΔG_{sol} are shown in Table 2.1 together with minimum energies after zero-point vibrational energy (ZPVE) corrections, E_{min}, of optimized molecules. The values of ΔG_{gas}, $\Delta\Delta G_{sol}$ and ΔG_{aq} were calculated using well-known relationships (eqs 6-8)

$$\Delta G_{gas} = \sum G_{gas}(\text{products}) - \sum G_{gas}(\text{reactants}) \quad (6)$$

$$\Delta\Delta G_{sol} = \sum \Delta G_{sol}(\text{products}) - \sum \Delta G_{sol}(\text{reactants}) \quad (7)$$

$$\Delta G_{aq} = \Delta G_{gas} + \Delta\Delta G_{sol} \quad (8)$$

and dissociation constants $K_a^{(n)}$ were obtained from eq. 9 and eq. 10[11, 69] in case of TC-1 and TC-2, respectively.

$$\Delta G_{aq}^{(n)} = -RT \ln K_a^{(n)} \quad (9)$$

$$\Delta G_{aq}^{(n)} - RT \ln [H_2O] = -RT \ln K_a^{(n)} \quad (10)$$

As shown in Figure 2.3, Scheme 1, in case of proton ion, the experimental values of –6.28 and –264.61 kcal/mol for G_{gas} and ΔG_{sol} were used, respectively.[12, 32] Also, in final calculations of dissociation constants, appropriate correction of –1.89 kcal mol^{-1} (corresponding to the free energy change accompanied by the reversible state change of 1 mol of gas from 1 atm (24.47 L mol^{-1}) to 1 M (1 mol L^{-1})) was made to the calculated solvation free energy, as discussed thoroughly by Jang et al.[30]

Table 2.1: Selected thermochemical data (E_{min} stands for ZPVE-corrected energy) obtained for indicated NTPA species, H_2O, and H_3O^+.

	Gas-phase optimized structures		SPC in solvent	Solvent optimized structures	SPC in solvent
Species	E_{min} [a]	G_{gas} [a]	ΔG_{sol} [b]	E_{min} [a]	ΔG_{sol} [b]
L^{3-}	–856.307384	–856.358126	–335.26	–856.825492	–338.92
HL^{2-}	–857.007834	–857.058753	–184.84	–857.290671	–191.7
H_2L^-	–857.615989	–857.664227	–71.96	–857.742252	–87.05
H_2O	–76.437174	–76.454816	–6.72	–76.452207	–6.90
H_3O^+	–76.696787	–76.714893	–107.35	–76.840274	–109.78

[a]) In atomic unit, Hartree (1 Hartree = 627.5095 kcal/mol)
[b]) In kcal/mol

The calculated first two protonation constants for NTPA ligand using TC 1 and 2 are summarized in Table 2.1 – results obtained are far from satisfactory (see δ, the difference

between the computed and experimental values). In attempt to improve the prediction of computed protonation constants, slightly modified procedure was employed. It involved full energy optimization of the ligand species in solvent (PCM/UA0) followed by single point calculations in the solvent (CPCM/UAHF) in order to generate ΔG_{sol}; similar approach[2] resulted in some improvement in computed dissociation constants. Data obtained here is included in Tables 2.1-2.2.

Table 2.2: Comparison of experimental and calculated protonation constants, as log K_H, using gas-phase and solvent optimized structures, seen in Figure 2.2.

		Gas-phase structure				Solvent structure			
Reaction	**Exp[a]**	**TC 1**	***δ***	**TC 2**	***δ***	**TC 1**	***δ***	**TC 2**	***δ***
$L^{3-} + H^+ =$	9.49	12.06[b]	2.57	15.49[b]	6.00	14.40[b]	4.91	16.18[b]	6.69
HL^{2-}		11.20[c]	1.71	30.13[c]	20.64	13.58[c]	4.09	31.67[c]	22.18
$HL^{2-} + H^+ =$ H_2L^-	4.22	–4.19[b]	–8.41	–0.76[b]	–4.98	1.84[b]	–2.38	3.62[b]	–0.60

[a]) Experimental protonation constants[4] at 20 °C and ionic strength = 0.1 M.
[b]) The CPCM-UAHF model was used for SPC.
[c]) The PCM-UA0 model was used for SPC.

From Table 2.2 it is seen that (i) TC 1 worked better for first protonation constant for both, gas-phase and solvent, structures, (ii) gas-optimized structures generated smaller errors in computed first protonation constant when compared with equivalent solvent structures values, (iii) smaller errors in the second protonation constant were obtained from solvent structures which is opposite to what is observed for the first protonation constant. Also the influence of the solvation model used at the SPC was tested; the use of PCM/UA0 somewhat improved prediction generated from TC1 and made them erroneous when TC2 was employed. For some reasons, the simplified solvation model did not work at all for the second protonation reaction. There seems to be no obvious pattern in the data seen in Table 2.2 and, on average, results obtained are totally unacceptable due to large differences (δ) between the computed and experimental values. There are several possible sources for such large errors in computed values and some of them were discussed extensively elsewhere.[11, 46] Regardless of the reasons applicable to this particular case, it was concluded that since it was impossible to optimize all necessary protonated forms of the ligand NTPA in gas phase, further investigations involving different levels of theory, larger basis sets, other solvation models, or different thermodynamic cycles was not a justifiable option.

2.3.3. *Isodesmic Reactions*

Since protonation constants could not be computed with acceptable accuracy by employing TCs, attention was turned to applicability of methodology based on isodesmic reaction (IRn). An IRn can be applied when the total number of each type of a bond is identical in the reactants and products.[63] To date, IRns have been used to predict enthalpies of formation[70-78] and in some cases they have been incorporated within a thermodynamic cycles to predict dissociation constants.[1, 6-9] IRn is commonly used to investigate processes in solvent (such as water) as it should minimize (or systematically cancel of) errors related to the solvation model used[51] (similar error should be introduced into each component within the reaction) provided that the same level of theory, basis set and solvation model is used for each component involved in the reaction. It can only be employed if accurate experimental data, such as protonation constants or enthalpy, is available for the reference species used in the IRn.[51, 70] Interestingly, an explicit application of IRns in theoretical study of protonation/dissociation constants has not been found; for whatever reasons, that is not understood fully, only different kinds of TC were employed till now and almost exclusively in case of mono-dissociable organic acids. It was assumed that (i) weakness of presently available solvation models, particularly when poly-negatively charged ions are investigated, is mainly responsible for large errors in protonation constants generated from TC-based computations and (ii) the use of IRn methodology might eliminate to a significant degree errors typically associated with the use of TC.

Main challenge associated with the use of IRn is the selection of appropriate reference molecule. Taking into account structural properties of NTPA (called further $L_{(1)}$) NTA and IDA were chosen as reference compounds (called further $L_{(2)}$) because they have the same number (in case of NTA) and kind of electron donor atoms that can be protonated in a solution. Also, protonation constants of NTA and IDA are known as they are widely-studied ligands.[4, 5]

IRn can be seen here as simply a competition reaction between two ligands for a proton and for the first protonation constant of NTPA it can be written as

$$L_{(1)}(aq) + HL_{(2)}(aq) = HL_{(1)}(aq) + L_{(2)}(aq) \qquad \Delta G_{aq} \qquad (11)$$

Each of the two ligands (NTPA and reference molecule) is involved in several stepwise protonation reaction; for simplicity, only first is shown

$$H^+ + L_{(1)} \leftrightarrow HL_{(1)} \qquad \Delta G_1(aq) = -\Delta G_{1d}(aq) \qquad (12)$$

$$H^+ + L_{(2)} \leftrightarrow HL_{(2)} \qquad \Delta G_2(aq) = -\Delta G_{2d}(aq) \qquad (13)$$

where $\Delta G_{nd}(aq)$ refers to reverse and relevant stepwise dissociation reaction. The change in Gibbs energies for each protonation reaction can be written as

$$\Delta G_1(aq) = G_{aq}(HL_{(1)}) - G_{aq}(H^+) - G_{aq}(L_{(1)}) \qquad (14)$$

$$\Delta G_2(aq) = G_{aq}(HL_{(2)}) - G_{aq}(H^+) - G_{aq}(L_{(2)}) \qquad (15)$$

The IRn of interest (eq 11) can be obtained from subtracting eq 13 from 12, and hence from subtracting eq 15 from 14 one obtains expression for the change in Gibbs energy, $\Delta G_{aq} = \Delta G_1(aq) - \Delta G_2(aq)$ applicable to this IRn, where the uncertainty related to $G_{aq}(H^+)$ is no more applicable as this term cancels of

$$\Delta G_{aq} = \Delta G_1(aq) - \Delta G_2(aq) = G_{aq}(HL_{(1)}) - G_{aq}(L_{(1)}) - G_{aq}(HL_{(2)}) + G_{aq}(L_{(2)}) \qquad (16)$$

Equation 16 was used to calculate ΔG_{aq} of IRn (eq 11) from appropriate Gibbs energies obtained for relevant and fully solvent-optimized structures of the ligand NTPA and reference ligand $L_{(2)}$. Table A2 in Appendix A provides the ZPVE-corrected minimum energies E_{min} as well as Gibbs free energies for NTPA and IDA (energy solvent-minimized structures of IDA and relevant data for the reference ligand NTA are shown in Figure A1 and Table A3, respectively, in Appendix A). There are two energies listed for H_3L form of the ligand NTPA, one of which refers to the optimized crystal structure H_3L^*; they were both utilized in the calculation of stepwise protonation constants of NTPA in order to determine which yielded better results as these two structures have some distinguishable features.

The value of $\Delta G_2(aq)$ was obtained from eq 9 using reported protonation constants (at 25 °C and μ = 0.0 and 0.1 M) of the $L_{(2)}$ ligands, NTA and IDA. Having ΔG_{aq} and $\Delta G_2(aq)$ one can obtain $\Delta G_1(aq)$ (needed to calculate protonation constants of NTPA) from eq 16. Table 2.3 provides the values of functions required to calculate protonation constants, calculated and experimental protonation constants of the ligand NTPA, along with the differences between calculated and experimental protonation constants (δ). For all IRns seen in Table 2.3 the reference ligand is singly protonated, hence only first

protonation constant of IDA[4] was used to calculate ΔG_2(aq); values obtained were –13.358 and –12.744 kcal mol^{-1} from experimental protonation constants (log K_H = 9.79 and 9.34) at ionic strength 0.0 and 0.1 M, respectively, both at 25 °C.

Table 2.3: Comparison of experimental[4] (Exp) at μ = 0.1 M and 20 °C and calculated stepwise protonation constants of NTPA, as log K_H, using first protonation constant of the reference molecule IDA at ionic strength μ = 0.0 and 0.1 M and 25 °C.

		μ = 0.0 M, 25 °C (IDA)				**μ = 0.1 M, 25 °C (IDA)**			
Reaction	**ΔG_{aq}**	**ΔG_1(aq)**	**log K_H**	**Exp**	**δ**	**ΔG_1(aq)**	**log K_H**	**Exp**	**δ**
$L_{(1)}^{3-} + HL_{(2)}^{-} = HL_{(1)}^{2-} + L_{(2)}^{2-}$	–1.223	–14.581	10.69	9.49	1.20	–13.967	10.24	9.49	0.75
$HL_{(1)}^{2-} + HL_{(2)}^{-} = H_2L_{(1)}^{-} + L_{(2)}^{2-}$	6.719	–6.639	4.87	4.22	0.65	–6.025	4.42	4.22	0.20
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)} + L_{(2)}^{2-}$	11.350	–2.008	1.47	3.68	–2.21	–1.394	1.02	3.68	–2.66
$H_3L_{(1)} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L_{(2)}^{2-}$	9.060	–4.298	3.15	2.71[a]	0.44	–3.684	2.70	2.71[a]	–0.01
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)}^{*} + L_{(2)}^{2-}$	7.602	–5.755	4.22	3.68	0.54	–5.141	3.77	3.68	0.09
$H_3L_{(1)}^{*} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L_{(2)}^{2-}$	12.807	–0.550	0.40	2.71[a]	–2.31	0.064	–0.05	2.71[a]	–2.76

[a]) Experimental NTPA protonation constant[4] at μ = 0.5 M and 25 °C.
All energies are reported in kcal/mol.

A number of different isodesmic reactions have been tested, but only those that produced best results have been reproduced in Table 2.3 (The remaining results, also involving NTA ligand, are provided in Table A4 of Appendix A). It is seen from Table 2.3 that application of IRns resulted in much better overall prediction for the protonation constants when compared with results generated from TCs (Table 2.2). It is important to stress here that the available experimental protonation constants[4] of NTPA at μ = 0.1 M and 20 °C (except for the forth one, μ = 0.5 M and 25 °C) were compared against computed values generated with inclusion of protonation constants of the reference molecule IDA at μ = 0.0 and 0.1 M, both at 25 °C. Paying attention to ionic strength at which experimental values were obtained is not common practice in the literature when TCs were employed but based on the assumption that due to the inherent property of the IRn the prediction of computationally generated protonation constant, if possible, must be at the same ionic strength as the experimental values used for the reference ligand. In other words, the values of ΔG_1(aq) (computed) and ΔG_2(aq) (experimental value of the reference molecule) should be at the same ionic strength because by subtracting them the influence of ionic strength should cancel of and standard state function ΔG_{aq}, as computed

for IRn 11, is obtained.

From Table 2.3 it can also be seen that $K_H^{(1)}$, $K_H^{(2)}$, and $K_H^{(4)}$ values, as log K_H, of NTPA (when protonation constants of IDA at ionic strength of 0.1 M were utilized) are predicted with excellent accuracy; the H_3L structure was employed in case of the 3rd and 4th protonation constants. Incorporating the H_3L^* structure resulted in excellent prediction of $K_H^{(3)}$, the third protonation constant, but rather poor result was obtained for $K_H^{(4)}$. Interestingly, the δ values of poor predictions are almost identical, 2.66 and 2.76 log unit in case of self-constructed and crystallographic structure of H_3L used to calculate $K_H^{(3)}$, and $K_H^{(4)}$, respectively. This observation and analysis of results seen in Table 2.3 lead us to the conclusion that the structural differences between self-constructed H_3L and crystallographic H_3L^* are responsible for the observed significant differences between experimental and computed protonation constants. If this is indeed the case then one should perform full conformational analysis in solvent of all possible protonated forms of molecules involved prior to application of IRn. Interestingly full conformational analysis was not utilized when dissociation constants were computed from TCs; this is a time consuming exercise particularly when performed in Gaussian and this was embarked on as a separate project in order to investigate the influence of small structural changes on accuracy of computed protonation constants, but at the same time it was decided that a simplified test be conducted here. All forms of NTPA and IDA were subjected to the Schrödinger's Maestro[79] conformational analysis. This software automatically generates hundreds of possible conformers and estimates their energies in a short time based on MM/MD principles. It was of great interest to us to find out about predictions made by MM/MD analysis and whether this kind of conformational analysis would be of any use and help in this study. The structures of NTPA seen in Figure 2.2 were used as inputs for MM/MD conformational analysis in solvent (generated structures of lowest energy are shown in Figure A2 in Appendix A). Table 2.4(a) provides energies (in kJ mol^{-1}) of five lowest in energy MM/MD-conformers (C-1 to C-5) of all protonated forms of NTPA seen in Figure 2.2. Also SPC on the structures from Figure 2.2 were performed in solvent using MM in order to compare these energies with lowest energy MM/MD relevant conformer – obtained data is also included in Table 2.4(a). It was of some concern to see that all the SPC MM-structures were of considerably lower energy with the difference δ reaching over 20 kcal mol^{-1} (equivalent of about 14 log units in protonation constants) in case of H_3L (much lower difference was obtained for H_3L^*). Because of that the MM/MD-generated C-1 conformers were fully solvent-optimized in Gaussian using the same

procedure as described above for protonated NTPA species – results obtained are shown in Table 2.4(b).

Table 2.4: Part (a). Minimum energies of MM/MD-generated conformers in solvent (C1-C5) and energies obtained from MM-based SPC performed on the NTPA structures seen in Figure 2.2. Part (b). Solvent-optimized energies of all protonated forms of the ligand NTPA obtained from DFT calculations (E_{min} = ZPVE-corrected energy) of structures seen in Figure 2.2 and lowest energy MM/MD-generated C-1 conformers.

(a)

L = NTPA	E_{SPC}	E_{C-1}	δE kJ/mol	δE kcal/mol	C-2	C-3	C-4	C-5
L^{3-}	-947.64	-962.90	15.26	3.65	-962.89	-960.04	-960.04	-959.97
HL^{2-}	-1156.19	-1208.80	52.61	12.57	-1208.80	-1207.27	-1207.27	-1207.24
H_2L^{-}	-910.03	-957.14	47.11	11.26	-957.13	-956.01	-956.00	-955.67
H_3L	-615.06	-701.41	86.35	20.64	-701.19	-701.19	-700.21	-698.73
H_3L^{*}	-658.92	-716.44	57.52	13.75	-715.10	-713.39	-713.26	-712.96
H_4L^{+}	-379.64	-435.17	55.53	13.27	-435.16	-433.09	-433.09	-430.78

$\delta E = E_{SPC} - E_{C-1}$

(b)

	Structures seen in Fig. 2		C-1			
L = NTPA	E_{min} (Hartree)	G_{aq} (Hartree)	E_{min} (Hartree)	G_{aq} (Hartree)	δG_{aq} Hartree	δG_{aq} kcal/mol
L^{3-}	-856.825492	-856.875606	-856.828229	*-856.877193*	0.001587	1.00
HL^{2-}	-857.290671	-857.339621	-857.294193	*-857.342045*	0.002424	1.52
H_2L^{-}	-857.742252	-857.790980	-857.745414	*-857.792633*	0.001653	1.04
H_3L	-858.185011	-858.234959	-858.195297	*-858.241670*	0.006711	4.21
H_3L^{*}	-858.192859	-858.240931	-858.184953	-858.229956	0.010975	6.89
H_4L^{+}	-858.633248	*-858.682587*	-858.635423	-858.681482	0.001105	0.69

$\delta G_{aq} = G_{aq}$(structure in Fig. 2) – G_{aq}(C-1)

It was gratifying to see that, even though all energies obtained from SPC (involving MM) were lower in value, the differences came down to a single kcal-range except H_3L and H_3L^{*} for which δG = 4.21 and 6.89 kcal mol^{-1}, respectively. Similar procedure was applied to all the protonated forms of IDA and results are shown in Table 2.5. All SPC-generated energies of IDA were again lower in value when compared with energies of fully optimized structures seen in Figure 2.2 (Table 2.5(a) where $\delta E = E - E_{C-1} > 0$). However, when C-1 structures of IDA were fully DFT-optimized the resultant energies were lower in value when compared with the lowest energies of C-1 conformers (see Table 2.5(b) where $\delta G < 0$ for all except HL) and it was gratifying to see that the δG values were rather small. From that it might follow that to perform proper structural analysis in search of the lowest energy conformer, one would have to analyze a number of

structures obtained from the MM/MD optimisation. The aim here was to prove the point that indeed it is possible to theoretically predict four consecutive protonation constants with acceptable accuracy (the interest in this work was not to find out how accurate that prediction might be) hence full analysis of all the MM/MD-generated conformers was not performed – it appears, however, that MM/MD analysis might be a useful tool in search of conformers for the purpose of this kind of studies.

Table 2.5: Part (a). Minimum energies of MM/MD-generated conformers in solvent (C1-C5) and energies obtained from MM-based SPC performed on the IDA structures seen in Figure A1. Part (b). Solvent-optimized energies of all protonated forms of the ligand IDA obtained from DFT calculations (E_{min} = ZPVE-corrected energy) of structures seen in Figure A1 and lowest energy MM/MD-generated C-1 conformers.

(a)

L = IDA	E_{SPC}	E_{C-1}	δE kJ/mol	δE kcal/mol	C-2	C-3	C-4	C-5
L^{2-}	-663.08	-676.62	13.53	3.23	-676.52	-676.52	-657.54	-657.53
HL^{-}	-1002.94	-1034.18	31.24	7.47	-1034.17	-1031.99	-	-
H_2L	-759.59	-794.46	34.88	8.34	-794.46	-793.52	-793.51	-792.31
H_3L^{+}	-483.78	-517.66	33.88	8.10	-517.64	-517.38	-514.21	-514.20

$\delta E = E_{SPC} - E_{C-1}$

(b)

	Structures seen in Fig. A1		C-1			
L = IDA	E_{min} (Hartree)	G_{aq} (Hartree)	E_{min} (Hartree)	G_{aq} (Hartree)	δG_{aq} Hartree	δG_{aq} kcal/mol
L^{2-}	-511.483326	*-511.519032*	-511.483246	-511.518986	-0.000046	-0.03
HL^{-}	-511.945458	-511.981098	-511.945929	*-511.982547*	0.001449	0.91
H_2L	-512.387398	*-512.423365*	-512.385598	-512.422578	-0.000787	-0.49
H_3L^{+}	-512.825900	*-512.862287*	-512.821788	-512.858264	-0.004023	-2.52

$\delta G_{aq} = G_{aq}$(structure in Fig. S1) – G_{aq}(C-1)

Relevant NTPA and IDA structures were selected from Tables 2.4 and 2.5 that had lowest DFT-computed G_{aq} (printed in Italic in Tables 2.4-2.5) and were used for the protonation constants calculations based on the IRn approach discussed earlier – results obtained are shown in Table 2.6. The differences between experimental and computed values obtained at both ionic strengths are within ±1 log unit; this must be seen as an exceptional result since poly-negatively charged structures were investigated here. Similar, in magnitude, departures from experimental values were reported only for protonation constants of singly protonated common organic acids. An additional important fact is that the three experimental protonation constants (log $K_H^{(2)}$, log $K_H^{(3)}$, and log $K_H^{(4)}$) do not differ in

value by less than a log unit. Computed values, even though with errors of the same dimension, follow the experimental trend correctly, namely log $K_H^{(2)}$ > log $K_H^{(3)}$ > log $K_H^{(4)}$. Comparison of data seen in Tables 2.3 and 2.6 supports the above supposition that accuracy in predicted (computed) protonation constants depends on conformational structure used for both molecules (the studied and reference one); it was possible to improve (on average) computed protonation constants by use of simplified, MM/MD conformational analysis.

Table 2.6: Comparison of experimental[4] (Exp) at μ = 0.1 M and 20 °C and calculated stepwise protonation constants of NTPA, as log K_H, using the lowest energy structures from Tables 2.4-2.5 (seen in Italic) with protonation constants of the reference molecule IDA at ionic strength μ = 0.0 and 0.1 M.

	μ = 0.0 M, 25 °C (IDA)			**μ = 0.1 M, 25 °C (IDA)**		
Reaction	**log K_H**	**Exp**	**δ**	**log K_H**	**Exp**	**δ**
$L_{(1)}^{3-} + HL_{(2)}^{-} = HL_{(1)}^{2-} + L_{(2)}^{2-}$	10.40	9.49	0.91	9.95	9.49	0.46
$HL_{(1)}^{2-} + HL_{(2)}^{-} = H_2L_1^{-} + L_{(2)}^{2-}$	3.84	4.22	–0.38	3.39	4.22	–0.83
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)} + L_{(2)}^{2-}$	3.13	3.68	–0.55	2.68	3.68	–1.00
$H_3L_{(1)} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL_{(2)}^{-}$	2.89	2.71[a]	0.18	2.67	2.71[a]	–0.04

[a]) Experimental NTPA protonation constant[4] at μ = 0.5 M and 25 °C.

It is important to realize that energy optimization operation of all molecules, even though it was performed in solvent, most likely does not result in exact molecular structure (and hence the computed minimum energy) as it would exist in solution. This is most likely the reason why exact (to the second decimal place) prediction of protonation constants is not possible when the described protocol is implemented. Nevertheless, results obtained are very encouraging and suitable for many applications, where exact value might not be of absolute necessity. The conclusion reached was that implementation of IRn methodology provides, or might provide, computed protonation constants at acceptable accuracy. This broadens up a scope of studying protonation constants computationally and opens up a new field of applications for poly-charged ligands.

Incorporation of IDA rather than NTA in IRn has resulted in much better prediction of protonation constants. The reason for that is not clear at this stage as one would expect that NTA, being structurally much closer to NTPA, should generate better results – work is in progress to explain this phenomenon. It is evident; however, that the selection of the reference molecule plays crucial role and most likely it should be selected individually

according to structural properties of a molecule under investigation. It is not surprising, then, that the use of acetic acid as a single reference molecule in the study of a large number and structurally different mono-protonated organic acids (and also without conformational analysis) often resulted in very large errors in the predicted dissociation constants.[1]

2.4. Conclusion

This chapter has shown that prediction of several consecutive protonation constants for the highly and negatively charged molecules, such as NTPA, is possible with acceptable accuracy when isodesmic reaction methodology, instead of commonly employed thermodynamic cycle, is employed. Four stepwise protonation constants of NTPA were computed to within ±1 log unit of experimental data with an average error in the protonation constant of about 0.5 log unit. This good agreement was achieved for minimum energy structures of NTPA (studied ligand) and IDA (reference molecule) obtained from MM/MD conformational analysis followed by full energy optimization in solvent by Gaussian. It is reasonable to assume that even better estimates might be generated computationally when a number of lowest energy MM/MD-generated conformers were subjected to the DFT optimization in search for global minimum energy conformers. It appears that full conformational analysis should be seen as prerequisite for computing protonation/dissociation constants from IRn and possibly also from TC.

Results obtained in this study from TC agree with the literature reports in that TC methodology does not provide acceptably accurate results for negatively poly-charged molecules. Also, it was demonstrated that in the gas phase the proton prefers to be bound to $–COO^-$ group (to form –COOH) rather than to nitrogen in NTPA; this severely restricts the use of TC in the study of numerous ligands.

2.5. References

1. M. Namazian, S. Halvani, *J. Chem. Therm.* **2006**, *38*, 1495–1502.

2. G. A. A. Saracino, R. Improta, V. Barone, *Chem. Phys. Lett.* **2003**, *373*, 411.

3. G. Schuurmann, M. Cossi, V. Barone, J. Tomasi, *J. Phys. Chem. A* **1998**, *102*, 6706.

4. 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, US Department of Commerce, National Institute of Standards and Technology, 2004.

5. The IUPAC Stability Constants Database, http://www.iupac.org distributed and maintained by Academic Software, Sourby Old Farm, Timble, Otley, Yorks, LS21, 2PW, UK, (http://www.acadsoft.co.uk/scdbase/),

6. M. Namazian, H. Heidary, *J. Mol. Struct. (Theochem)* **2003**, *620*, 257.

7. M. Namazian, S. Halvani, M. R. Noorbala, *J. Mol. Struct. (Theochem)* **2004**, *711*, 13.

8. M. Namazian, F. Kalantary-Fotooh, M. R. Noorbala, D. J. Searles, M. C. Coote, *J. Mol. Struct. (Theochem)* **2006**, *758*, 275.

9. M. Namazian, M. Zakery, M. R. Noorbala, M. L. Coote, *Chem. Phy. Lett.* **2008**, *451*, 163.

10. I. E. Charif, S. M. Mekelleche, D. Villemin, N. Mora-Diez, *J Mol. Struct. (Theochem)* **2007**, *818*, 1.

11. M. D. Liptak, G. C. Shields, *Int. J. Quantum Chem.* **2001**, *85*, 727.

12. M. D. Liptak, G. C. Shields, *J. Am. Chem. Soc.* **2001**, *123*, 7314.

13. C. O. Silva, E. C. d. Silva, M. A. C. Nascimento, *J. Phys. Chem. A* **2000**, *104*, 2402.

14. D. M. Chipman, *J. Phys. Chem. A* **2002**, *106*, 7413.

15. A. Klamt, F. Eckert, M. Diedenhofen, M. E. Beck, *J. Phys. Chem. A* **2003**, *107*, 9380.

16. C. P. Kelly, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. A* **2006**, *110*, 2493.

17. J. R. Pliego Jr., J. M. Riveros, *J. Phys. Chem. A* **2002**, *106*, 7434.

18. J. J. Klicic, R. A. Friesner, S. Liu, W. C. Guida, *J. Phys. Chem. A* **2002**, *106*, 1327.

19. K. R. Adam, *J. Phys. Chem. A* **2002**, *106*, 11963.

20. K. B. Wiberg, S. Clifford, W. L. Jorgensen, M. J. Frisch, *J. Phys. Chem. A* **2000**, *104*, 7625.

21. W. Sang-Aroon, V. Ruangpornvisuti, *Int. J. Quantum Chem.* **2008**, *108*, 1181.

22. V. S. Bryantsev, M. S. Diallo, W. A. Goddard III, *J. Phys. Chem. A* **2007**, *111*, 4422.

23. J. I. Mujika, J. M. Mercero, X. Lopez, *J. Phys. Chem. A* **2003**, *107*, 6099.

24. D. Gao, P. Svoronos, P. K. Wong, D. Maddalena, J. Hwang, H. Walker, *J. Phys. Chem. A* **2005**, *109*, 10776.

25. H. Li, A. W. Hains, J. E. Everts, A. D. Robertson, J. H. Jensen, *J. Phys. Chem. B* **2002**, *106*, 3486.

26. E. E. Dahlke, C. J. Cramer, *J. Phys. Org. Chem.* **2003**, *16*, 336.

27. M. D. Liptak, K. C. Gross, P. G. Seybold, S. Feldgus, G. C. Shields, *J. Am. Chem. Soc.* **2002**, *124*, 6421.

28. G. da Silva, E. M. Kennedy, B. Z. Dlugogorski, *J. Phys. Chem. A* **2006**, *110*, 11371.

29. C. O. Silva, E. C. d. Silva, M. A. C. Nascimento, *Chem. Phys. Lett.* **2003**, *381*, 244.

30. J. Ulander, A. Broo, *Int. J. Quantum Chem.* **2005**, *105*, 866.

31. Y. H. Jang, L. C. Sowers, T. Cagin, W. A. Goddard III, *J. Phys. Chem. A* **2001**, *105*, 274.

32. Y. H. Jang, W. A. Goddard III, K. T. Noyes, L. C. Sowers, S. Hwang, D. S. Chung, *J. Phys. Chem. B* **2003**, *107*, 344.

33. L. C. Jitariu, A. J. Masters, I. H. Hiller, *J. Chem. Phys* **2004**, *121*, 7795.

34. K. C. Gross, P. G. Seybold, *Int. J. Quantum Chem.* **2000**, *80*, 1107.

35. K. C. Gross, P. G. Seybold, *Int. J. Quantum Chem.* **2001**, *85*, 569.

36. K. Byun, Y. Mo, J. Gao, *J. Am. Chem. Soc.* **2001**, *123*, 3974.

37. K. Murlowska, N. Sadlej-Sosnowska, *J. Phys. Chem. A* **2005**, *109*, 5590.

38. G. Li, Q. Cui, *J. Phys. Chem. B* **2003**, *107*, 14521.

39. F. Himo, L. Noodleman, M. R. A. Blomberg, P. E. M. Siegbahn, *J. Phys. Chem. A* **2002**, *106*, 8757.

40. A. M. Magill, K. J. Cavell, B. F. Yates, *J. Am. Chem. Soc.* **2003**, *126*, 8717.

41. I. A. Topol, S. K. Burt, A. A. Rashin, J. W. Erickson, *J. Phys. Chem. A* **2000**, *104*, 866.

42. C. Scharnagl, R. A. Raupp-Kossmann, *J. Phys. Chem. B* **2004**, *108*, 477.

43. I. Lee, C. K. Kim, I. Y. Lee, C. K. Kim, *J. Phys. Chem. A* **2000**, *104*, 6332.

44. J. D. Kubicki, *J. Phys. Chem. A* **2001**, *105*, 8756.

45. Z. B. Maksic, R. Vianello, *J. Phys. Chem. A* **2002**, *106*, 419.

46. J. E. Davies, N. L. Doltsinis, A. J. Kirby, C. D. Roussev, M. Sprik, *J. Am. Chem. Soc.* **2002**, *124*, 6594.

47. Y. J. Kim, A. Streitwieser, *J. Am. Chem. Soc.* **2002**, *124*, 5757.

48. N. Sadlej-Sosnowska, *Theor. Chem. Acc.* **2007**, *118*, 281.

49. G. N. Merrill, G. D. Fletcher, *J. Mol. Struct. (Theochem)* **2008**, *849*, 84.

50. E. Soriano, S. Cerdan, P. Ballesteros, *J. Mol. Struct. (Theochem)* **2004**, *684*, 121.

51. C. J. Cramer, 'Essentials of Computational Chemistry', John Wiley & Sons Ltd, 2002.

52. V. Barone, M. Cossi, *J. Phys. Chem. A* **1998**, *102*, 1995.

53. J. A. Montgomery, M. J. Frisch, J. W. Ochterski, G. A. Petersson, *J. Chem. Phys.* **1999**, *110*, 2822.

54. J. W. Ochterski, G. A. Petersson, J. A. Montgomery, Jr., *J. Chem. Phys.* **1996**, *104*, 2598.

55. G. A. Petersson, D. K. Malick, W. G. Wilson, J. W. Ochterski, J. A. Montgomery, M. J. Frisch, *J. Chem. Phys.* **1998**, *109*, 10570.

56. I. A. Topol, G. J. Tawa, S. K. Burt, A. A. Rashin, *J. Chem. Phys.* **1999**, *111*, 10998.

57. P. R. Varadwaj, I. Cukrowski, H. M. Marques, *J. Phys. Chem. A* **2008**, *112*, 10657.

58. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Pettersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian 03, Revision D.01, Gaussian, Inc., Wallingford CT, 2004.

59. P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623.

60. E. Cances, B. Mennucci, J. Tomasi, *J. Chem. Phys.* **1997**, *107*, 3032.

61. M. Cossi, V. Barone, R. Cammi, J. Tomasi, *Chem. Phys. Lett.* **1996**, *255*, 327.

62. S. Miertus, E. Scrocco, J. Tomasi, *Chem. Phys.* **1981**, *55*, 117.

63. J. B. Foresman, A. Frisch, 'Exploring Chemistry with Electronic Structure Methods', Gaussian, Inc., Pittsburgh, PA, 1996.

64. M. Cossi, N. Rega, G. Scalmani, V. Barone, *J. Comput. Chem.* **2003**, *24*, 669.

65. E. H. Walker, Jr., A. W. Apblett, R. Walker, A. Zachary, *Chem. Mater.* **2004**, *16*, 5336.

66. F. H. Allen, *Acta Cryst.* **2002**, *B58*, 380.

67. L. A. Curtiss, K. Raghavachari, *Theor. Chem. Acc.* **2002**, *108*, 61.

68. W. Sang-Aroon, V. Ruangpornvisuti, *J. Mol. Struct. (Theochem)* **2006**, *758*, 181.

69. J. R. Pliego Jr., *Chem. Phys. Lett.* **2003**, *367*, 145.

70. G. da Silva, J. W. Bozzelli, *J. Phys. Chem. C* **2007**, *111*, 5760.

71. G. da Silva, J. W. Bozzelli, *J. Phys. Chem. A* **2006**, *110*, 13058.

72. Y. Guo, H. Gao, B. Twamley, J. M. Shreeve, *Adv. Mater.* **2007**, *19*, 2884.

73. J. Espinosa-Garcia, J. C. Garcia-Bernaldez, *Phys. Chem. Chem. Phys.* **2002**, *4*, 4096.

74. I. Krossing, I. Raabe, *Chem-Eur. J.* **2004**, *10*, 5017.

75. A. Kutt, V. Movchun, T. Rodima, T. Dansauer, E. B. Rusanov, I. Leito, I. Kaljurand, J. Koppel, V. Pihl, I. Koppel, G. Ovsjannikov, L. Toom, M. Mishima, M. Medebielle, E. Lork, G.-V. Roschenthaler, I. A. Koppel, A. A. Kolomeitsev, *J. Org. Chem.* **2008**, *73*, 2607.

76. E. M. Nolan, R. G. Linck, *J. Am. Chem. Soc.* **2000**, *122*, 11497.

77. D. Peeters, G. Leroy, C. Wilante, *J. Mol. Struct.* **1997**, *416*, 21.

78. L. Wang, D. E. Heard, M. J. Pilling, P. Seakins, *J. Phys. Chem. A* **2008**, *112*, 1832.

79. L. L. C. Schrodinger, Schrodinger Maestro, 32nd Floor, Tower 45, 120 West Forty-Fifth Street, New York, 10036, 2003.

Chapter 3

This is a typed copy of a manuscript entitled “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”, *J. Phys. Chem. A.* (under review)

3. Protonation Constants of NTA

3.1. 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 a 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] Theoretical predictions performed thus far utilized thermodynamic cycles (Born-Haber cycles), which involve a two step operation, namely (i) a full gas-phase optimization of each molecule involved in a dissociation reaction, followed by (ii) a single point calculation in solvent on the gas-phase optimized structures using different solvation models and levels of theory.[1-52] Results reported to date predominantly describe the calculations of singly charged molecules, either anions[1-28] (a study of doubly charged anions is very rare), or cations[29-34]; this is due to inaccurate computational evaluations of hydration and/or solvation energies for highly charged ions. Accuracies achieved thus far for computed dissociation constants are within ±1.0 log unit, on average, when compared with experimentally available values, but differences of several log units are not uncommon.[3, 4]

Recently a paper has been published were the DFT-predicted four stepwise protonation constants, expressed as log $K_H^{(n)}$, for a highly charged molecule nitrilotripropionic acid (NTPA)[53] was determined. NTPA can exist in 5 forms in a solution, from a triply charged anion (when fully deprotonated at high pH), through a neutral molecule, to a singly charged cation (when fully protonated at very low pH). It was established that the application of thermodynamic cycles (TCs) yielded unacceptable results that varied from experimental log $K_H^{(n)}$ values by many log units, and discovered that application of a computational procedure based on a concept of an isodesmic reaction (IRn) resulted in far more accurate predicted values (with the average difference between predicted and experimental stepwise protonation constants being ± 0.5 log unit). This proved that, contrary to previous belief,[54] accurate determination of stepwise protonation constants for highly charged molecules is possible.

Since there are no other reports in which highly charged molecules, such as NTPA or NTA, have been investigated, it is important to investigate many more poly-charged ligands

in order to establish (i) whether indeed application of IRn might always result in much better values than those obtained from TCs, (ii) how significant the selection of a reference molecule is (its structural similarity to the studied compound) when accuracy in computed protonation constants is concerned, (iii) how significant the selection of a protonated form of a reference molecule (its charge in relation to the charge on a studied compound in a particular isodesmic reaction) is when accuracy in computed protonation constants is concerned, (iv) which one, structural similarity or charge on a protonated form of a molecule, play more important role in a isodesmic reaction, and (v) due to unavoidable errors in computed values, if it is possible to predict protonation constants in correct order as determined from an experiment.

In this chapter studies are conducted on the ligand nitrilotriacetic acid (NTA), an important derivative of glycine that is widely studied due to its excellent chelating abilities.[55] This is a ligand that has enjoyed numerous applications in medicine[56, 57], biochemistry[56, 57] and industry.[58-60] 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.[61] NTA has also been used as a transient phytoextraction agent that combines high biodegradability and low phytotoxicity with chelating strength.[62]

Since the protonation/dissociation of a ligand is what determines its biological significance, it was only fitting that focus be placed on the prediction of stepwise protonation constants for the ligand NTA with a number of reference molecules involved in IRns. The influence of not only a kind of a reference molecule (its structural similarity to the studied compound), but also its protonated form (positively or negatively charged molecule) on the theoretical prediction of stepwise protonation constants of NTA is investigated. The reference molecules shown in Figure 3.1, Iminodiacetic acid (IDA), N-Methyliminodiacetic acid (MIDA), N-Ethyliminodiacetic acid (EIDA), N-Propyliminodiacetic acid (PIDA) and N-(2-Hydroxyethyl)iminodiacetic acid (HIDA) have many structural similarities with the molecule of interest (NTA) and their experimental stepwise protonation constants are well known.[55] For comparison purposes, this study also evaluates applicability of TCs in order to determine which of the two methodologies is more applicable to the systems studied.

3.2. Computational Details

All calculations were performed using GAUSSIAN 03, revision D.01[63] on a 64-bit Linux

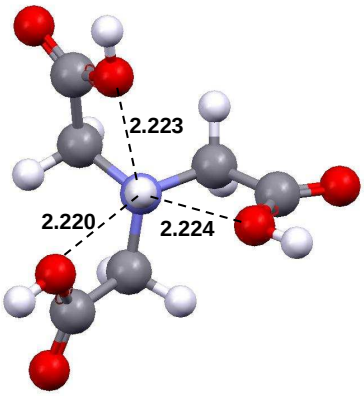
workstation in parallel environment (Opensuse 10.3). Molecular visualizations were accomplished with the aid of GaussView 4[64]. Since it is of paramount importance to include diffuse functions for anions,[16] both gas-phase and solvent (water, ε = 78.39) optimization of protonated NTA forms was performed at the RB3LYP/6-311+G(d,p) level of theory.[65] Full optimization in solvent was performed with the default solvation model provided by Gaussian, i.e. Tomasi's Polarized Continuum Model (PCM),[66-68] and UA0 radii (United Atom Topological Model). This model was chosen because it resulted in acceptable results in the prediction of log $K_H^{(n)}$ values for NTPA.[53] Single point calculations (SPCs) were carried out in solvent at the same (RB3LYP/6-311+G(d,p)) and HF/6-311+G(d,p) level of theory on the gas-phase structures and structures fully optimized in solvent, using the PCM-UAHF as well as CPCM-UAHF[69, 70] (polarisable conductor model in combination with the United Atom for Hartree-Fock radii) solvation models. With CPCM, the solute cavities are modelled on the optimized molecular shape, and include both electrostatic and non-electrostatic contributions to energies.[9] The HF level of theory also was used for single point calculations since the UAHF radii were optimized for HF.[18]

Full geometry optimization of all protonated forms of the reference molecules, seen in Figure 3.1, 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).

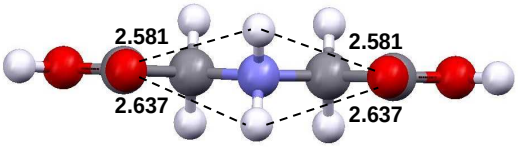
3.3. Results and Discussion

3.3.1: *Level of Theory*

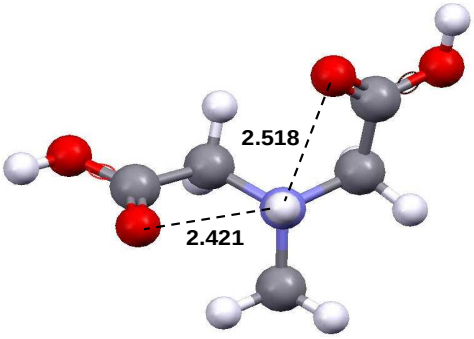
In the Cambridge Structural Database (CSD)[71] there are only two crystallographic structures of NTA[72, 73] and they both are of H_3L form of the ligand (referred to as H_3L^* from here onwards). It is important to stress here that H_3L^*, even though it has no overall charge, has two charged centres with opposite polarities, the positive one on the protonated N-atom and negative one on the de-protonated $–COO^-$ group – see Figure 3.2. It was decided that a level of theory and basis set should be chosen in such a way that it would reproduce the structure of H_3L^*.



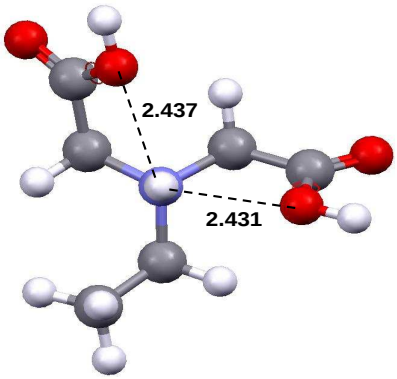
NTA (H_4L^+)



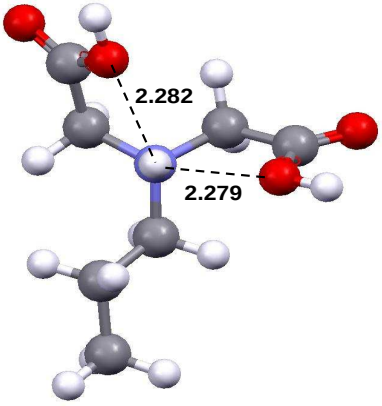
IDA (H_3L^+)



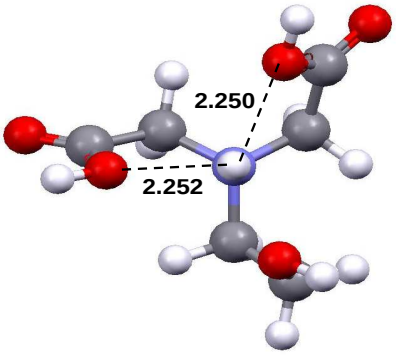
MIDA (H_3L^+)



EIDA (H_3L^+)



PIDA (H_3L^+)

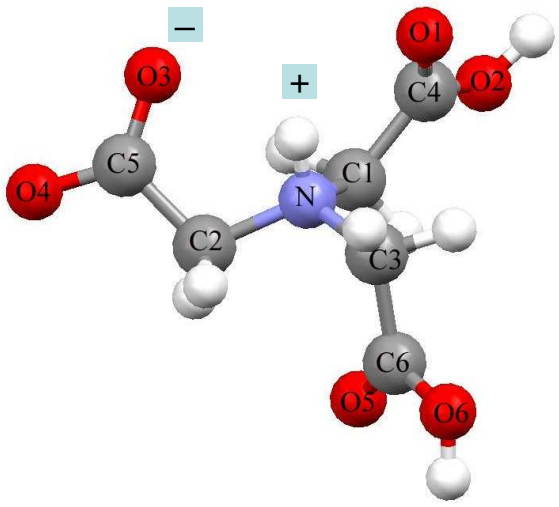


HIDA (H_3L^+)

Figure 3.1: Top view of the ligands (in fully protonated forms) discussed in this work.

Structures for the remaining 4 possible forms of NTA, namely H_2L^-, HL^{2-}, L^{3-} (fully deprotonated ligand), and H_4L^+ (fully protonated form of the ligand) had to be self-constructed as there is no crystallographic information available for them. To maintain

consistency and deal with a full set of self-constructed forms of the NTA ligand, the H_3L molecule had to also be constructed. From this it follows that it was imperative to perform protonation constant calculations not only on the crystallographic H_3L^* structure, but also on the self-constructed H_3L as well. If successful, this should provide us with some sort of assurance when optimization of the remaining self-constructed structures is concerned. The H_2L^- was generated from the energy-optimized H_3L structure (by removing a dissociable proton from the –COOH group) and H_4L^+ was generated by adding a proton to the remaining $–COO^-$ group in the energy-optimized H_3L structure. A similar procedure was followed to generate HL^{2-} and L^{3-}, were a proton was removed from the preceding protonated and energy-optimized structure to generate the product of stepwise dissociation reaction. Very much the same procedure was applied to generate all the protonated forms of the reference molecules seen in Figure 3.1.



NTA (H_3L)

Figure 3.2: Fully labelled reported crystal structures of the H_3L form of NTA[71].

The computed structural matrix of the solvent-optimized H_3L^* and H_3L of NTA together with the data available from the CSD[71] is given in Table 3.1 (The relevant data obtained for IDA, MIDA, and HIDA are provided in Appendix B, Tables B1-B3; the fully labelled crystal structures of H_2L^* forms of IDA, MIDA and HIDA, whose labelling was used in Tables B1-B3, is given in Figure B1).

Table 3.1: Comparison of experimental (CSD) and computed at the RB3LYP/6-311+G(d,p) level of theory in conjunction with PCM/UA0 solvation model selected bond distances and angles for the H_3L^* and H_3L forms of NTA. Bond lengths and angles are in Å and deg, respectively.

Atoms	CSD[a]	H_3L^*	Δ[b]	H_3L	Δ[b]	δ[c]
N-C1	1.501	1.500	0.001	1.500	0.001	0.000
N-C2	1.493	1.514	–0.021	1.513	–0.020	0.001
N-C3	1.497	1.497	0.000	1.502	–0.005	–0.005
C1-C4	1.513	1.518	–0.005	1.521	–0.008	–0.003
C2-C5	1.513	1.550	–0.037	1.552	–0.039	–0.002
C3-C6	1.519	1.520	–0.001	1.520	–0.001	0.000
C4-O1	1.204	1.209	–0.005	1.207	–0.003	0.002
C5-O3	1.252	1.245	0.007	1.246	0.006	–0.001
C6-O5	1.209	1.210	–0.001	1.206	0.003	0.004
C4-O2	1.310	1.327	–0.017	1.328	–0.018	–0.001
C5-O4	1.250	1.258	–0.008	1.257	–0.007	0.001
C6-O6	1.304	1.325	–0.021	1.330	–0.026	–0.005
C1-N-C2	112.4	111.5	0.9	110.8	1.6	0.7
C1-N-C3	113.7	114.5	–0.8	113.6	0.1	0.9
C2-N-C3	113.2	114.3	–1.1	111.8	1.4	2.5
N-C1-C4	110.2	108.6	1.6	108.8	1.4	–0.2
N-C2-C5	111.2	113.7	–2.5	114.1	–2.9	–0.4
N-C3-C6	110.7	111.1	–0.4	114.3	–3.6	–3.2
C2-C5-O3	122.1	123.7	–1.6	121.6	0.5	2.1
C1-C4-O2	117.1	115.6	1.5	115.5	1.6	0.1
C3-C6-O5	122.1	125.1	–3.0	121.6	0.5	3.5
C1-C4-O2	117.5	115.3	2.2	115.2	2.3	0.1
C2-C5-O4	111.7	109.5	2.2	112.6	–0.9	–3.1
C3-C6-O6	111.5	110.5	1.0	112.4	–0.9	–1.9

[a]) Average bond lengths and angles of two crystal structures obtained from the CSD [71]
[b]) Δ = (CSD – computed) value
[c]) δ = (H_3L^* – H_3L) value

From Table 3.1 it is seen that the H_3L^* structure of NTA, energy-optimized in solvent when the input was that of the reported crystallographic structure H_3L (shown in Figure 3.2), can be regarded as satisfactory when the bond lengths are considered – they were predicted to within –0.009 ± 0.013 Å. On average, the computed values are slightly overestimated, as is expected

for B3LYP optimization.[53] However, some of the computed bond angles (that on average were reproduced to within 0.00 ± 1.81°) differ significantly from that found in the two reported crystal structures (similar observation applies to the dihedral angles – they are not provided in the table). The simplest and reasonable explanation of observed differences might be the fact that modelling performed on a single molecule did not account for crystal packing forces as well as intermolecular non-bonding interactions in the lattice.

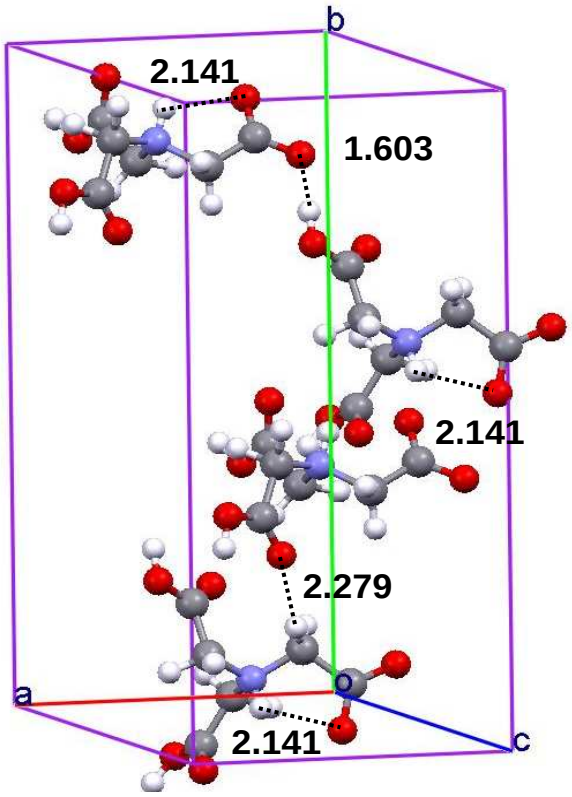
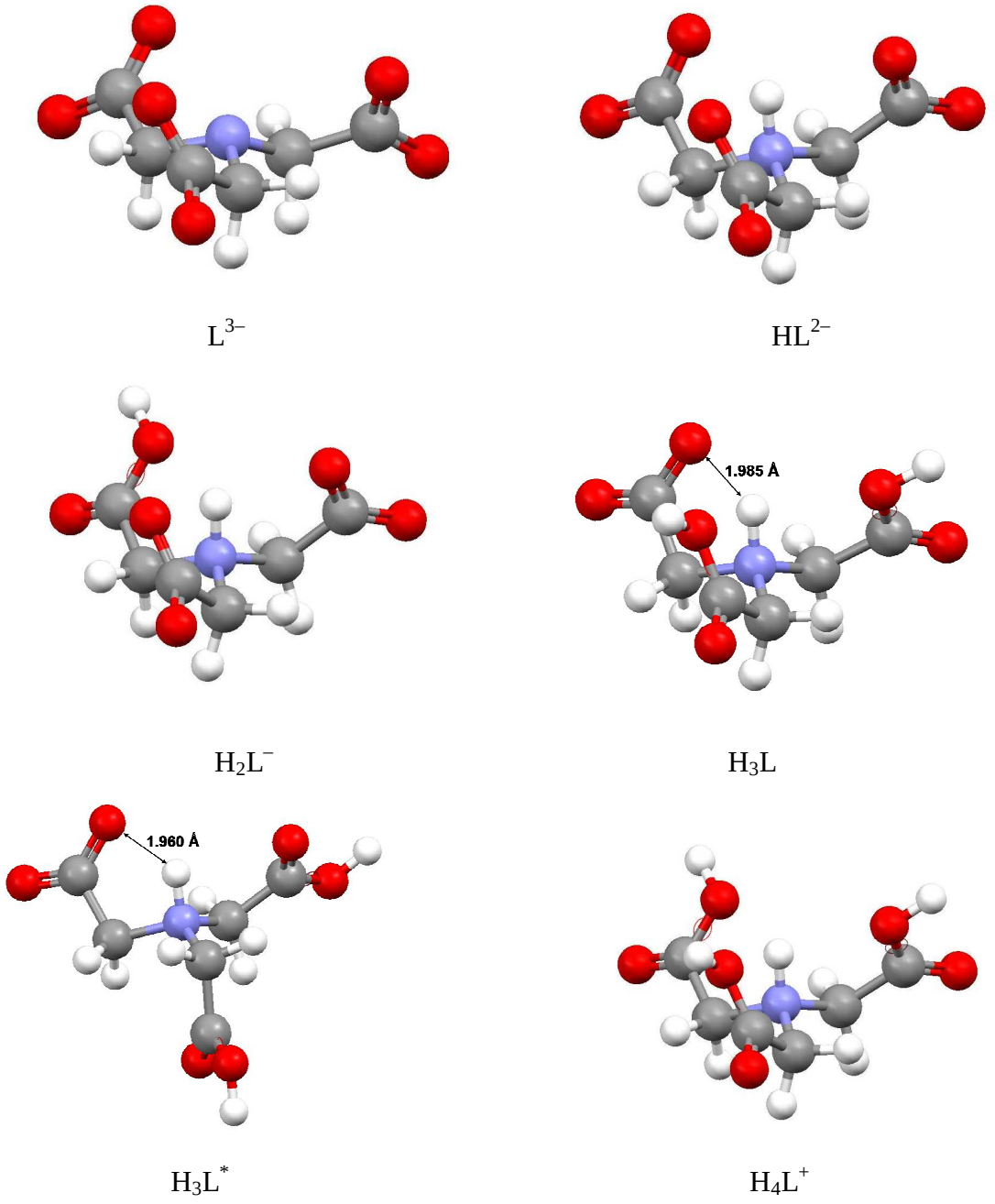


Figure 3.3: Crystallographic structure[71] of NTA (molecules within a unit cell) with selected intra- and intermolecular non-bonding interactions marked by dashed lines and distances in Å.

Figure 3.3 can be used in support of the supposition where one can clearly see strong hydrogen bond –OH···O (1.603 Å) between N-C3-C6-O5-H arm of one NTA molecule and N-C1-C4-O1 arm of another NTA molecule. There is also a close contact –CO···HC– (2.279 Å) and, for clarity, only these two intermolecular interactions are shown in Figure 3.3; all carboxylic groups are involved in numerous intermolecular interactions. Intermolecular interactions are also responsible for the intramolecular H-bond of 1.960 Å, found in the

energy-optimized H_3L^* form of NTA (shown in Figure 3.4), being shorter when compared with that in crystal structure (2.141 Å – see Figure 3.3).



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

H_2L^- H_3L

H_3L^* H_4L^+

Figure 3.4: Structures of all protonated forms of NTA fully optimized at the RB3LYP/6-311+G(d,p) level of theory in solvent (PCM/UA0).

The energy minimized H_3L molecule also has small (and of the same order of magnitude as reported above for optimized crystallographic structure H_3L^*) differences when compared

with crystallographic data; bond lengths and angles were reproduced, on average, to within –0.010 ± 0.013 Å and 0.09 ± 1.85°, respectively. The differences between the two H_3L optimized structures ($\delta = H_3L^* - H_3L$) can be considered as negligible in case of bond lengths (δ = –0.001 ± 0.003) and acceptable in case of bond angles (δ = 0.09 ± 2.08°) when one accounts for the intermolecular interactions discussed above; as a matter of fact, due to strong intermolecular interactions present in the lattice it would be surprising if better agreement was observed. Similar observations apply to the data generated for optimized H_3L structure of HIDA (Table B3 in Appendix B). Differences in computed and experimental angles of HIDA can be easily rationalised when close contacts seen in Figure B2 (Appendix B) are considered; as an example, exceptionally strong intermolecular interactions of 1.027 Å are observed between –COOH···HOOC–. Based on results obtained (Table 3.1 and Tables B1-B3 in Appendix B) it was concluded that (i) all self-constructed structures could be seen as sufficiently reliable for further theoretical calculations, and (ii) the RB3LYP/6-311+G(d,p) level of theory in conjunction with PCM/UA0 solvation model could be regarded as appropriate for the purpose of these studies. It is also clear that crystallographic structures can be used as a guide in constructing molecules but experimental data (bond lengths and angles) must be treated with caution when compared with DFT-optimized structures.

All solvent-optimized protonated forms of NTA, including the crystallographic structure H_3L^*, are shown in Figure 3.4 (those of the reference molecules are provided in Appendix B, Figures B3-B8). All protonated forms of the ligand NTA have considerably strong hydrogen bonds between oxygen on the $–COO^-$ or –COOH groups and a proton on nitrogen and they vary in length between about 2.0 and 2.35 Å. The shortest H-bond of 1.985 and 1.960 Å was found in H_3L and H_3L^*, respectively. The self-constructed and energy-minimized H_3L molecule differs significantly from H_3L^*. This is expected as all the carboxylic groups in H_3L form strong intramolecular H-bonds with the H-atom on nitrogen whereas one carboxylic group in H_3L^* (that is involved in very strong intermolecular interaction with another NTA molecule) is bend ‘downwards’ (it does not face the central nitrogen atom). It is then expected that these two structures of H_3L form of NTA should generated different prediction in protonation constants and it was of utmost interest to us to find out which form predicts more accurate values.

3.3.2: *Thermodynamic cycle*

In the field of metal-ligand equilibria studies, which is the focus of these studies, it is the protonation (K_H) instead of dissociation (K_a) constants that are used in solving mass-balance equations needed to develop most likely metal-ligand model and refine stability constants of all metal-containing species. There are a number of important compilations of ligand protonation and complex formation constants, among them are Martell and Smith[55] (which is used in this work) and a very recent one by IUPAC[74].

NTA has four protonation constants

$$L^{3-} + H^{+} \leftrightarrow HL^{2-} \qquad K_H^{(1)} = \frac{[HL^{2-}]}{[L^{3-}][H^{+}]} \tag{1}$$

$$\vdots \qquad\qquad \vdots$$

$$H_3L + H^{+} \leftrightarrow H_4L^{+} \qquad K_H^{(4)} = \frac{[H_4L^{+}]}{[H_3L][H^{+}]} \tag{2}$$

and there are several values reported at different ionic strengths (μ) and temperatures.[55, 74] The dissociation reactions of NTA can be written as

$$H_4L^{+} \leftrightarrow H_3L + H^{+} \qquad K_a^{(1)} = \frac{[H_3L][H^{+}]}{[H_4L^{+}]} \tag{3}$$

$$\vdots \qquad\qquad \vdots$$

$$HL^{2-} \leftrightarrow L^{3-} + H^{+} \qquad K_a^{(4)} = \frac{[L^{3-}][H^{+}]}{[HL^{2-}]}. \tag{4}$$

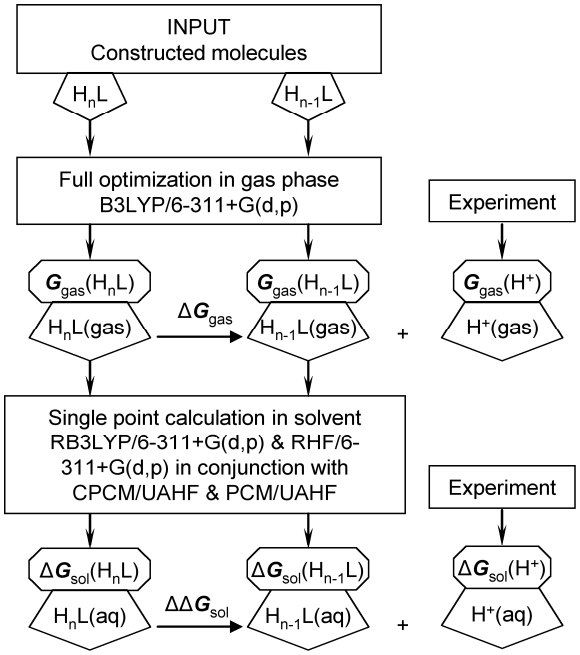
and the value of ΔG^{o}(aq), computed for the reactions given in eqs. 3-4, can be used, when TCs are employed, to calculate the *n*th stepwise dissociation constant at 25 ^{o}C, which is commonly reported as pK_a. A protonation reaction, on the other hand, is the reverse of a weak acid dissociation reaction and in the case of stepwise reactions the following holds

$$\log K_H^{(k)} = \log \frac{1}{K_a^{(n)}} = pK_a^{(n)} \tag{5}$$

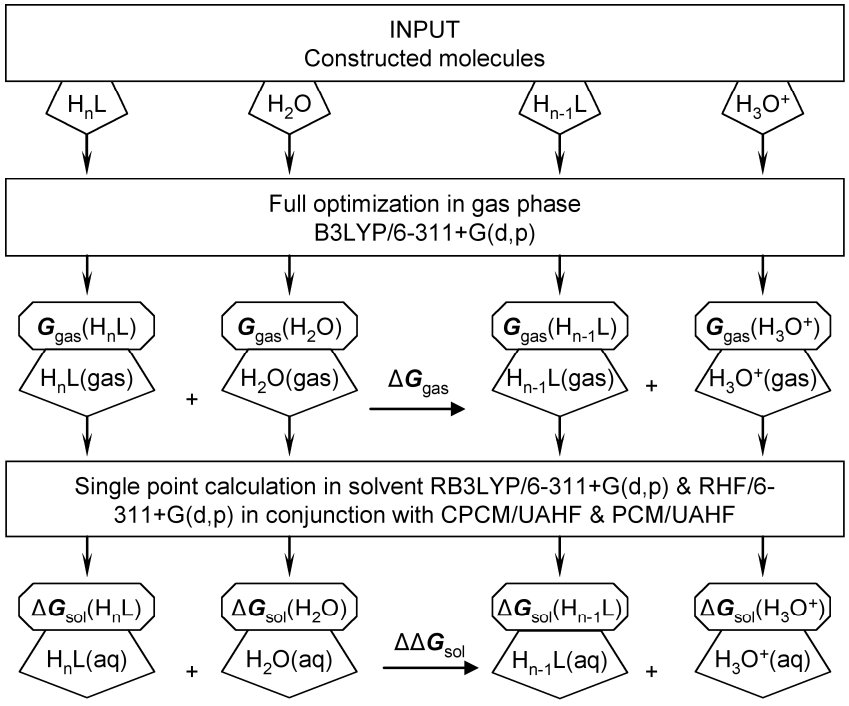
where $k = 1 + m - n$, m stands for the highest dissociation constant (here $m = 4$), n indicates an nth consecutive dissociation constant, $1 \leq n \leq m$, and k applies to a kth consecutive protonation constant, $1 \leq k \leq m$. Note that the ligand NTA has three acidic groups and only three dissociation constants would be reported and hence dissociation reaction (eq 3) most likely would not be considered as the first dissociation reaction, log $K_a^{(1)}$, in theoretical prediction of pK_a values employing TC-based methodology. However, due to the protonation/de-protonation of N-atom in NTA it is of paramount importance in the field of metal-ligand equilibria studies to consider also the first protonation constant, log $K_H^{(1)}$, as described by protonation reaction (eq 1). From this it follows that the 4th dissociation constant of NTA is linked through eq 5 with the 1st protonation constant of this ligand. The above is provided here for convenience and to assure clarity in nomenclature used in this chapter.

There were two, most commonly employed and best-performing, TCs considered in this work and they are shown in a self-explanatory fashion in Figure 3.5 as Scheme 1 and 2 (The charges on the ligand species are omitted for simplicity). In order to apply these TCs to the ligand NTA, each of the protonated forms needs to be optimized in gas phase first. However, there was a problem experienced when H_3L form of NTA was optimized in gas phase; the proton situated on the N-atom shifted to the $–COO^-$ group thereby forming the carboxylic group –COOH. A number of input structures were tested, but each time the proton on the N-atom was not preserved. The same phenomenon was also observed in literature for the ligand aspartic acid[12, 75] as well as in a recent report on NTPA.[53] This is not surprising as the moment solid NTA is placed in water (the H_3L solid reagent has three protons present on carboxylic groups, no proton is present on the N-atom) the N-atom is protonated instantly and at least one proton from an acetic acid arm dissociates fully; an instantaneous proton replacement takes place in water.

Unfortunately, in gas phase the molecule H_3L does not exist in zwitterionic form, and because of this the TCs could not be applied to the H_3L and H_4L forms of NTA and theoretical studies were restricted to only the first two protonation reactions from which HL and H_2L were formed. Table 3.2 shows minimum energies after zero-point vibrational energy (ZPVE) corrections, $E_{min,}$ together with the values of G_{gas} of the optimized molecules. The single point energy values (ΔG_{sol}) that were calculated using different levels of theory and solvation models are given in Table 3.3.



SCHEME 1



SCHEME 2

Figure 3.5: Schematic representation of TCs employed in this chapter.

Values of ΔG_{gas}, $\Delta\Delta G_{sol}$, and ΔG_{aq} were calculated using well known reported relationships (eqs 6-8),[6] and dissociation constants $K_a^{(n)}$ were obtained from eq 9 and 10[8, 47] in combination with TC-1 and TC-2, respectively.

$$\Delta G_{gas} = \sum G_{gas}(\text{products}) - \sum G_{gas}(\text{reactants}) \quad (6)$$

$$\Delta\Delta G_{sol} = \sum \Delta G_{sol}(\text{products}) - \sum \Delta G_{sol}(\text{reactants}) \quad (7)$$

$$\Delta G_{aq} = \Delta G_{gas} + \Delta\Delta G_{sol} \quad (8)$$

$$\Delta G_{aq}^{(n)} = -RT \ln K_a^{(n)} \quad (9)$$

$$\Delta G_{aq}^{(n)} - RT \ln [H_2O] = -RT \ln K_a^{(n)} \quad (10)$$

It can be seen in Figure 3.5 (Scheme 1) that a free proton is involved, but since theoretical prediction of G_{gas} and ΔG_{sol} is considerably difficult, experimental values for these two quantities are commonly used now. The values used were –6.28 and –263.98 kcal/mol for G_{gas} and ΔG_{sol}, respectively, which are the most recent reported values to date.[11, 38] A correction of –1.89 kcal/mol (corresponding to a state change of 1 mol of gas from 1 atm (24.47 L/mol) to 1 M (1 mol/L)) was made to the calculated solvation free energy, which is discussed thoroughly by Jang et al.[26]

Table 3.2: Selected thermochemical data (E_{min} stands for ZPVE-corrected energy) obtained for indicated NTA species, H_2O, and H_3O^+ computed at the RB3LYP/6-311+G(d,p) level of theory in conjunction with PCM/UA0 solvation model.

	Gas-phase optimized structures		**Solvent optimized structures**
Species	E_{min} [a]	G_{gas} [a]	E_{min} [a]
L^{3-}	–738.348636	–738.390043	–738.936972
HL^{2-}	–739.124533	–739.166015	–739.410772
H_2L^-	–739.727610	–739.769183	–739.852748
H_2O	–76.437174	–76.454816	–76.452207
H_3O^+	–76.696787	–76.714893	–76.840274

[a]) In atomic unit, Hartree (1 Hartree = 627.5095 kcal mol^{-1})

Table 3.3: Single point calculated, at different levels of theory (a 6-311+G(d,p) basis set was used) in conjunction with indicated solvation model, values of ΔG_{sol}, reported in kcal mol^{-1}.

	Gas-phase optimized structures				Solvent optimized structures			
	RB3LYP		RHF		RB3LYP		RHF	
Species	ΔG_{sol}[a]	ΔG_{sol}[b]	ΔG_{sol}[a]	ΔG_{sol}[b]	ΔG_{sol}[a]	ΔG_{sol}[b]	ΔG_{sol}[a]	ΔG_{sol}[b]
L^{3-}	–377.02	–377.31	–387.48	–387.81	–383.71	–384.09	–395.27	–395.70
HL^{2-}	–186.51	–186.78	–196.24	–196.56	–192.72	–193.08	–203.52	–203.95
H_2L^-	–78.26	–78.60	–87.14	–87.53	–88.26	–88.69	–97.99	–98.47
H_2O	–6.67	–6.72	–7.01	–7.07	–6.84	–6.90	–7.22	–7.28
H_3O^+	–107.21	–107.35	–107.65	–107.79	–109.63	–109.78	–110.23	–110.39

[a]) PCM/UAHF, [b]) CPCM/UAHF.

The calculated first two protonation constants for the ligand NTA, using TC-1 and TC-2 are summarized in Table 3.4, but unfortunately, most of the results obtained are far from satisfactory with differences (δ) reaching 10 log units in case of the first protonation constant.

In an attempt to improve the prediction of the computed protonation constants, a slightly modified approach was applied. This involved full energy optimisation of each of the ligand species, water and the hydronium ion in solvent at the RB3LYP/6-311+G(d,p) level of theory in conjunction with PCM/UA0 solvation model followed by a single point calculation (SPC) in solvent using both, RHF/6-311+G(d,p) and RB3LYP/6-311+G(d,p), levels of theory in combination with the PCM/UAHF and CPCM/UAHF solvation models, respectively, in order to generate required ΔG_{sol}. A similar approach[18] resulted in somewhat improved computed dissociation constants and, in general, the same was observed here, but still most of results obtained are unacceptable – only the second protonation constant, when TC-2 was utilised and computation was performed at the RB3LYP/6-311+G(d,p) level of theory in solvent (either PCM/UAHF or CPCM/UAHF), was within one log unit of the experimental data. No consistency in computed protonation constants was achieved.

From Table 3.4 it is seen that (i) TC-1 worked better for the first protonation constant for both, the gas-phase and solvent structures, as was the case with previous studies[53] , (ii) solvent-optimized structures generated smaller errors in computed first and second protonation constants when compared with the equivalent gas-phase structure values.

Table 3.4: Comparison of experimental[55] and calculated (from thermodynamic cycles) protonation constants, as log K_H, using gas-phase and solvent optimized structures of NTA seen in Figure 3.4. 6-311+G(d,p) basis set was used.

	Gas-phase structure					Solvent Structure			
					RHF				
Reaction	**Exp**[a]	**TC-1**	δ[b]	**TC-2**	δ[b]	**TC-1**	δ[b]	**TC-2**	δ[b]
$L^{3-} + H^+ =$	10.334	17.25[c]	6.92	20.22[c]	9.89	16.88[c]	6.55	18.10[c]	7.77
HL^{2-}		17.25[d]	6.92	20.15[d]	9.82	16.88[d]	6.55	18.03[d]	7.70
$HL^{2-} + H^+ =$	2.94	–2.02[c]	–4.96	0.94[c]	–2.00	0.60[c]	–2.34	1.82[c]	–1.12
H_2L^-		–1.97[d]	–4.91	0.93[d]	–2.01	0.63[d]	–2.31	1.78[d]	–1.16
					RB3LYP				
$L^{3-} + H^+ =$	10.334	17.79[c]	7.46	20.82[c]	10.49	17.44[c]	7.11	18.82[c]	8.49
HL^{2-}		17.78[d]	7.45	20.74[d]	10.41	17.42[d]	7.09	18.74[d]	8.41
$HL^{2-} + H^+ =$	2.94	–1.40[c]	–4.34	1.64[c]	–1.30	1.38[c]	–1.56	2.77[c]	–0.17
H_2L^-		–1.35[d]	–4.29	1.62[d]	–1.32	1.43[d]	–1.51	2.75[d]	–0.19

[a]) μ = 0.0 M and 20 °C, [b]) δ = (computed – experimental) value, [c]) The PCM/UAHF solvation model was used for SPCs, [d]) The CPCM/UAHF solvation model was used for SPCs.

When the influence of different levels of theory and solvation models used for SPCs were tested, in case of the first protonation constant, when TC-1 was used, the RHF/6-311+G(d,p) worked better than the RB3LYP/6-311+G(d,p) level of theory, and in the case of TC-2 the CPCM/UAHF model yielded slightly improved values when compared to the PCM/UAHF solvation model. For some reason the RB3LYP/6-311+G(d,p) level of theory in combination with PCM/UAHF model applied to TC-2 yielded the best results for the second protonation constant. This was unexpected as the UAHF radii were optimized for HF[18] and one would expect the RHF/6-311+G(d,p) level of theory to yield better results, as was the case with the prediction of the first protonation constant.

Nevertheless, there seems to be no obvious pattern in the data obtained in Table 3.4 and, on average, results obtained are totally unacceptable due to large differences (δ) between computed and experimental values; possible sources of errors in the computed values are discussed extensively elsewhere.[8, 76] Since all necessary protonated forms of NTA could not be optimized in gas phase, further investigations involving the TCs, not only those seen in Figure 3.5, were not considered a justifiable option.

3.3.3: *Isodesmic Reactions*

Since application of TCs provided results that were unacceptable, attention was placed on the applicability of methodology based on isodesmic reactions (IRns)[77]. To date, IRns have been utilized in the prediction of enthalpies of formation[77-85] and in some cases they have been incorporated within TCs to predict dissociation constants.[1, 3-6] It is expected that the implementation of IRn 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 and every component involved in the reaction of interest. An application of IRn is dependent on the availability of experimental data, such as protonation constants, for the reference species used.[54, 78] To date there has been only one explicit application of IRn in theoretical studies of protonation/dissociation constants;[53] only different kinds of TCs were employed till now and this is probably due to the fact that studies have been restricted to mono-dissociable organic acids, for which TCs might work reasonably well. When considering poly-negatively charged ions, such as the fully deprotonated NTA ligand, for example, the presently available solvation models introduce large errors in protonation constants generated from TC-based computations. This is predominantly due to their inability to contain the large amounts of charge within the solvation sphere. With IRns these errors are expected to be significantly eliminated and by doing so they should provide more accurately predicted protonation constant values.

However, the main challenge associated with the use of IRn is the selection of appropriate reference molecule and, particularly in this case, the most appropriate protonated form of the reference molecule to be included in one of the fourth isodesmic reactions needed to compute four protonation constants of NTA. By considering the structural properties of NTA (called further $L_{(1)}$) the ligands IDA, MIDA, EIDA, PIDA and HIDA were used 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 solution. In addition, protonation constants for all of the chosen reference molecules are well known as they are widely-studied ligands.[55, 74] The impact of the kind of a reference molecule and its protonated form used in the prediction of protonation constants is evaluated here.

IRn 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

$$L_{(1)}(aq) + HL_{(2)}(aq) = HL_{(1)}(aq) + L_{(2)}(aq) \qquad \Delta G_{IRn}^{(1)}(aq) \qquad (11)$$

and for each *k*th consecutive IRn (there must be at least four IRns involved here since NTA has four protonation constants, but one might consider more IRns when different protonated forms of a reference molecule are used) one can obtain relevant change in *G*, $\Delta G_{IRn}^{(k)}(aq)$. Each of the two ligands (NTA and a reference molecule) is involved in several stepwise protonation reactions and for simplicity only the first ones are shown as eqs 12 and 13

$$H^+ + L_{(1)} \leftrightarrow HL_{(1)} \qquad \Delta G_{p_{L(1)}}^{(1)}(aq) = -\Delta G_{d_{L(1)}}^{(4)}(aq) \qquad (12)$$

$$H^+ + L_{(2)} \leftrightarrow HL_{(2)} \qquad \Delta G_{p_{L(2)}}^{(1)}(aq) = -\Delta G_{d_{L(2)}}^{(3)}(aq) \qquad (13)$$

where $\Delta G_{p_{L(1)}}^{(k)}(aq)$ and $\Delta G_{d_{L(1)}}^{(n)}(aq)$ refers to the *k*th (here $k = 1$) stepwise protonation (p) reaction and the reverse and relevant *n*th (here $n = 4$ or 3 for NTA and reference molecule, respectively) stepwise dissociation (d) reaction, respectively, as described by eq 5. The change in Gibbs energies for each protonation reaction 12 and 13 can be written, respectively, as

$$\Delta G_{p_{L(1)}}^{(1)}(aq) = G_{aq}(HL_{(1)}) - G_{aq}(H^+) - G_{aq}(L_{(1)}) \qquad (14)$$

$$\Delta G_{p_{L(2)}}^{(1)}(aq) = G_{aq}(HL_{(2)}) - G_{aq}(H^+) - G_{aq}(L_{(2)}) \qquad (15)$$

The isodesmic reaction of interest (eq 11) can be obtained from subtracting eq 13 from 12, and hence from subtracting eq 15 from 14 one obtains an expression for the change in Gibbs energy, $\Delta G_{IRn}^{(1)}(aq) = \Delta G_{p_{L(1)}}^{(1)}(aq) - \Delta G_{p_{L(2)}}^{(1)}(aq)$ applicable to this isodesmic reaction, where the uncertainty related to $G_{aq}(H^+)$ is no more applicable as this term cancels of, and hence any error that might have been introduced by the use of an experimental value for this quantity is eliminated

$$\begin{aligned}\Delta G_{IRn}^{(1)}(aq) &= \Delta G_{p_{L(1)}}^{(1)}(aq) - \Delta G_{p_{L(2)}}^{(1)}(aq) \\ &= G_{aq}(HL_{(1)}) - G_{aq}(L_{(1)}) - G_{aq}(HL_{(2)}) + G_{aq}(L_{(2)}). \qquad (16)\end{aligned}$$

Eq 16 was used to calculate $\Delta G_{IRn}^{(1)}(aq)$ of IRn (eq 11) from appropriate Gibbs energies obtained for relevant and fully solvent-optimized structures of the ligand NTA and reference ligand $L_{(2)}$. Table B4 in Appendix B 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. One will notice that there are two energies reported for the H_3L form of the ligands NTA, one of which refers to the optimized crystal structure H_3L^* and the other to the optimized self-constructed structure H_3L, both of which were utilized in the calculation of stepwise protonation constants of NTA in order to determine which yielded better results. The value for $\Delta G_{p_{L(2)}}^{(1)}(aq)$, or in general $\Delta G_{p_{L(2)}}^{(k)}(aq)$, was obtained from eq 9 using the reported protonation constants[55] (at 20 and 25 ^oC and μ = 0.0 and 0.1 M) of the reference ligand $L_{(2)}$, either IDA, MIDA, EIDA, PIDA, or HIDA. Once $\Delta G_{IRn}^{(1)}(aq)$ and $\Delta G_{p_{L(2)}}^{(1)}(aq)$ have been calculated, the value of $\Delta G_{p_{L(1)}}^{(1)}(aq)$, which is needed to calculate the protonation constants for NTA, was obtained by simply rearranging eq 16 ($\Delta G_{p_{L(1)}}^{(1)}(aq) = \Delta G_{IRn}^{(1)}(aq) - \Delta G_{p_{L(2)}}^{(1)}(aq)$).

Table 3.5 provides values for the functions required to calculate protonation constants, calculated and experimental protonation constants of the ligand NTA, along with differences between calculated and experimental protonation constants (δ). Values for $\Delta G_{p_{L(2)}}^{(k)}(aq)$ were calculated from the experimentally available protonation constants[55], which have been reproduced in Appendix B, Table B5. There were a number of different IRns that were tested, but only those that produced the best results are shown in Table 3.5 (The remaining results are provided in Appendix B, Table B6). From Table 3.5 it can be seen that only the reactions with the optimized self-constructed H_3L form of NTA are shown. Reactions with the optimized crystallographic structure (H_3L^*) were also tested, but they didn't produce results that were considered acceptable; the computed values differed from experimental protonation constants by more than one log unit (these results can be obtained from Appendix B, Table B6). Poor predictions in protonation constants when H_3L^* was used was most likely due to the fact that in a solution the intermolecular interactions (present in the solid state) must disappear and hence the structure of H_3L^* is not suitable for the protonation constants calculations in solvent. It appears that, since H_3L of NTA generated acceptabe protonation constants, the H_3L conformer of NTA might be much closer approximation of structural configuration of the ligand in solvent when bound to three protons.

Table 3.5: Comparison of experimental (Exp)[55] 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)	$\Delta G_{p_{L_{(1)}}}$(aq)	log K_H	Exp[a]	δ	$\Delta G_{p_{L_{(1)}}}$(aq)	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
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)} + HL_{(2)}^{-}$	1.054	–1.948	1.43	2.00[c]	–0.57	–1.948	1.43	1.81	–0.38
$H_3L_{(1)} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L_{(2)}$	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.

This is not surprising when one is analyzing the H_3L^* structure shown in Figures 3.2 and 3.3, where one of the acetic acid arms is bent out of the N-atom and is involved in strong H-bond with another NTA molecule in solid state. The above strongly suggests that in order to generate most accurate theoretically computed protonation constants one would have to generate structures of all molecules (the studied and reference ones) that are as close as possible to the ones existing in solvent, but this is not a trivial and easy to overcome problem.

Even though some results obtained with most of the reference molecules are not very accurate (there are differences between calculated and experimental protonation constants that are larger than one log unit) the IRns have still proven to yield results that are superior when compared to those produced by application of TCs (Table 3.4). There are several interesting observations one can make from the analysis of data seen in Table 3.5. As one goes from IDA to HIDA, it is seen that the prediction of protonation constants gets more accurate, with HIDA yielding the most accurate estimates (all predicted values have less than one log unit difference when compared to the experimental protonation constants of NTA at μ = 0.0 M). The observed trend can be linked with an increase in structural similarity between the reference and studied molecules. Another and very important fact is that all IRns seen in Table 3.5 predicted four protonation constants in correct order, with the first and the fourth one being the largest and smallest one, respectively, and with consecutive decrease in value. One must realise 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. There is also another pattern that emerges with these predictions, namely, the use of the HL^-, H_2L, and H_3L^+ protonated form of any reference molecule studied here (equivalent to the use of the first, second and third protonation constants of these molecules) resulted in the best estimates of the first, second as well third, and the forth protonation constant of NTA, respectively, involving L^{3-} + HL^{2-}, HL^{2-} + H_2L^- as well as H_2L^- + H_3L, and H_3L + H_4L^+, respectively, protonated forms of NTA. This strongly suggests that not only structural similarity but also the size of a charge on a reference molecule plays an important role. It appears that the more similar charges are on the studied and reference molecule, the better prediction is achieved. The least accurate computed log K_H, regardless of the reference molecule used, was obtained always for the first protonation reaction of NTA that involves the most negatively charge form of the ligand NTA, $L_{(1)}^{3-}$, and the reference molecule $L_{(2)}^{2-}$.

For all reference molecules, except HIDA, the difference between the computed and experimental values (δ) was above one log unit and the value of δ (computed with experimental protonation constants of NTA at μ = 0.0 M) decreases systematically from 4.81 to 0.94 when going from IDA to HIDA. Interestingly, for all reference molecules used the smallest errors in the predicted log K_H values was 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 rationalise this observation in terms of a charge placed on the studied and reference molecules. In these IRns, 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 $K_H^{(3)}$ were obtained regardless of the reference molecule studied here, one might conclude that the size of a charge on a molecule place more decisive role than its structural similarity to NTA when accuracy in predicted protonation constants is considered. Unexpectedly, as far as generally accepted knowledge related to poor performance of TC in the case of negative charges present on a molecule is taken into consideration[54], somewhat worse estimates in log $K_H^{(4)}$ values (when compared with accuracy achieved for the second and third protonation constants) were obtained when neutral and singly positively charged forms of molecules were involved in IRns – this applies to all systems studied here. This is significant observation because many important ligands (among them macrocyclic ligands) do not have carboxylic groups (they are neutral in their fully deprotonated forms) and when protonated they might have multiple and positive charge on them; clearly this requires a dedicated investigation in order to establish whether IRn-based protocol (or TCs) can be applied successfully.

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 (in HIDA) being part of the –OH instead of –COOH group. All the other reference molecules have only two O-atoms. The structural similarity of HIDA to NTA correlates well with results seen in Table 3.5 as the best estimates were generated with the involvement of HIDA in the IRns. 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 a 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 of atoms, especially (in this case) to the positioning of heteroatoms, because 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.

From the data seen in Table 3.5 it was established that HIDA is the best reference molecule to use for the purposes of these studies, but it was also of interest to determine if conformational analysis of the molecules involved would yield even more accurate results. All forms of NTA and reference molecules were subjected to Schrödinger's Maestro[86] conformational analysis. This software automatically generates hundreds of possible conformers and estimates their energies in a short period of time based on MM/MD (Molecular Mechanics/Molecular Dynamics) principles. The structures of NTA seen in Figure 3.4 were used as inputs to Maestro (input structures of the reference molecules are provided in Appendix B, Figure B3-B8) and MM/MD-based conformational analysis was performed in solvent.

Table 3.6(a) provides energies (in kJ/mol) of five lowest in energy MM/MD-conformers (C-1 to C-5) of all the protonated forms of NTA seen in Figure 3.4. MM/MD-based SPCs in solvent were also performed on the DFT-structures shown in Figure 3.4 in order to compare these energies with the MM/MD-generated lowest energy conformers – obtained data are also included in Table 3.6(a). Since all of the MM/MD C-1 conformers of all the protonated forms of NTA were of significantly lower energy than the MM/MD SPC- energies obtained on the initial structures, all of the C-1 conformers were fully DFT-optimized in solvent – results obtained are shown in Table 3.6(b). Surprisingly, only two out of six of the DFT-optimized C-1 conformers had lower energies when compared with the energies of the original structures seen in Figure 3.4 (optimized with Gaussian). Most important and gratifying was the fact that the differences δG in Table 3.6(b) between the relevant structures became almost negligibly small; for all protonated forms, except H_4L^+, the value of δG is about ± 0.1 kcal mol^{-1} that is equivalent of about 0.07 of the computed log K_H value, a typical uncertainty in experimentally determined protonation constants.

Table 3.6: Part (a). Minimum energies of MM/MD-generated conformers in solvent (E_{C1}-E_{C5}) and energies obtained from MM-based SPC (E_{SPC}) performed on the NTA structures seen in Figure 3.4. Part (b). DFT-calculated solvent-optimized energies of all protonated forms of the ligand NTA (E_{min} = ZPVE-corrected energy) of structures seen in Figure 3.4 and lowest energy MM/MD-generated C-1 conformers.

(a)

L = NTA	E_{SPC}	E_{C-1}	δE[a] kJ/mol	δE[a] kcal/mol	E_{C-2}	E_{C-3}	E_{C-4}	E_{C-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]) $\delta E = E_{SPC} - E_{C-1}$

(b)

	Structures seen in Fig. 4		C-1 structures			
L = NTA	E_{min} (Hartree)	G_{aq} (Hartree)	E_{min} (Hartree)	G_{aq} (Hartree)	δG[b] Hartree	δG[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]) $\delta G_{aq} = G_{aq}$(structure in Fig. 4) – G_{aq}(C-1)

Similar procedure was applied to all the protonated forms of IDA, MIDA, EIDA, PIDA and HIDA, but for simplicity and since the use of HIDA resulted in the best predictions for the protonation constants of NTA only the results for HIDA are shown in Table 3.7 (results for the other reference molecules studied in this chapter are provided in Appendix B, Table B7). Although all MM/MD-structures of HIDA were lower in energy when compared with the relevant MM/MD SPC-energies obtained on the appropriate DFT-structures (see Table 3.7(a), where $\delta E = E - E_{C-1} > 0$), the results after full-energy optimization in Gaussian suggest opposite for HL^{-} and H_2L where $\delta G < 0$ (see Table 3.7(b)). It was again gratifying to see that the δG values were much smaller than the δE values but this time the largest difference is observed for the free ligand HIDA, L^{2-}.

Table 3.7: Part (a). Minimum energies of MM/MD-generated conformers in solvent (E_{C1}-E_{C5}) and energies obtained from MM-based SPC performed on the HIDA structures seen in Figure B8. Part (b). DFT-calculated solvent-optimized energies of all protonated forms of the ligand HIDA (E_{min} = ZPVE-corrected energy) of structures seen in Figure B8 and lowest energy MM/MD-generated C-1 conformers.

(a)

L = HIDA	E_{SPC}	E_{C-1}	δE[a] kJ/mol	δE[a] kcal/mol	E_{C-2}	E_{C-3}	E_{C-4}	E_{C-5}
L^{2-}	–505.84	–557.49	51.66	12.35	–557.49	–555.33	–555.33	–552.63
HL^{-}	–869.14	–906.75	37.61	8.99	–906.75	–905.39	–905.39	–905.17
H_2L	–639.60	–681.53	41.93	10.02	–681.17	–681.13	–680.21	–680.06
H_3L^{+}	–379.57	–426.26	46.69	11.16	–426.25	–422.83	–422.82	–421.98

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

(b)

	Structures seen in Fig. S4		C-1 structures			
L = HIDA	E_{min} (Hartree)	G_{aq} (Hartree)	E_{min} (Hartree)	G_{aq} (Hartree)	δG[b] Hartree	δG[b] kcal/mol
L^{2-}	–665.299655	–665.340516	–665.305277	*–665.344936*	0.004420	2.77
HL^{-}	–665.767828	*–665.808578*	–665.766303	–665.807676	–0.000902	–0.57
H_2L	–666.207584	*–666.248534*	–666.205011	–666.246706	–0.001828	–1.15
H_3L^{+}	–666.642422	–666.683405	–666.642406	*–666.683411*	0.000006	0.004

[b]) $\delta G_{aq} = G_{aq}$(structure in Fig. 4) – G_{aq}(C-1)

With a hope on improving predictions in protonation constants, data from Tables 3.6 and 3.7 was selected that had the lowest DFT-computed G_{aq} values (printed in Italic) for NTA and HIDA and used them for the protonation constant calculations based on the IRn approach discussed in this work – results obtained are shown in Table 3.8. Even though, as it can be seen from Table 3.6(b), the H_3L^{*} structure of NTA possess an energy that is lower than that of H_3L, only H_3L was used and this is due to the fact that, as discussed earlier, the H_3L^{*} form is most likely not the one present in a solution, it had not generated acceptable results and hence it was not included in Table 3.5 already. Also, in Table 3.8 only the results obtained when HIDA was used as a reference molecule are presented (the remaining results are provided in Appendix B, Table B8). From Table 3.8 it is seen that the differences, when compared with the results obtained in Table 3.5, show that there has actually been a drop in accuracy with a very significant error in prediction observed for the first protonation constant of NTA. This decrease in accuracy might be due to the fact that either (i) analysis on a

number of MM/MD generated conformers in search for a global minimum energy structure was not performed, (ii) conformers generated by conformational analysis (performed on a single molecule in a simulated solvent environment) still differ significantly from those existing in a real solution, or (iii) computational solvation models used in energy optimization do not perform well enough and generate small, but significant, errors depending particularly on a charge present on a molecule. One might assume that most accurate protonation constants could be computed with the use of real conformers formed in a solution but, on one hand there is not an easy and obvious to follow protocol to find them, and, on the other hand, there is no guaranty that when optimized with a use of solvation models available, the computed energies would not carry significant errors.

Table 3.8: Comparison of experimental (Exp)[55] and calculated stepwise protonation constants of NTA ($L_{(1)}$), as log K_H, using lowest energy structures from Tables 3.6 and 3.7 (seen in Italic) with protonation constants of the reference molecule HIDA ($L_{(2)}$) at ionic strength μ = 0.1 M and 25 °C.

Reaction	**log K_H**	**Exp[a]**	**δ**	**log K_H**	**Exp[b]**	**δ**
$L_{(1)}^{3-} + HL_{(2)}^{-} = HL_{(1)}^{2-} + L_{(2)}^{2-}$	13.39	10.334	3.05	13.39	9.66	3.73
$HL_{(1)}^{2-} + H_2L_{(2)} = H_2L_{(1)}^{-} + HL_{(2)}^{-}$	3.22	2.94	0.28	3.22	2.52	0.70
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)} + HL_{(2)}^{-}$	1.35	2.00[c]	–0.65	1.35	1.81	–0.46
$H_3L_{(1)} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L_{(2)}$	0.42	1.00[b]	–0.58	0.42	1.00	–0.58

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

From this it was concluded that in order to perform proper structural analysis in search of the global minimum energy conformer in a solution, three possible and obvious procedures could be followed; (i) an analysis of a number of the MM/MD conformational structures generated, for instance the first five lowest energy conformers, by performing full DFT-based energy optimization on them, (ii) a full DFT-based conformational analysis in solvent, which is a considerably time consuming task and due to erratic performance of solvation models there would still be uncertainty related to the selection of a structure that indeed predominates in a solution, or (iii) energy optimization performed on a ligand with an inclusion of a primary solvation layer. The latter option would have to include an investigation of the influence of a

number of water molecules directly interacting with a molecule of interest (the number would have to be established by try and error approach) on conformational analysis of all possible forms of NTA and reference molecules and this is an extremely time consuming exercise. However, the main aims of this work were (i) to prove the point that it is indeed possible to theoretically predict four consecutive protonation constants with acceptable accuracy with an average error in prediction below one log unit, (ii) an evaluation of an impact the selection of reference molecule (and its protonated form) has on accuracy of the computed log K_H values, and (iii) establishing which one, structural similarity or charge distribution, makes larger impact on the prediction of protonation constants. Since there was no attempt to find out how accurate that prediction might be, neither the full conformational analysis nor considering of a primary solvation layer was performed. One might realise that the reported predictions are already within one log unit for all four stepwise protonation constants and hence one might question the necessity of such time-consuming possible approaches in this study when compared with relatively small gain in the predicted values (a fraction of a log unit). From a theoretical point of view and in order to find out most rigid procedure necessary to establish unknown protonation constants of a molecule of interest it would be of particular interest and importance to include water molecules in conformational analysis and appropriate investigation is currently under way. Nevertheless, it appears that MM/MD might be a useful tool in a quick and preliminary search of conformers for the purpose of this kind of study. The self-constructed and MM/MD generated lowest energy conformers, for all molecules utilized in this work after DFT energy minimisation, are provided in Appendix B, Figure B3-B8).

3.4. Conclusions

This chapter has successfully 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 possible with acceptable accuracy when isodesmic reaction methodology is employed. The application of the commonly employed thermodynamic cycles once again was proved to be unreliable methodology that most likely should not be employed in this kind of a study. In this chapter it was established that in order to obtain results that are within ± 1 log unit from the experimental data for all protonation constants, a reference molecule must be chosen in such a way that it is structurally (the size it occupies in solvation models

utilised, the cavity, kind and a number of charges as well as their distribution within a molecule) almost identical to the molecule of interest. Results obtained here strongly suggest that the charge distribution within the molecule is of utmost importance – all reference molecules investigated here (IDA, MIDA, EIDA, PIDA and HIDA) were suitable for the prediction of the second and third protonation constants of NTA – and this is most likely due to the solvation models available for energy minimisation in polar solvents, such as water. From all reference molecules tested during these studies HIDA provided the most accurate predicted protonation constants, particularly for the first and the forth protonation constants of NTA – it appears that structural (the kind and a number of donor atoms) and charge similarities with NTA played most important role here. Available crystallographic structures of H_3L form of NTA and HIDA were found to be unsuitable (when used in isodesmic reaction they generated poor predictions in protonation constants of NTA) and this is most likely because of strong intermolecular interactions in a solid phase that made their structures very much different when compared with expected structures in solvent; the use of crystallographic structures must be treated with caution. It is clear that the next necessary and logical step in the theoretical studies of stepwise protonation (or dissociation) constants, with an aim of achieving computed values within a small fraction of a log unit when compared with experimental values, is the development of methodologies suitable in the prediction of conformers that are most likely formed in a solution as they are involved in real protonation (dissociation) reactions.

3.5. References

1. M. Namazian, F. Kalantary-Fotooh, M. R. Noorbala, D. J. Searles, M. C. Coote, *J. Mol. Struct. (Theochem)* **2006**, *758*, 275.

2. D. Jacquemin, E. A. Perpete, I. Ciofini, C. Adamo, *J. Phys. Chem. A* **2008**, *112*, 794.

3. M. Namazian, H. Heidary, *J. Mol. Struct. (Theochem)* **2003**, *620*, 257.

4. M. Namazian, S. Halvani, M. R. Noorbala, *J. Mol. Struct. (Theochem)* **2004**, *711*, 13.

5. M. Namazian, S. Halvani, *J. Chem. Thermodyn.* **2006**, *38*, 1495.

6. M. Namazian, M. Zakery, M. R. Noorbala, M. L. Coote, *Chem. Phys. Lett.* **2008**, *451*, 163.

7. I. E. Charif, S. M. Mekelleche, D. Villemin, N. Mora-Diez, *J. Mol. Struct. (Theochem)* **2007**, *818*, 1.

8. M. D. Liptak, G. C. Shields, *Int. J. Quantum Chem.* **2001**, *85*, 727.

9. M. D. Liptak, G. C. Shields, *J. Am. Chem. Soc.* **2001**, *123*, 7314.

10. M. D. Liptak, K. C. Gross, P. G. Seybold, S. Feldgus, G. C. Shields, *J. Am. Chem. Soc.* **2002**, *124*, 6421.

11. M. T. Bensen, M. L. Moser, D. R. Peterman, A. Dinescu, *J. Mol. Struct. (Theochem)* **2008**, *867*, 71.

12. W. Sang-Aroon, V. Ruangpornvisuti, *Int. J. Quantum Chem.* **2008**, *108*, 1181.

13. Y. H. Jang, W. A. Goddard III, K. T. Noyes, L. C. Sowers, S. Hwang, D. S. Chung, *J. Phys. Chem. B* **2003**, *107*, 344.

14. L. C. Jitariu, A. J. Masters, I. H. Hiller, *J. Chem. Phys* **2004**, *121*, 7795.

15. V. S. Bryantsev, M. S. Diallo, W. A. Goddard III, *J. Phys. Chem. A* **2007**, *111*, 4422.

16. D. Gao, P. Svoronos, P. K. Wong, D. Maddalena, J. Hwang, H. Walker, *J. Phys. Chem. A* **2005**, *109*, 10776.

17. K. Murlowska, N. Sadlej-Sosnowska, *J. Phys. Chem. A* **2005**, *109*, 5590.

18. G. A. A. Saracino, R. Improta, V. Barone, *Chem. Phys. Lett.* **2003**, *373*, 411.

19. C. O. da Silva, E. C. da Silva, M. A. C. Nascimento, *J. Phys. Chem. A* **1999**, *103*, 11194.

20. G. Li, Q. Cui, *J. Phys. Chem. B* **2003**, *107*, 14521.

21. E. E. Dahlke, C. J. Cramer, *J. Phys. Org. Chem.* **2003**, *16*, 336.

22. A. Klamt, F. Eckert, M. Diedenhofen, M. E. Beck, *J. Phys. Chem. A* **2003**, *107*, 9380.

23. C. P. Kelly, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. A* **2006**, *110*, 2493.

24. J. R. Pliego Jr., J. M. Riveros, *J. Phys. Chem. A* **2002**, *106*, 7434.

25. J. J. Klicic, R. A. Friesner, S. Liu, W. C. Guida, *J. Phys. Chem. A* **2002**, *106*, 1327.

26. Y. H. Jang, L. C. Sowers, T. Cagin, W. A. Goddard III, *J. Phys. Chem. A* **2001**, *105*, 274.

27. F. Himo, L. A. Eriksson, M. R. A. Blomberg, P. E. M. Siegbahn, *Int. J. Quantum Chem.* **2002**, *76*, 714.

28. F. Himo, L. Noodleman, M. R. A. Blomberg, P. E. M. Siegbahn, *J. Phys. Chem. A* **2002**, *106*, 8757.

29. D. M. Chipman, *J. Phys. Chem. A* **2002**, *106*, 7413.

30. J. I. Mujika, J. M. Mercero, X. Lopez, *J. Phys. Chem. A* **2003**, *107*, 6099.

31. C. Scharnagl, R. A. Raupp-Kossmann, *J. Phys. Chem. B* **2004**, *108*, 477.

32. I. Lee, C. K. Kim, I. Y. Lee, C. K. Kim, *J. Phys. Chem. A* **2000**, *104*, 6332.

33. Z. B. Maksic, R. Vianello, *J. Phys. Chem. A* **2002**, *106*, 419.

34. A. M. Magill, K. J. Cavell, B. F. Yates, *J. Am. Chem. Soc.* **2003**, *126*, 8717.

35. H. Li, A. W. Hains, J. E. Everts, A. D. Robertson, J. H. Jensen, *J. Phys. Chem. B* **2002**, *106*, 3486.

36. C. O. da Silva, E. C. da Silva, M. A. C. Nascimento, *J. Phys. Chem. A* **2000**, *104*, 2402.

37. C. O. da Silva, E. C. da Silva, M. A. C. Nascimento, *Chem. Phys. Lett.* **2003**, *381*, 244.

38. M. Alvarado-Gonzalez, E. Orrantia-Borunda, D Glossman-Mitnik, *J. Mol. Struct. (Theochem)* **2008**, *869*, 105.

39. M. Sulpizi, M. Sprik, *Phys. Chem. Chem. Phys.* **2008**, *10*, 5238.

40. L. Tao, J. Han, F. M. Tao, *J. Phys. Chem. A* **2008**, *112*, 775.

41. G. Kieseritzky, E. W. Knapp, *J. Comput. Chem.* **2008**, *29*, 2575.

42. G. da Silva, E. M. Kennedy, B. Z. Dlugogorski, *J. Phys. Chem. A* **2006**, *110*, 11371.

43. J. Ulander, A. Broo, *Int. J. Quantum Chem.* **2005**, *105*, 866.

44. K. C. Gross, P. G. Seybold, *Int. J. Quantum Chem.* **2000**, *80*, 1107.

45. K. C. Gross, P. G. Seybold, *Int. J. Quantum Chem.* **2001**, *85*, 569.

46. J. R. Pliego Jr., J. M. Riveros, *Chem. Phys. Lett.* **2000**, *332*, 597.

47. J. R. Pliego Jr., *Chem. Phys. Lett.* **2003**, *367*, 145.

48. K. R. Adam, *J. Phys. Chem. A* **2002**, *106*, 11963.

49. Y. J. Kim, A. Streitwieser, *J. Am. Chem. Soc.* **2002**, *124*, 5757.

50. G. Schuurmann, M. Cossi, V. Barone, J. Tomasi, *J. Phys. Chem. A* **1998**, *102*, 6706.

51. E. Soriano, S. Cerdan, P. Ballesteros, *J. Mol. Struct. (Theochem)* **2004**, *684*, 121.

52. N. Sadlej-Sosnowska, *Theor. Chem. Acc.* **2007**, *118*, 281.

53. K. K. Govender, I. Cukrowski, *J. Phys. Chem. A* **2009**, *113*, 3639.

54. C. J. Cramer, 'Essentials of Computational Chemistry', John Wiley & Sons Ltd, 2002.

55. 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, US Department of Commerce, National Institute of Standards and Technology, 2004.

56. M. S. Grigoriev, C. D. Auwer, D. Meyer, P. Moisy, *Acta Cryst.* **2006**, *C62*, m163.

57. L. C. Yu, S. L. Liu, E. X. Liang, C. L. Wen, *J. Coord. Chem.* **2007**, *60*, 2097.

58. D. Malevich, Z. Wang, P. R. Tremaine, *J. Sol. Chem.* **2006**, *35*, 1303.

59. M. S. Vohra, A. P. Davis, *J. Colloid Interf. Sci.* **1997**, *194*, 59.

60. J. Karhu, L. Harju, A. Ivaska, *Anal. Chimi. Acta* **1999**, *380*, 105.

61. E. R. Souaya, W. G. Hanna, E. H. Ismail, N. E. Milad, *J. Coord. Chem.* **2004**, *57*, 825.

62. M. F. Quartacci, B. Irtelli, A. J. M. Baker, F. Navari-Izzo, *Chemosphere* **2007**, *68*, 1920.

63. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Pettersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V.

Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian 03, Revision D.01, Gaussian, Inc., Wallingford CT, 2004.

64. GaussView 4.1, Gaussian, Inc., 340 Quinnipiac St. Bldg. 40 Wallingford, CT, 06492, USA.

65. P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623.

66. E. Cances, B. Mennucci, J. Tomasi, *J. Chem. Phys.* **1997**, *107*, 3032.

67. M. Cossi, V. Barone, R. Cammi, J. Tomasi, *Chem. Phys. Lett.* **1996**, *255*, 327.

68. S. Miertus, E. Scrocco, J. Tomasi, *Chem. Phys.* **1981**, *55*, 117.

69. V. Barone, M. Cossi, *J. Phys. Chem. A* **1998**, *102*, 1995.

70. M. Cossi, N. Rega, G. Scalmani, V. Barone, *J. Comput. Chem.* **2003**, *24*, 669.

71. F. H. Allen, *Acta Cryst.* **2002**, *B58*, 380.

72. E. Skrzypczak-Jankun, D. A. Smith, H. Maluszynska, *Acta Cryst.* **1994**, *C50*, 1097.

73. M. Kaneyoshi, A. Bond, W. Jones, *Acta Cryst.* **1999**, *C55*, 1260.

74. L. A. Curtiss, K. Raghavachari, *Theor. Chem. Acc.* **2002**, *108*, 61.

75. W. Sang-Aroon, V. Ruangpornvisuti, *J. Mol. Struct. (Theochem)* **2006**, *758*, 181.

76. J. E. Davies, N. L. Doltsinis, A. J. Kirby, C. D. Roussev, M. Sprik, *J. Am. Chem. Soc.* **2002**, *124*, 6594.

77. G. da Silva, J. W. Bozzelli, *J. Phys. Chem. A* **2006**, *110*, 13058.

78. G. da Silva, J. W. Bozzelli, *J. Phys. Chem. C* **2007**, *111*, 5760.

79. Y. Guo, H. Gao, B. Twamley, J. M. Shreeve, *Adv. Mater.* **2007**, *19*, 2884.

80. J. Espinosa-Garcia, J. C. Garcia-Bernaldez, *Phys. Chem. Chem. Phys.* **2002**, *4*, 4096.

81. I. Krossing, I. Raabe, *Chem-Eur. J.* **2004**, *10*, 5017.

82. A. Kutt, V. Movchun, T. Rodima, T. Dansauer, E. B. Rusanov, I. Leito, I. Kaljurand, J. Koppel, V. Pihl, I. Koppel, G. Ovsjannikov, L. Toom, M. Mishima, M. Medebielle, E. Lork, G.-V. Roschenthaler, I. A. Koppel, A. A. Kolomeitsev, *J. Org. Chem.* **2008**, *73*, 2607.

83. E. M. Nolan, R. G. Linck, *J. Am. Chem. Soc.* **2000**, *122*, 11497.

84. D. Peeters, G. Leroy, C. Wilante, *J. Mol. Struct.* **1997**, *416*, 21.

85. L. Wang, D. E. Heard, M. J. Pilling, P. Seakins, *J. Phys. Chem. A* **2008**, *112*, 1832.

86. L. L. C. Schrodinger, Schrodinger Maestro, 32nd Floor, Tower 45, 120 West Forty-Fifth Street, New York, 10036, 2003.

Chapter 4

This is a typed copy of a manuscript with an aim of submitting it for publication.

4. Stability of metal complexes of NTA and NTPA

4.1. Introduction

Many metallic elements play a crucial role in all living systems. An important characteristic of metals is that they easily lose electrons from the elemental or metallic state to form positively charged ions.[1] This means that metal ions are usually electron deficient which results in them having a general tendency to bind to and interact with biological molecules, such as NTA and NTPA, which are electron rich molecules. This same principle applies to the affinity of metal ions for many small molecules and ions crucial to life, such as O_2.[1] There are numerous tasks performed by metals, such as carrying oxygen throughout the body and shuttling electrons.[2] The number of ligands, usually between two and nine, that can bind (coordinate) to a metal ion vary depending on size, charge, and electron configuration of both the metal ion and the ligands. The geometries around the central metal ion can therefore vary from linear for two coordinated ligands, to tri-capped trigonal prismatic, for nine coordinated ligands.

Within this chapter focus is placed on metal-ligand complexes of Zn^{2+} and Ni^{2+} with both NTA and NTPA. Zinc is able to constitute strong, but readily exchangeable and flexible, complexes with organic molecules, thereby enabling it to modify the three-dimensional structure of nucleic acids, specific proteins and cellular membranes and influence catalytic properties of many enzyme systems and intracellular signalling.[3] Another interesting feature of this metal ion is that it can be transported in biological systems without inducing oxidant damage, as can occur with trace elements such as iron and copper.[3] Nickel, on the other hand is both essential and toxic for both humans and animals.[4] In fact, an excess of Nickel induces Zinc deficiency symptoms which are similar to parakeratosis, a symptom found in pigs.[4] The choice of these metals for the purposes of these studies was influenced by this relation.

Since most studies involving complexes of Zn^{2+} and Ni^{2+} with NTA have reported octahedral ML complexes,[5-9] where two water molecules complete the metal coordination sphere, it was decided that these studies will be focused solely on octahedral (six-coordinate) complexes of NTA and NTPA. The complexes, which can formally be written as $[Zn(NTA)(H_2O)_2]^-$, $[Ni(NTA)(H_2O)_2]^-$, $[Zn(NTPA)(H_2O)_2]^-$ and $[Ni(NTPA)(H_2O)_2]^-$, will be referred to as Zn(NTA), Ni(NTA), Zn(NTPA) and Ni(NTPA), respectively, for the remainder

of this chapter. As it has been mentioned in previous chapters, metal complexes of NTA are orders of magnitude stronger than those of NTPA. In the case of Zn(NTA) and Ni(NTA) complexes the stability constants, as log K_1, are 10.45 and 11.51 at ionic strength 0.1 M and 25 ^{o}C, respectively,[10] and those for Zn(NTPA) and Ni(NTPA) are, as log K_1, 5.8 and 5.3 at ionic strength 0.1 M and 30 ^{o}C, respectively.[10] In this chapter a number of aspects is considered with the ultimate goal of explaining why complexes with NTA are orders of magnitude more stable than those of NTPA. The aspects considered were:

a) Theoretical prediction of stability constants for Zn(NTPA) and Ni(NTPA) complexes with the aid of isodesmic reactions. A study was conducted by Hancock et al,[11] in which a prediction of formation constants for complexes of ammonia was done, but the means by which they went about predicting the formation constants is considerably different to the methodology used for these studies and their results were not very accurate implying that the theoretical prediction of the stability constants is not an easy task.

b) Structural comparative analysis of each of the metal-ligand complexes.

c) Full natural bond orbital (NBO)[12] and Bader's charge analysis of specific atoms within the metal-ligand complexes. In particular, focus is placed on the central metal ion and atoms involved in formation of bonds with the metal ion.

d) The analysis based on the quantum theory of atoms in molecules (QTAIM) of Bader[13] was performed. QTAIM provides a great deal of information about the nature of the bonding environment. The topological properties of the electron charge density, (ρ) and the Laplacian ($\nabla^2\rho$) of the charge density at various bond, ring and cage critical points was explored. Previous studies with the aid of the QTAIM have been useful in determining the presence of intermolecular hydrogen bonds,[14] halogen bonding interactions,[15] uncommon hydrogen bonds,[16] amongst others.

e) Single point calculations on selected fragments of complexes studied.

It shall be seen that all of the methods mentioned above, which were the only techniques available at the time, complement one another with respect to explaining why complexes formed with NTA are orders of magnitude more stable than those formed with NTPA. In addition, it was discovered that so-called hydrogen clashes (C–H···H–C close contacts)[17] are present only in the more crowded NTPA complexes. They are used by inorganic chemists to

explain smaller stability of the M-NTPA complexes (were here M = Zn^{2+} or Ni^{2+}, but it applies to all metal ions). However, according to the AIM studies, these are actually so-called H–H stabilizing interactions,[18] without which the M-NTPA complexes most likely would not have formed at all.

4.2. Computational Details

All molecular modelling was performed with the aid of GAUSSIAN 03 software package[19]. The software was installed and ran on Intel Xeon computer systems, possessing 8 processors, 16 GB of RAM and a 64-bit linux based operating system (OpenSuse 10.3). GaussView 4.1[20] was used for all molecular construction and visualization purposes. All molecules, after being self-constructed (referred to as constructed structures from this point onwards), were optimized in solvent (water, ε = 78.39) using the B3LYP/6-311+G(d,p) level of theory in combination with the PCM-UA0,[21-23] PCM-UAHF and CPCM-UAKS[24, 25] solvation models; since all molecules analyzed possessed negative charges it was essential that diffuse functions be included.[26] Gas phase optimization was also tested, but in the case of the Zn^{2+} complexes the structures went from being 6-coordinate octahedral complexes to being 5-coordinate trigonal bipyrimidal structures. Because the focus of these studies is based on analysis of octahedral complexes, gas phase optimization could not be utilized. In the case of Ni^{2+} complexes of NTA and NTPA, all structures had to be optimized using a triplet spin multiplicity, as optimization using a singlet spin multiplicity is used for Ni^{2+} complexes possessing square-planar geometry.[27] Structures of Zn^{2+} and Ni^{2+} complexes of NTA and NTPA were also generated, in solvent, with the aid of Schrödinger Maestro software package[28] (referred to as MM/MD structures from this point onwards). These structures were submitted to Gaussian for energy optimization using the same level of theory and PCM-UA0 solvation model.

The NBO analyses has been applied to assess the transfer of electron density between the electron donor and the acceptor of an idealized Lewis structure into an empty non-Lewis orbital by computing the second order stabilization (perturbation) energies $E^{(2)}$ (caused by charge transfer interactions between various donor-acceptor pairs of orbitals) by natural localized molecular orbitals (NLMO) analysis using the NBO 3.0 set of programs[29] available in GAUSSIAN 03. The charges were generated by performing single point calculations (SPCs) on the solvent optimized structures, i.e. both the constructed and

MM/MD generated structures, using the B3LYP/6-311+G(d,p) level of theory and PCM-UA0 solvation model.

A topological analysis of the electron charge density (ρ(r)) and the Laplacian of the charge density ($\nabla^2\rho$(r)) at all the bond critical points (BCPs), ring critical points (RCPs), and cage critical points (CCPs) using B3LYP level of theory in conjunction with the 6-311+G(d,p) basis set was carried out on the solvent optimized molecular geometry of the complexes using the quantum theory of atoms in molecules of Bader.[13] The resulting wavefunction files (generated in Gaussian) were used as inputs to AIM 2000 software[30] for the calculation and visualization of the topological properties of ρ(r) and $\nabla^2\rho$(r) as well as Bader's charges, which were generated by integrating in natural coordinates rather than just integrating inside the Beta-Sphere.[31]

Whether the optimized structures corresponded to stable minima on the potential energy surface was determined by performing frequency calculations. With all of the structures used during these studies no negative frequencies were encountered.

4.3. Results and Discussion

4.3.1. *Level of Theory*

Since the RB3LYP/6-311+G(d,p) level of theory worked well, as far as the analysis of the ligands was concerned, it was decided that the same level of theory be tested on the Zn^{2+} and Ni^{2+} complexes with NTA and NTPA. Assumption was also made that this level of theory could reproduce the crystallographic data (available from the Cambridge Crystallographic Database (CSD)[32]). Unfortunately, there are no crystal structures in which Zn(NTA), Ni(NTA) or Ni(NTPA) have been crystallized on their own; instead the crystal structures that are available contain these complexes within an intertwined network of other co-crystallized molecules.[33-35] In addition, one of the crystal structures (Ni(NTA)) has a counter-ion present in the molecule (Figure C1 in Appendix C depicts the crystallographic structures for Ni(NTA) and Ni(NTPA)). Because this was the only data available, it had to be used when comparing the experimental (crystal structure) and theoretically determined metal complex structures. Figure 4.1 provides the theoretical structure for the Zn(NTA) complex with the atom labelling that was used for comparison purposes; only those atoms that were compared with the crystallographic structures are labelled (Labelled structures of Ni(NTA), Zn(NTPA) and Ni(NTPA) are provided in Figure C2 of Appendix C).

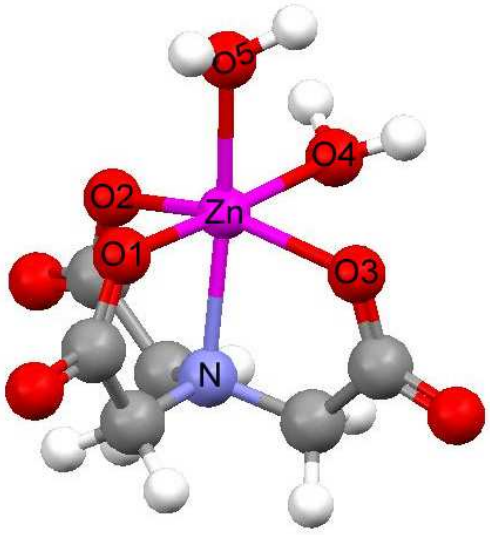


Figure 4.1: Partially labelled, solvent optimized, structure of Zn(NTA) metal complex

The structural matrix shown in Table 4.1 provides a comparison of the experimental and theoretical (solvent optimized) structures of Zn(NTA) (Structural matrices for Ni(NTA) and Ni(NTPA) are provided in Tables C1 and C2 of Appendix C, no experimental data was available for Zn(NTPA)). From the results produced in Table 4.1 it was found that the metal-ligand bond lengths (including the bonds to the water molecules) were reproduced to within –0.030 ± 0.1148 Å, and the *cis* and *trans* angles were reproduced to within –0.39 ± 5.63 and –3.43 ± 4.19, respectively. Figure 4.2 provides the unit cell for the crystal structure of Zn(NTA). In this figure it can be seen that a single COO group bonds to two Zinc atoms (illustrated by a circle), there are also oxygen bonds to Zinc from a neighbouring NTA arm (illustrated by arrow A) and there are oxygen bonds to both sodium and zinc (illustrated by arrow B) which also belong to a neighbouring NTA arm. This same sort of pattern is observed for all of the molecules shown in the unit cell. The oxygen atoms illustrated with arrows A and B belong to coordinated water molecules when looking at the optimized Zn(NTA) complex (Figure 4.1), but in the crystal structure this is not the case. The optimized Zn(NTA) complex also does not have any additional bonds present on the NTA arms as is the case with the crystal structure. It is clear that the crystal structure differs significantly from the optimized Zn(NTA) structure and it is this large deviation that results in the large differences seen in Table 4.1. This indicates that comparison with the available crystal structure is not going to provide the information necessary to determine if the RB3LYP/6-311+G(d,p) level of theory is sufficient. Since the RB3LYP/6-311+G(d,p) level of theory

was sufficient for ligand related calculations (as discussed in previous chapters), it was decided to use it also for the purpose of these studies.

Table 4.1: Comparison of experimental (CSD) and computed at the RB3LYP/6-311+G(d,p) level of theory and PCM-UA0 solvation model selected bond distances and angles for the Zn(NTA) complex. Bond lengths and angles are in Å and °, respectively.

Atoms	CSD data[b]	Input structure	Difference (Δ)[a]
Zn-N	2.147	2.219	–0.072
Zn-O1	2.071	2.075	–0.004
Zn-O2	2.124	2.081	0.043
Zn-O3	2.188	2.093	0.095
Zn-O4	2.001	2.236	–0.235
Zn-O5	2.128	2.135	–0.007
N-Zn-O5	163.7	171.6	–7.9
O1-Zn-O4	174.2	177.0	–2.8
O2-Zn-O3	155.9	155.5	0.4
N-Zn-O1	94.7	100.1	–5.4
N-Zn-O2	78.3	80.3	–2.0
N-Zn-O3	78.2	80.1	–1.9
N-Zn-O4	83.1	82.9	0.2
O1-Zn-O2	83.9	84.5	–0.6
O1-Zn-O3	88.9	88.8	0.1
O1-Zn-O5	90.5	95.9	–5.4
O2-Zn-O4	91.8	96.2	–4.4
O2-Zn-O5	87.8	84.6	3.2
O3-Zn-O4	93.0	99.4	–6.4
O3-Zn-O5	92.0	88.2	3.8
O4-Zn-O5	116.2	102.1	14.1

[a]) Δ = (experimental – computed) value
[b]) Average bond lengths and angles of crystal structures obtained from the CSD[32]

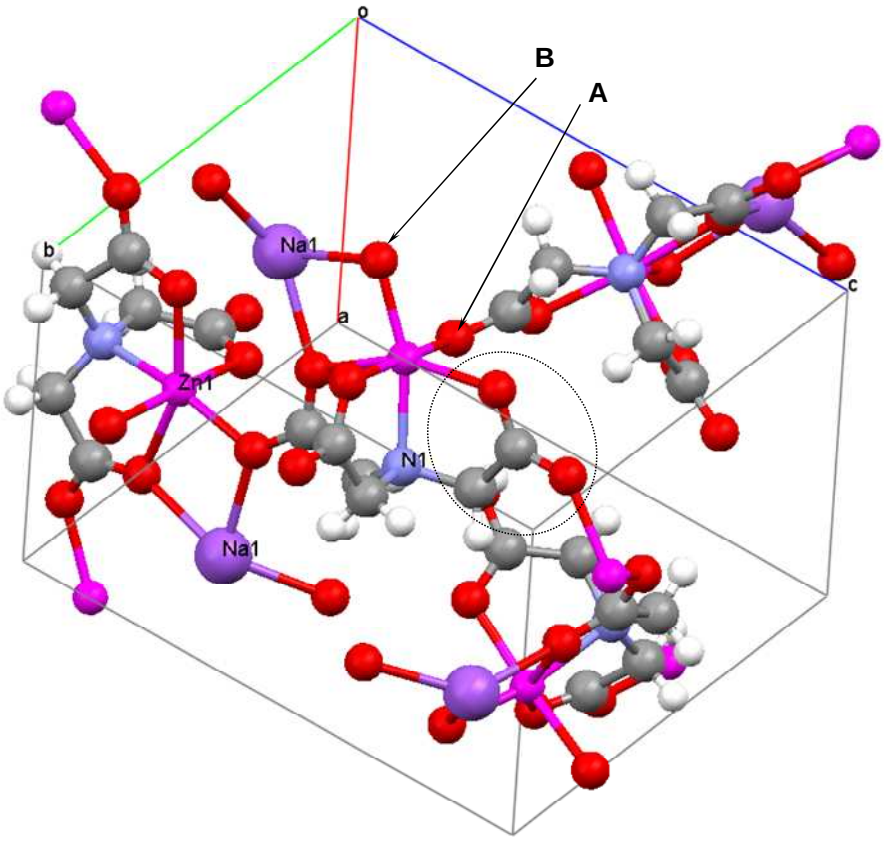


Figure 4.2: Crystallographic structure[32] of Zn(NTA) (molecules within a unit cell).

4.3.2. *Stability constants of Zn(II) and Ni(II) with NTPA from Isodesmic Reaction*

Since the use of isodesmic reactions (IRns) was so successful in the prediction of protonation constants for NTA and NTPA, this methodology was also applied to the prediction of stability constants for both Zn(NTPA) and Ni(NTPA) complexes. An isodesmic reaction that represents competing of two ligands ($L_{(1)}$ and $L_{(2)}$ that represent the studied and reference ligand, respectively) for the same metal ion can be written as

$$L_{(1)}^{3-}(aq) + [ML_{(2)}(H_2O)_2]^-(aq) = [ML_{(1)}(H_2O)_2]^-(aq) + L_{(2)}^{3-}(aq) \quad \Delta G_{aq} \qquad (1)$$

Each ligand (NTPA = $L_{(1)}$ and NTA = $L_{(2)}$) is involved in complex formation reactions, which can be written as

$$[M(H_2O)_6]^{2+} + L_{(1)}^{3-} = [ML_{(1)}(H_2O)_2]^- + 4H_2O \qquad \Delta G_1(aq) \qquad (2)$$

$$[M(H_2O)_6]^{2+} + L_{(2)}^{3-} = [ML_{(2)}(H_2O)_2]^- + 4H_2O \qquad \Delta G_2(aq) \qquad (3)$$

and the change in Gibbs energies for each complex formation reaction (eq 2 and 3) can be written as

$$\Delta G_1(\text{aq}) = G_{\text{aq}}([\text{ML}_{(1)}(\text{H}_2\text{O})_2]^-) + 4\times G_{\text{aq}}(\text{H}_2\text{O}) - G_{\text{aq}}([\text{M}(\text{H}_2\text{O})_6]^{2+}) - G_{\text{aq}}(\text{L}_{(1)}{}^{3-}) \quad (4)$$

$$\Delta G_2(\text{aq}) = G_{\text{aq}}([\text{ML}_{(2)}(\text{H}_2\text{O})_2]^-) + 4\times G_{\text{aq}}(\text{H}_2\text{O}) - G_{\text{aq}}([\text{M}(\text{H}_2\text{O})_6]^{2+}) - G_{\text{aq}}(\text{L}_{(2)}{}^{3-}) \quad (5)$$

The isodesmic reaction of interest (eq 1) can be obtained by subtracting eq 3 from 2, and hence from subtracting eq 5 from 4 one obtains the expression for the Gibbs energy, $\Delta G_{\text{aq}} = \Delta G_1(\text{aq}) - \Delta G_2(\text{aq})$, were the term $G_{\text{aq}}[\text{M}(\text{H}_2\text{O})_6]^{2+}$ and $G_{\text{aq}}(\text{H}_2\text{O})$ now cancel out. This results in a final expression for this isodesmic reaction, which is written as

$$\begin{aligned}\Delta G_{\text{aq}} = \Delta G_1(\text{aq}) - \Delta G_2(\text{aq}) = \\ G_{\text{aq}}([\text{ML}_{(1)}(\text{H}_2\text{O})_2]^-) - G_{\text{aq}}(\text{L}_{(1)}{}^{3-}) - \\ G_{\text{aq}}([\text{ML}_{(2)}(\text{H}_2\text{O})_2]^-) + G_{\text{aq}}(\text{L}_{(2)}{}^{3-})\end{aligned} \quad (6)$$

Eq 6 was used to calculate ΔG_{aq} of the isodesmic reaction (eq 1) from appropriate Gibbs energies for the relevant and fully solvent-optimized structures of all the molecules involved in the reaction. Table 4.2 lists the ZPVE-corrected minimum energies, E_{min}, as well as the Gibbs free energies for NTA, NTPA and the Zn^{2+} and Ni^{2+} metal complexes with these ligands, which were obtained by making use of a number of different solvation models. The value of $\Delta G_2(\text{aq})$ was calculated from eq 7 using the reported stability constants (at 25 $^{\text{o}}$C and μ = 0.1 M) of the $\text{ML}_{(2)}(\text{H}_2\text{O})_2$ complexes.

$$\Delta G(\text{aq}) = -\text{RT} \ln K_{\text{ML}} \quad (7)$$

Eq 7 can also be written as

$$\Delta G(\text{aq}) = -2.303\text{RT} \log K_1 \quad (8)$$

Once the values for ΔG_{aq} and $\Delta G_2(\text{aq})$ have been calculated one can obtain $\Delta G_1(\text{aq})$ by rearranging eq 6 and this in turn can be used to calculate the stability constants for the $\text{ML}_{(1)}(\text{H}_2\text{O})_2$ complexes.

Table 4.3 provides the values of the quantities required to calculate stability constants, calculated and experimental stability constants of both Zn(NTPA) and Ni(NTPA), as well as the differences between calculated and experimental stability constants (δ). The values used for ΔG_2(aq), which are not presented in Table 4.3, were –14.258 and –15.704 kcal/mol for the Zn(NTA) and Ni(NTA) metal complexes, respectively. These values were obtained from experimental data[10] with log K_1 = 10.45 and 11.51 at ionic strength 0.1 M and 25 ^{o}C for the Zn(NTA) and Ni(NTA) metal complexes, respectively. It is important to mention that all of the information provided in Tables 4.2 and 4.3 were generated from constructed structures. The results produced from the MM/MD structures are provided in Tables C3 and C4 of Appendix C. From the results shown in Table 4.3 it is clear that by increasing the solvation model from PCM-UA0 to CPCM-UAKS the calculated stability constant improved by very little. Because of this it was concluded that computing the MM/MD structures using larger solvation models is merely going to consume a lot of time without providing results that are significantly more accurate when compared with PCM-UA0.

Unfortunately, due to time constraints only one reference molecule (ligand $L_{(2)}$) was tested during the application of the IRns and in light of the fact that these studies were focused on NTA and NTPA, the ligand chosen was NTA. The results obtained in Table 4.3 show that prediction of stability constants with IRns is not very accurate, at least not as accurate as values predicted for protonation constants (as discussed in previous chapters). For each predicted stability constant there is significant underestimation indicating that, according to theory, complexes formed with NTPA are much weaker than those which have been determined experimentally. This underestimation (similar or even larger differences between computed and experimental values are commonly reported also for protonation constants[43, 44] that do not involve a metal ion in computations) is probably due to the use of conformers that were not totally correct, as they exist in real solvent. Involving a full conformation analysis possibly might produce results that are more accurate. However, the predicted values do follow the trend set by experiment, in that both the Zn^{2+} and Ni^{2+} metal complexes with NTPA are far less stable than the complexes formed with NTA. As was expected, values predicted with PCM-UA0 solvation model are not as accurate as those predicted with CPCM-UAKS.[26, 36-40] Interestingly, it was noticed that the PCM-UAHF solvation model predicted stability constant that was worse than that of PCM-UA0 in the case of Zn(NTPA), but it predicted a value that was better in the case of Ni(NTPA).

Table 4.2: ZPVE-corrected minimum and Gibbs free energies of NTA, NTPA, Zn(NTA), Zn(NTPA), Ni(NTA) and Ni(NTPA) obtained in solvent (H_2O) at the RB3LYP/6-311+G(d,p) level of theory with different solvation models.

	Solvent optimized constructed structures					
	PCM-UA0		**PCM-UAHF**		**CPCM-UAKS**	
Species	E_{min}	G_{aq}	E_{min}	G_{aq}	E_{min}	G_{aq}
NTA^{3-}	–738.936972	–738.978069	–738.962253	–739.002625	–738.962832	–739.003179
$NTPA^{3-}$	–856.828229	–856.877193	–856.849447	–856.897237	–856.849754	–856.897797
Zn(NTA)	–2671.017506	–2671.065930	–2671.017601	–2671.066428	–2671.022898	–2671.072144
Zn(NTPA)	–2788.895561	–2788.948938	–2788.892233	–2788.944754	–2788.897482	–2788.951271
Ni(NTA)	–2399.984155	–2400.031805	–2399.984352	–2400.032423	–2399.989762	–2400.037883
Ni(NTPA)	–2517.860414	–2517.911881	–2517.856183	–2517.908016	–2517.862129	–2517.913905

All energies are reported in atomic units, Hartree (1 Hartree = 627.5095 kcal/mol)

Table 4.3: Comparison of experimental[5] (Exp) and calculated stability constants, as log *K*, of Zn(NTPA) and Ni(NTPA) metal complexes.

		PCM-UA0				**PCM-UAHF**				**CPCM-UAKS**			
Reaction	**Exp**	ΔG_{aq}	ΔG_1(aq)	log K_1	Δ	ΔG_{aq}	ΔG_1(aq)	log K_1	δ	ΔG_{aq}	ΔG_1(aq)	log K_1	δ
$[ZnL_{(2)}(H_2O)_2]^- + L_{(1)}^{3-} = [ZnL_{(1)}(H_2O)_2]^- + L_{(2)}^{3-}$	5.3	9.117	–5.141	3.04	2.26	10.220	–4.039	2.96	2.34	9.721	–4.537	3.33	1.97
$[NiL_{(2)}(H_2O)_2]^- + L_{(1)}^{3-} = [NiL_{(1)}(H_2O)_2]^- + L_{(2)}^{3-}$	5.8	10.957	–4.747	2.75	3.05	11.935	–3.770	2.76	3.04	11.669	–4.035	2.96	2.84

All energies are reported in kcal/mol

The difference in predicted stability constants between these two solvation models was considerably small, which is a good indication that these two models don't differ that much from one another as far as optimization is concerned, but this is something which possibly only occurs for the metal complexes that are studied here. In order to confirm if this is a pattern that continues, both the PCM-UA0 and PCM-UAHF solvation models would have to be applied to other metal complexes.

4.3.3. *Structural analysis*

Since the trend in stability of metal complexes with NTA and NTPA has been predicted correctly, attention was focused now on determining why the metal complexes of NTA are orders of magnitude stronger than complexes formed with NTPA. The first step was to compare the optimized structures of metal complexes with each other, in order to identify any unusual or complex-specific structural features. Since solvent optimization was not only performed on constructed structures, but also on MM/MD structures, it was decided that a comparison of all of these structures should be done in order to locate any significant structural differences. Table 4.4 provides such a comparison for bond lengths between the central metal ion and a donor atom of a ligand, along with the difference (Δ) between the NTA and NTPA metal complexes, as well the respective bite angles. The MM/MD structures (generated in solvent with the aid of Schrödinger Maestro) were fully optimized with the aid of Gaussian at the RB3LYP/6-311+G(d,p) level of theory in conjunction with the PCM/UA0 solvation model. The partially labelled constructed and MM/MD structures of the metal complexes are shown in Figures C2 and C3 of Appendix C (the constructed structure of Zn(NTA) can be obtained from Figure 4.1). The data in Table 4.4 reveals that there are no large differences between the metal to ligand bond lengths of Zn(NTA) and Zn(NTPA), nor are there large differences between the bond lengths of Ni(NTA) and Ni(NTPA). It is expected that the longer a metal-ligand bond the weaker that complex would be[41] since the transfer of electrons from the metal to the ligand will be reduced. By taking the average bond lengths (with the average being for both Zn-ligand and Zn-H_2O bonds) of Zn(NTA) constructed structure one obtains 2.140 Å and for the Zn(NTPA) constructed structure the average is 2.170 Å. This indicates that the Zn(NTPA) constructed structure has an average bond length that is 0.030 Å longer than the Zn(NTA) constructed structure. Similarly the Zn(NTPA) MM/MD structure was found to have an average bond length that was 0.015 Å

longer than the Zn(NTA) MM/MD structure. In the case of the Ni(NTA) and Ni(NTPA) constructed structures it was found that Ni(NTPA) possessed an average bond length that was 0.022 Å longer than Ni(NTA). Finally for the Ni(NTA) and Ni(NTPA) MM/MD structures it was discovered that Ni(NTPA) had an average bond length that was 0.013 Å longer than Ni(NTA). This result therefore confirms, theoretically, that the complexes formed with NTA are stronger, and hence more stable, than those with NTPA.

Table 4.4: Comparison of selected bond lengths (in Å) and bite angles (in °) of solvent optimized, at the RB3LYP/6-311+G(d,p) level in combination with the PCM-UA0 solvation model, Zn^{2+} and Ni^{2+} complexes of NTA and NTPA, respectively.

	Constructed structures			MM/MD structures		
Atoms	**NTA**	**NTPA**	**Δ[a]**	**NTA**	**NTPA**	**Δ[a]**
			M = Zn^{2+}			
Zn–N	2.219	2.185	0.034	2.215	2.191	0.024
Zn–O1	2.075	2.051	0.024	2.082	2.067	0.015
Zn–O2	2.081	2.063	0.018	2.080	2.071	0.009
Zn–O3	2.093	2.043	0.050	2.078	2.060	0.018
Zn–O4*	2.236	2.456	–0.220	2.262	2.326	–0.064
Zn–O5*	2.135	2.221	–0.086	2.135	2.224	–0.089
N-Zn-O1	82.9	92.6	–9.7	79.9	95.5	–15.6
N-Zn-O2	80.2	94.3	–14.1	81.0	93.6	–12.6
N-Zn-O3	80.1	94.1	–14.0	82.7	92.6	–9.9
			M = Ni^{2+}			
Ni–N	2.103	2.128	–0.025	2.097	2.121	–0.024
Ni–O1	2.077	2.039	0.038	2.069	2.039	0.030
Ni–O2	2.028	2.038	–0.010	2.035	2.048	–0.013
Ni–O3	2.074	2.042	0.032	2.068	2.047	0.021
Ni–O4*	2.116	2.256	–0.140	2.120	2.201	–0.081
Ni–O5*	2.157	2.183	–0.026	2.167	2.179	–0.012
N-Ni-O1	81.6	92.3	–10.7	82.3	97.3	–15.0
N-Ni-O2	85.2	93.9	–8.7	85.0	93.5	–8.5
N-Ni-O3	82.4	94.3	–11.9	82.7	93.2	–10.5

[a]) Δ = (NTA – NTPA), * = Oxygen atoms from water molecules.

It is important at this point to mention that since the difference between the bond lengths of the complexes is so minute, it was decided that this was an aspect that most likely was not the only factor influencing the stability of the molecules. Another aspect that should be taken into account is the M–O_{H_2O} bond lengths (O4* and O5*) in Zn^{2+} and Ni^{2+} complexes. It was found that the bond lengths with water molecules present in the Ni^{2+} complexes (for both NTA and NTPA) are shorter than those present in the Zn^{2+} complexes by approximately 0.08 Å. This shows that due to the water molecules being closer in the Ni^{2+} complexes there

should be a larger electron transfer occurring on the Ni^{2+} atom as compared to the Zn^{2+} and hence there should be a smaller charge present on Ni^{2+} as compared to Zn^{2+}. This is an aspect that shall be addressed in the section that follows.

It has previously been determined that the bite angle of an ideal five membered ring is 69^{o} and that of a six membered ring is 109.5^{o}.[17] From Table 4.4 it is clear that the bite angles of the five membered Zn(NTA) and Ni(NTA) complexes are similar but considerably larger than 69^{o} and the bite angles of the six membered Zn(NTPA) and Ni(NTPA) complexes are also similar in values but considerably smaller than 109.5^{o}. Clearly, the rule of preferential bite angle does not hold in this case.

4.3.4. *NBO analysis*

The structural analysis has not provided sufficient information as to why NTA metal complexes are more stable than NTPA complexes. However, as it was mentioned previously, this is definitely not the only factor influencing the stability of the molecules. It is for this reason that focus was then turned to performing NBO analysis. Table C5 of Appendix C provides a list of all the NBO and Bader charges for the different ML complexes and the labelling used within Table C5 is provided in Figures C2 and C3. The charges revealed that, as was suggested in the previous section, charges present on Ni^{2+} were smaller than those present on Zn^{2+}. Unfortunately, a comparison of selected charges in M(NTA) and M(NTPA) complexes did not provide any sought after information as the charges present in the different complexes were almost identical; in order to provide an explanation in terms of charges, the charges need to be considerably different from one another. In other words, charges on the central metal ion are not the cause of the experimentally observed differences in stability constants – there must be then another reason.

Since the charges on different atoms have not provided required information it was decided that an analysis of the second order perturbation energy ($E^{(2)}$) might prove to be a more useful option. Table 4.5 provides the total $E^{(2)}$ values (due to electron transfer from different lone pairs on oxygen atoms or on nitrogen atom to the anti-bonding orbital of Zn^{2+} or Ni^{2+}) obtained for all of the metal complexes considered during this study (the atom labeling utilized can be obtained from Figure 4.1 as well as Figures C2 and C3).

Table 4.5: Total second order perturbation energy ($E^{(2)}$) of Zn(NTA), Zn(NTPA), Ni(NTA) and Ni(NTPA) complexes.

	Constructed structure	**MM/MD structure**
Bond	**Total $E^{(2)}$**	**Total $E^{(2)}$**
	Zn(NTA)	
LP(N) → LP*(Zn)	22.63	22.60
LP(O1) → LP*(Zn)	35.68	35.73
LP(O2) → LP*(Zn)	36.63	36.65
LP(O3) → LP*(Zn)	36.06	36.93
LP(O4$^{\#}$) → LP*(Zn)	22.25	21.33
LP(O5$^{\#}$) → LP*(Zn)	26.62	27.13
	Zn(NTPA)	
LP(N) → LP*(Zn)	27.31	27.05
LP(O1) → LP*(Zn)	38.78	37.09
LP(O2) → LP*(Zn)	35.65	36.36
LP(O3) → LP*(Zn)	39.29	38.40
LP(O4$^{\#}$) → LP*(Zn)	14.45	18.01
LP(O5$^{\#}$) → LP*(Zn)	21.84	21.94
	Ni(NTA)	
LP(N) → LP*(Ni)	43.18	43.80
LP(O1) → LP*(Ni)	47.17	48.68
LP(O2) → LP*(Ni)	50.62	51.00
LP(O3) → LP*(Ni)	48.12	49.17
LP(O4$^{\#}$) → LP*(Ni)	39.78	40.22
LP(O5$^{\#}$) → LP*(Ni)	33.12	32.58
	Ni(NTPA)	
LP(N) → LP*(Ni)	10.29	10.67
LP(O1) → LP*(Ni)	46.38	46.30
LP(O2) → LP*(Ni)	16.26	16.68
LP(O3) → LP*(Ni)	46.73	46.34
LP(O4$^{\#}$) → LP*(Ni)	12.31	13.04
LP(O5$^{\#}$) → LP*(Ni)	13.55	13.43

All energies are reported in kcal/mol, $^{\#}$ = Oxygen atoms from water molecules.

An excellent correlation exists between $E^{(2)}$ and the bond length in the case of Zn complexes with NTA and NTPA – see Figure 4.3. The larger the $E^{(2)}$ value the shorter the bond. This applies to self-constructed as well as MM/MD-generated conformers indicating that this indeed might be a true physical property of the complexes. Figure 4.3 also shows that Zn–O_L bonds (were O_L represents the oxygen atoms of the ligand) of Zn(NTA) have $E^{(2)}$ values that are grouped together, whereas the same bonds for Zn(NTPA) are much more dispersed. In the case of the Zn–O_{H_2O} bonds for Zn(NTA) it is seen that the $E^{(2)}$ values are a bit more dispersed than those obtained for the Zn–O_L bonds, but the dispersion is not as large as that obtained for the Zn–O_{H_2O} bonds of Zn(NTPA). Clearly the $E^{(2)}$ values for Zn(NTPA)

are much more spread out than those of Zn(NTA) and it is possible that this dispersion occurs due to the difference in stability of the two complexes. But in order to confirm this one would have to perform NBO analysis on a number of different Zn complexes.

An interesting observation is related to the LP(N) → LP*(Zn) in case of NTPA – it is seen that the significantly larger $E^{(2)}$ in case of NTPA has not resulted in expected shortening of the Zn–N bonds. The question one would like to ask is why this is so. The simplified and possible reasonable explanation could be the fact of steric crowding present in the NTPA ligand when complexed to Zn. Clearly, the nitrogen atom cannot come closer to Zn as it would require further bending of the ligand-arms and significant clashes between numerous atoms in the ligand. It appears that to form this bond (the stabilizing contribution to the overall energy of the complex) significantly larger electron transfer had to take place in case of NTPA – without that contribution most likely this bond would be too weak to keep the metal ion and the ligand together. If the above supposition (or explanation) is correct then, as far as it is known, this is the first type of explanation involving the $E^{(2)}$ versus bond length relationship that was used to predict steric crowding within the molecule.

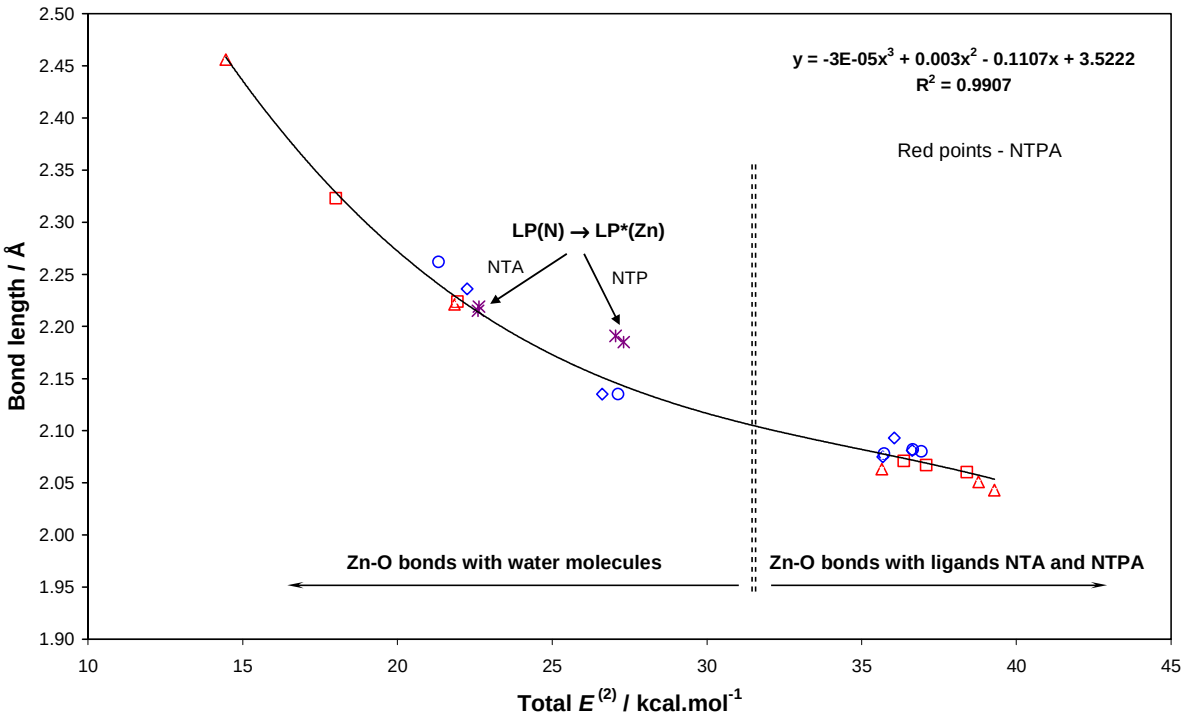
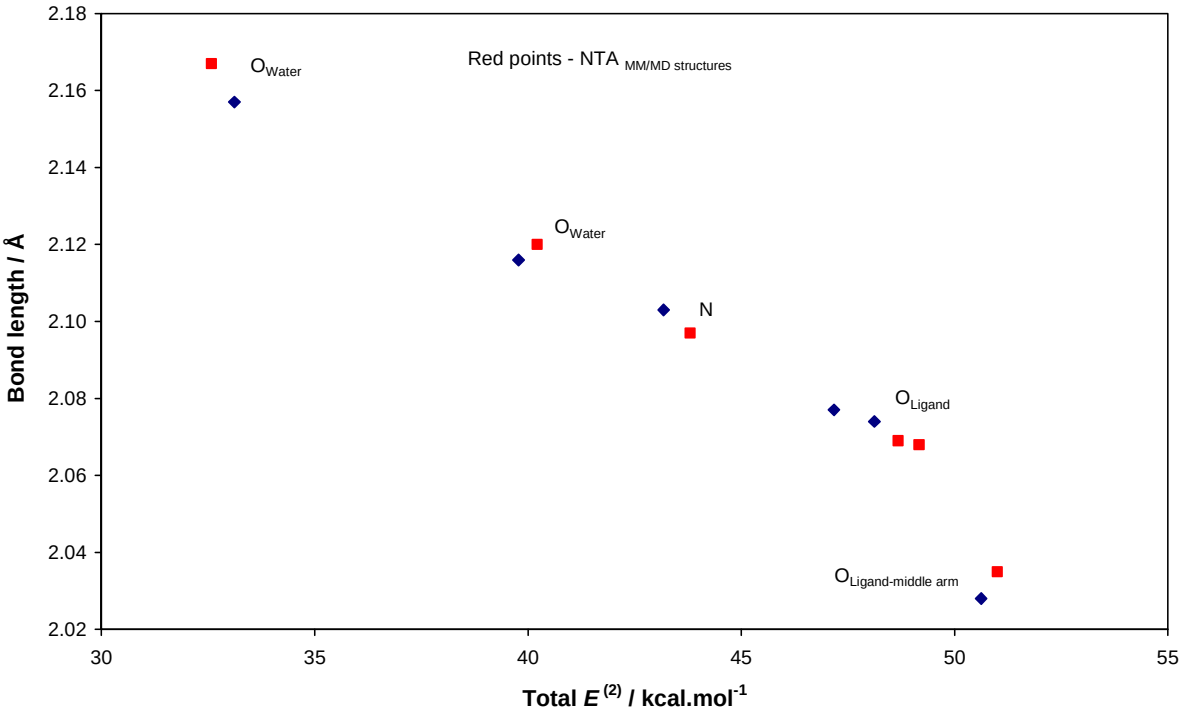


Figure 4.3: Total $E^{(2)}$ versus bond length for both Zn(NTA) and Zn(NTPA).

(a)



(b)

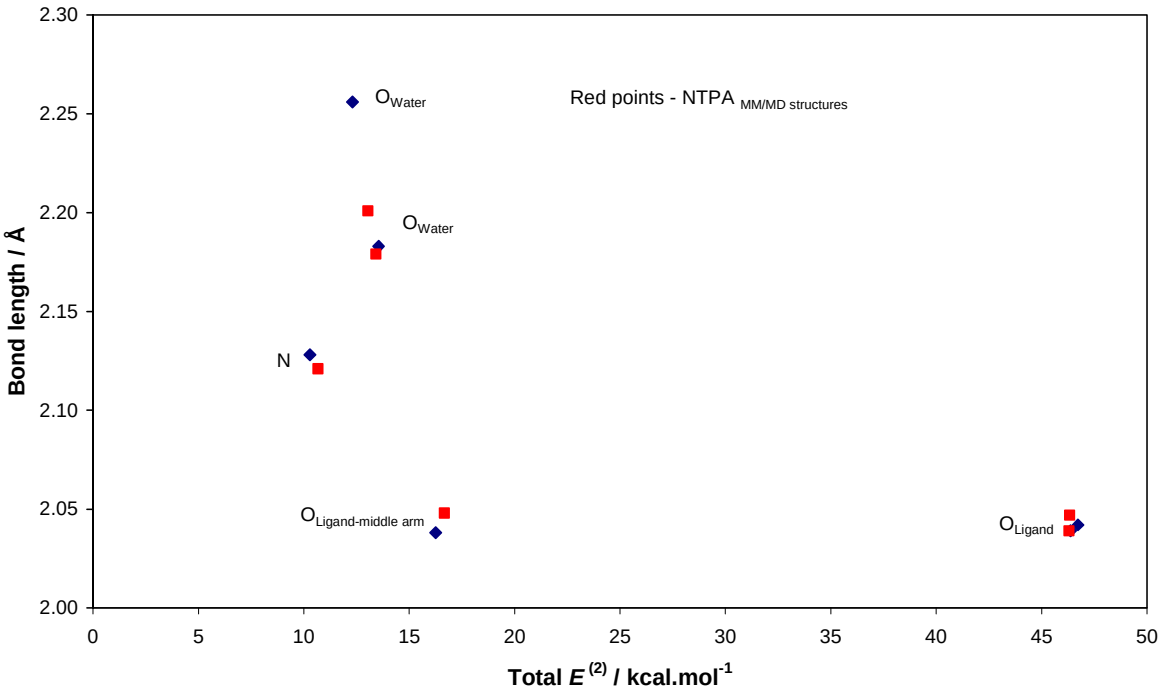


Figure 4.4: Total $E^{(2)}$ versus bond length for (a) Ni(NTA) and (b) Ni(NTPA).

Even though NBO analysis (involving $E^{(2)}$) has not provided a direct explanation of weaknesses of the Zn(NTPA) complexes, the departure of $E^{(2)}$ from the observed trend (longer bond lengths than predicted from relationship seen in Figure 4.3) most likely can be used as a predictive tool of steric hindrance in a molecule as well as a resultant weakness of complexes formed. In the case of the Ni^{2+} complexes it was found that NBO predicts transfers of $E^{(2)}$ to three antibonding orbital LP*: LP*(4), LP*(5) and LP*(6), but with Zn^{2+} only LP*(6) is predicted. In order to ensure a constant comparison between Ni^{2+} and Zn^{2+} only the LP*(6) values of Ni^{2+} are shown in Table 4.5. In the case of Ni(NTA), Table 4.5 shows a trend in $E^{(2)}$ values with LP(O_{H_2O}) → LP*(Ni) < LP(N) → LP*(Ni) < LP(O_L) → LP*(Ni) which is exactly the same trend as that observed in both Zn(NTA) and Zn(NTPA). Figure 4.4(a) provides the $E^{(2)}$ versus bond length of Ni(NTA) (were the red symbols represent the MM/MD structures and the blue symbols represent the constructed structures). This figure shows that the total $E^{(2)}$ follows the general trend, except for the middle arm of Ni(NTA) were the bond is “too short” or $E^{(2)}$ is “too small”. This “outlier” is the only obvious difference between the Zn(NTA) and Ni(NTA) metal complexes. A tentative explanation for this would be that this outlier is the result of the difference in the electron configuration of Ni as compared to Zn. Figure 4.4(b) provides the $E^{(2)}$ versus bond length of Ni(NTPA) and in this case both the nitrogen and the middle arm of Ni(NTPA) are outliers. At this point the significance and physical meaning of why this is the case is not fully understood and, unfortunately, no explanation can be made unless more Ni complexes are studied. What is interesting though is that regardless of the conformer used (constructed or MM/MD), the same trend is observed (Figure 4.4(b)).

4.3.5 *SPC on selected fragments of complexes*

There are two distinctive fragments in the complexes discussed, namely $M(H_2O)_2$ and the ligand that, when combined, forms the complex. The reaction can be written as

$$M(H_2O)_2 + L = ML(H_2O)_2 \qquad (9)$$

Single point frequency calculations in solvent on all of the molecular fragments using the RB3LYP/6-311+G(d,p) level of theory and PCM-UA0 solvation model was performed. This was done by taking the fully energy optimized metal complexes (both constructed and

MM/MD structures) and removing the metal and two waters, leaving behind just the ligand on which single point frequency calculation was performed.

Table 4.6: ZPVE-corrected Gibbs free energies of SPCs performed on molecular fragment.

	Constructed structure		MM/MD structure	
Molecule	E_{SPC}	G_{SPC}	E_{SPC}	G_{SPC}
NTA (from Zn(NTA))	–738.902280	–738.940058	–738.902404	–738.940123
NTA (from Ni(NTA))	–738.899486	–738.936693	–738.898007	–738.934238
NTPA (from Zn(NTPA))	–856.768706	–856.810282	–856.769478	–856.809358
NTPA (from Ni(NTPA))	–856.764484	–856.805992	–856.764193	–856.805777
$Zn(H_2O)_2^a$ (from Zn(NTA))	–1931.993755	–1932.023050	–1931.993220	–1932.022520
$Zn(H_2O)_2^a$ (from Zn(NTPA))	–1931.986784	–1932.015539	–1931.989675	–1932.021591
$Zn(H_2O)_2^b$ (from $Zn(H_2O)_6$)	–1931.997522	–1932.026598	–1931.997522	–1932.026598
$Ni(H_2O)_2^a$ (from Ni(NTA))	–1660.924162	–1660.953859	–1660.922446	–1660.951986
$Ni(H_2O)_2^a$ (from Ni(NTPA))	–1660.870761	–1660.901625	–1660.873602	–1660.903533
$Ni(H_2O)_2^b$ (from $Ni(H_2O)_6$)	–1660.927783	–1660.957362	–1660.927783	–1660.957362

All energies are reported in atomic units, Hartree (1 Hartree = 627.5095 kcal/mol)

A similar procedure was performed on the $M(H_2O)_2$ molecular fragment where the ligand was removed from the optimized metal complex before performing the SPC (this fragment is represented further as $M(H_2O)_2^a$). All results generated from calculations on molecular fragments are listed in Table 4.6. It must be pointed out that the energies of $M(H_2O)_2$ molecular fragment were calculated on the molecular fragments coming from either the complexes discussed in this chapter, or from the $M(H_2O)_6$ complex (this fragment is further represented as $M(H_2O)_2^b$). Figure C4 provides fully optimized structures of $Zn(H_2O)_6$ and $Ni(H_2O)_6$. In order to calculate the energies for the reaction provided in eq 9 the following was done

$$\Delta E = E_{min}(ML(H_2O)_2) - E_{SPC}(M(H_2O)_2) - E_{SPC}(L) \quad (10)$$

$$\Delta G_{aq} = G_{aq}(ML(H_2O)_2) - G_{SPC}(M(H_2O)_2) - G_{SPC}(L) \quad (11)$$

where E represents either the minimum energy, E_{min}, (when full optimisation was performed) or energy obtained from the SPC, E_{SPC}, and ΔG_{aq} is the Gibbs free energy change for the reaction (eq 9). The E_{min} and G_{aq} values for all of the $ML(H_2O)_2$ complexes can be obtained from Table 4.2 and Table C3. The results obtained after application of eq 10 and 11 are provided in Table 4.7. By looking at the values produced in Table 4.7 it is seen that larger

change in energy took place in case of the NTPA ligand. However, the complexes of NTA and NTPA are not identical. A Zn(NTA) complex differs from a Zn(NTPA) complex because the two ligands are not the same (NTPA contains more atoms) and the same applies to Ni(NTA) and Ni(NTPA), and hence a comparison of their energies is not viable. However, these results do provide enough information to say that formation of both Zn^{2+} and Ni^{2+} complexes of NTA and NTPA are thermodynamically favoured, due to the fact that all of the values produced in Table 4.7 are negative.

Table 4.7: Minimum and Gibbs free energies obtained for different complex formation reactions.

	Constructed structure		**MM/MD structure**	
Reaction	**Δ*E***	**Δ*G***	**Δ*E***	**Δ*G***
$Zn(H_2O)_2 + NTA^{3-} = Zn(NTA)$	–80.296	–64.522	–80.197	–66.689
$Zn(H_2O)_2 + NTPA^{3-} = Zn(NTPA)$	–92.423	–77.257	–90.322	–73.481
$Ni(H_2O)_2 + NTA^{3-} = Ni(NTA)$	–105.300	–88.638	–106.372	–92.741
$Ni(H_2O)_2 + NTPA^{3} = Ni(NTPA)$	–146.689	–128.178	–146.487	–128.782

All energies are reported in kcal/mol

In order to understand (or explain) why complexes of NTA are more stable than those of NTPA, a different approach has been adopted involving the SPC method. Instead of applying all of the SPC energies to the above reaction (eq 9) it was decided that they should be compared to energies that have been obtained from fully optimized molecules. The energies of the ligands provided in Table 4.6 were then compared to energies of their corresponding optimized molecules which can be obtained from Table 4.2 and Table C3. In the case of the $M(H_2O)_2$ fragment, the SPC energy obtained from the molecular fragment of the metal complex was compared to the SPC energy obtained from the $M(H_2O)_6$ complex ($M(H_2O)_2^{a}$ was compared to $M(H_2O)_2^{b}$). By comparing all of the SPC energies with those of the optimized molecules it was possible to obtain estimates in strain energies, which is the energy required for a molecule (or fragment) to go from its fully energy relaxed state to the state (structural configuration) it acquires when complexed to a metal ion. Table 4.8 provides the results for the aforementioned. It is clear from Table 4.8 that in order for a ligand (being NTA or NTPA) to reach the structural arrangement observed in a metal complex additional (strain) energy is required. The results obtained reveal that the strain energy for NTA is much smaller than that for NTPA and this correlates with the strength of metal complexes with these two ligands.

Table 4.8: Strain energies, calculated with both minimum energy (ΔE) and Gibbs free energy (ΔG_s), represented as the difference between complexed and fully energy optimized molecular structures.

	Strain Energy			
	Constructed structure		MM/MD structure	
Difference	ΔE	ΔG	ΔE	ΔG
NTA(from Zn(NTA)) – NTA_{FL}	22.355	23.852	22.373	23.811
NTPA(from Zn(NTPA)) –$NTPA_{FL}$	37.342	41.987	36.911	42.567
NTA(from Ni(NTA)) –NTA_{FL}	24.153	25.964	24.543	27.504
NTPA(from Ni(NTPA)) –$NTPA_{FL}$	39.964	44.679	40.202	44.814
$Zn(H_2O)_2^a$(from Zn(NTA)) – $Zn(H_2O)_2^b$(from $Zn(H_2O)_6$)	2.268	2.477	2.226	2.559
$Zn(H_2O)_2^a$ (from Zn(NTPA)) – $Zn(H_2O)_2^b$(from $Zn(H_2O)_6$)	6.553	6.940	4.738	3.142
$Ni(H_2O)_2^a$ (from Ni(NTA)) – $Ni(H_2O)_2^b$(from $Ni(H_2O)_6$)	2.155	2.198	3.028	3.373
$Ni(H_2O)_2^a$(from Ni(NTPA)) – $Ni(H_2O)_2^b$(from $Ni(H_2O)_6$)	35.870	34.976	33.769	33.778

All energies are reported in kcal/mol, FL = Fully energy optimized free ligand.

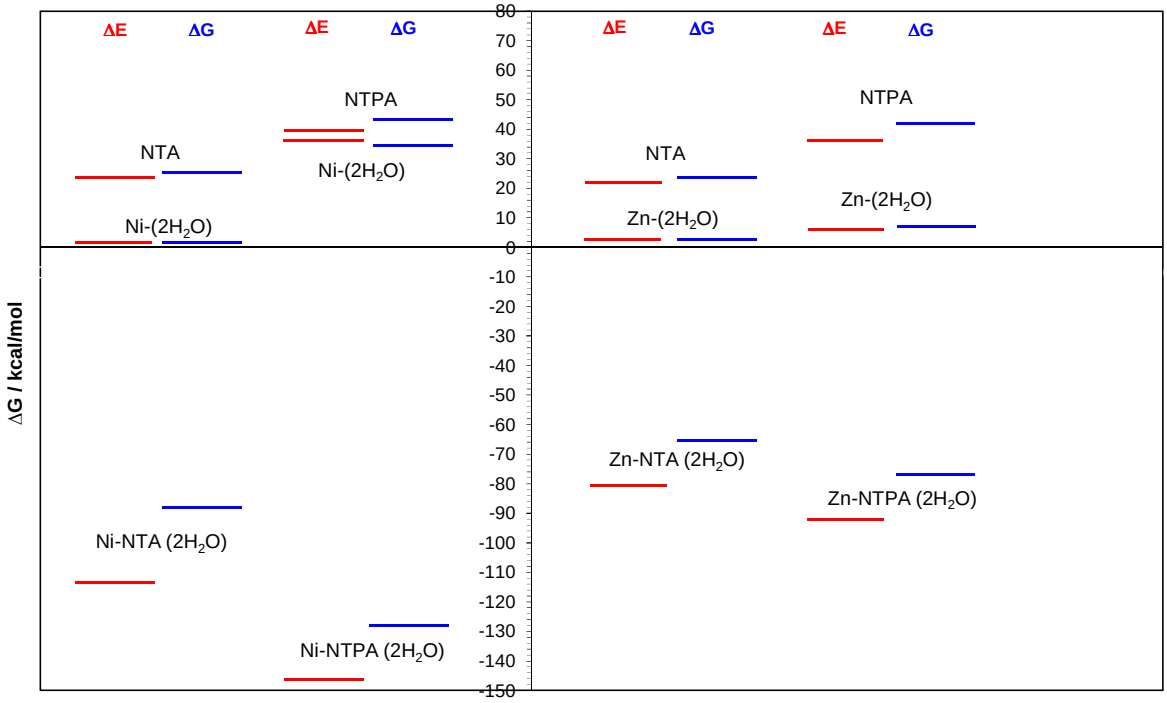


Figure 4.5: Energy of metal complexes and strain energies of different fragments present within each of the metal complexes.

This is illustrated in Figure 4.5, where (for clarity) only the energies for the constructed structures have been plotted. The energies related to the MM/MD structures are not shown

because the difference in energies for the constructed and MM/MD structures are rather small and a graph for the MM/MD structures would look exactly the same. The $\Delta G = 0$ in Figure 4.5 is an arbitrary point representing the energy levels of either fully optimized free ligands, sum of minimum energies of all components forming a complex, or the $M(H_2O)_2$ fragment from the aqueous metal complex – the change in energy for the fully energy optimised complexes and energies obtained from SPC performed on indicated fragments is presented. An additional, unexpected and very interesting observation was made when data presented in Table 4.8 were analysed. The ratio of strain energies for NTPA (ΔG = 41.987 kcal/mol) and NTA (ΔG = 23.852 kcal/mol) involved in Zn complexes is 1.76. This value compares well with the ratio of stability constants for the two ligands (log $K_{Zn(NTA)}$ = 10.45 and log $K_{Zn(NTPA)}$ = 5.3) which is 1.97. One must note that with an increase in strain energy the decrease in stability constant must be observed, hence the ratio of log $K_{M(NTA)}$ / log $K_{M(NTPA)}$ was calculated to compare the two ratios. Similar observation can be made for Ni, where the ratio of strain energies for NTPA (ΔG = 44.679 kcal/mol) and NTA (ΔG = 25.964 kcal/mol) involved in Ni complexes is 1.72. Here again, this value compares well with the ratio of stability constants for the two ligands (log $K_{Ni(NTA)}$ = 11.5 and log $K_{Ni(NTPA)}$ = 5.8) which is 1.98. From this one can conclude that the main contribution to the decrease in the stability of Ni and Zn complexes with NTPA, when compared with NTA, comes from the strain energy of ligands involved, the 'penalty' energy the ligand pays to form the complex. As far as one can assess, the above observation is reported for the first time and hence it is of utmost importance to find out whether the relationship between the computed strain energies (expressed as ratio) and the ratio of appropriate stability constants discovered here is a common feature. To establish that, many more analogous systems must be studied.

4.3.6. *AIM analysis*

With all of the analyses being performed up to this point there was just one more technique left to use, as far as available resources were concerned. This technique involved analysis of molecules with the aid of the QTAIM of Bader.[13] From these analyses the bond critical points (BCPs), ring critical points (RCPs) and cage critical points (CCPs), when present in each of the complexes, were generated. BCPs indicate that there is a bond present between two atoms, RCPs show the formation of a ring between three or more atoms and CCPs indicate the presence of a cage which is usually located within a number of RCPs.

Table 4.9 provides the topological properties of the electron charge density, (ρ) and the Laplacian ($\nabla^2\rho$) of the charge density, at selected BCPs (relevant to the central metal ion and intra-molecular interactions) and all ring as well as cage critical points found in the ML complexes examined here. Figure 4.6 provides the molecular graphs generated from AIM 2000 as well as the fully labelled Gaussian optimized structures whose atom labelling was utilized in Table 4.9; only the optimized constructed structures are provided (images for the optimized MM/MD structures can be obtained from Figure C5 of Appendix C). In both Table 4.9(a) and (b) the first three M–O bonds (under the BCP heading) refer to the bonds between the metal and the ligand (referred to as M–O_L), whereas the two immediately following M–O bonds represent the bonds between the metal and water (referred to as M–O_{H_2O}).

When comparing the ρ_{BCP} of Zn(NTA) with Zn(NTPA) it is seen that the ρ_{BCP} of the M–N bond in Zn(NTPA) is slightly larger than in Zn(NTA) suggesting that this bond in the weaker complex is somewhat stronger. The ρ_{BCP} at the M–O_L bonds, however, are identical between the two complexes. Unfortunately, the analysis of ρ_{BCP} related to four most important bonds in these complexes does not provide any insight and hence any explanation on why the Zn(NTA) complex is orders of magnitude stronger than the Zn(NTPA) complex. When looking at the M–O_{H_2O} ρ_{BCP} for the Zn complexes it was found that the Zn(NTPA) possessed values that are only half of that in Zn(NTA). It was then concluded that most likely the M–O_{H_2O} bonds must also be responsible for the reduced stability of Zn(NTPA). It shall also be seen in Table 4.9 that the Zn(NTPA) complex has both H–H and H–O BCPs that Zn(NTA) does not possess. The presence and importance of these bonds shall be discussed in more detail when considering the RCPs as these bonds play a much bigger role when present as part of a ring.

Attention is now turned to analysis of ρ_{RCP} values for the Zn(NTA) and Zn(NTPA) complexes. By focusing on the RCPs formed with the –COO^- containing arms it is seen that the Zn(NTA) values (both ρ_{RCP} and $\nabla^2\rho_{RCP}$) are twice as large than in Zn(NTPA). This is another factor that most likely also contributes to the higher stability of Zn(NTA). Moreover, it was also observed that the Zn(NTA) complex possessed only three RCPs, whereas the Zn(NTPA) complex possessed seven RCPs. This applies to both, the constructed and MM/MD, structures even though the size of the rings differs between the two structures.

Table 4.9: Rho (ρ) and Laplacian ($\nabla^2\rho$) values obtained for the different bond, ring and cage critical points of the solvent optimized (a) Zn^{2+} and (b) Ni^{2+} complexes of NTA and NTPA.

(a)

Atoms	Constructed structures		MM/MD structures	
Zn(NTA)				
BCP	ρ	$\nabla^2\rho$	ρ	$\nabla^2\rho$
Zn26–N1	0.05759	0.17739	0.05801	0.17941
Zn26–O12	0.06422	0.28561	0.06421	0.28484
Zn26–O8	0.06495	0.29290	0.06446	0.28901
Zn26–O10	0.06254	0.27407	0.06429	0.28663
Zn26–O21	0.04360	0.17270	0.04159	0.15909
Zn26–O20	0.05438	0.24071	0.05469	0.24061
RCP				
Zn26–N1–C2–C5–O10	0.02091	0.09528	0.02075	0.09460
Zn26–N1–C4–C6–O8	0.01966	0.08961	0.01980	0.09019
Zn26–N1–C3–C7–O12	0.02107	0.09653	0.02131	0.09830
Zn(NTPA)				
BCP				
Zn1–N2	0.06160	0.19957	0.06076	0.19479
Zn1–O20	0.06629	0.31625	0.06456	0.29920
Zn1–O18	0.06473	0.30498	0.06421	0.29579
Zn1–O19	0.06860	0.32503	0.06498	0.30593
Zn1–O3	0.02827	0.09411	0.03681	0.13124
Zn1–O4	0.04564	0.17944	0.04528	0.17786
H27–H28	0.01325	0.05039	0.01299	0.04715
H31–H32	0.01388	0.04848	–	–
H35–O3	0.00889	0.03009	–	–
H24–O3	0.00774	0.02638	–	–
H31–H34	–	–	0.00968	0.03326
H33–O3	–	–	0.01036	0.03503
RCP				
Zn1–N2–C12–C13–C14–O18	0.01146	0.04993	0.01131	0.05005
Zn1–N2–C15–C16–C17–O19	0.01198	0.05264	0.01174	0.05122
Zn1–N2–C9–C10–C11–O20	0.01161	0.05093	0.01152	0.04942
N2–C9–C10–H27–H28–C12	0.01311	0.05767	0.01279	0.05541
N2–C15–H32–H31–C13–C12	0.01313	0.05954	–	–
Zn1–O19–C17–C16–H35–O3	0.00621	0.02707	–	–
Zn1–N2–C9–H24–O3	0.00600	0.02538	–	–
Zn1–N2–C15–H33–O3	–	–	0.00754	0.03204
N2–C12–C13–H31–H34–C16–C15	–	–	0.00812	0.03553
Zn1–O19–C17–C16–H34–H31–C13–C14–O18	–	–	0.00547	0.02115
CCP				
Refer to Figure 4.6	0.00453	0.02104	0.00497	0.02204

(b)

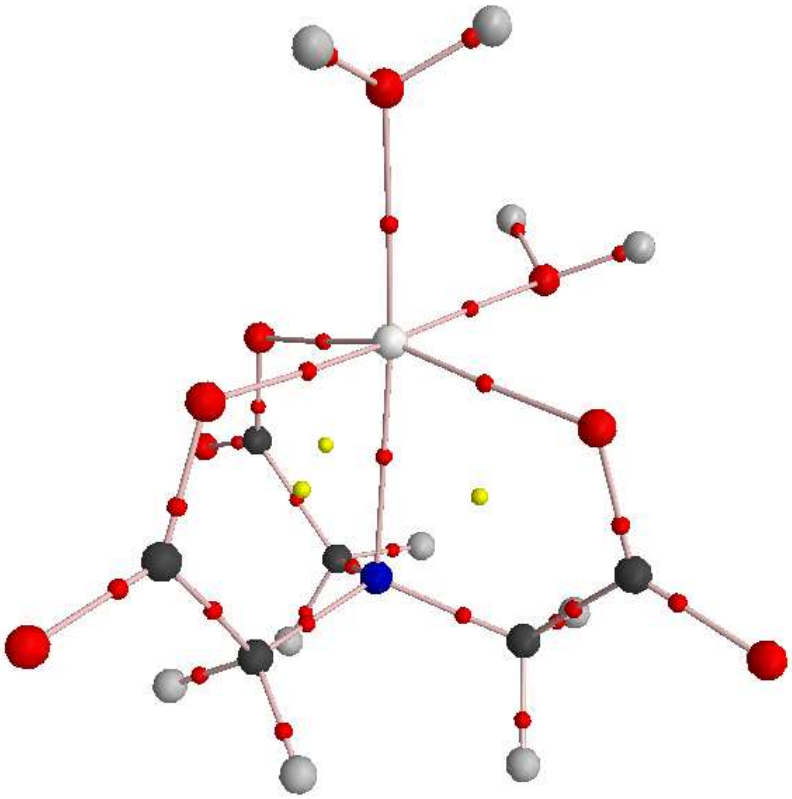
Atoms	Constructed structures		MM/MD structures	
	Ni(NTA)			
BCP	ρ	$\nabla^2\rho$	ρ	$\nabla^2\rho$
Ni1–N2	0.07113	0.30746	0.07220	0.31367
Ni1–O15	0.06971	0.40015	0.06849	0.39050
Ni1–O16	0.06289	0.33479	0.06408	0.34545
Ni1–O17	0.06328	0.34009	0.06427	0.34778
Ni1–O18	0.05310	0.30885	0.05301	0.30605
Ni1–O19	0.04820	0.26964	0.04762	0.26139
RCP				
Ni1–N2–C7–C10–O15	0.02204	0.10848	0.02208	0.10884
Ni1–N2–C11–C14–O16	0.02291	0.11267	0.02296	0.11300
Ni1–N2–C3–C6–O17	0.02263	0.11066	0.02278	0.11171
	Ni(NTPA)			
BCP				
Ni35–N1	0.06715	0.28892	0.06803	0.29528
Ni35–O11	0.06534	0.38459	0.06472	0.37401
Ni35–O13	0.06571	0.38354	0.06631	0.38445
Ni35–O15	0.06632	0.37962	0.06447	0.37544
Ni35–O17	0.03911	0.19033	0.04442	0.23003
Ni35–O18	0.04547	0.24369	0.04667	0.24791
H32–H25	0.01444	0.05060	–	–
H33–O17	0.00919	0.03196	–	–
H28–O17	0.01108	0.03855	–	–
H34–H25	–	–	0.00982	0.03427
H24–H29	–	–	0.01385	0.04914
H31–O17	–	–	0.01188	0.04159
RCP				
Ni35-N1-C8-C9-C10-O15	0.01261	0.05730	0.01234	0.05564
Ni35-N1-C2-C3-C4-O11	0.01229	0.05543	0.01212	0.05557
Ni35-N1-C5-C6-C7-O13	0.01206	0.05495	0.01213	0.05404
N1-C2-C3-H25-H32-C8	0.01355	0.06208	–	–
Ni35-O17-H33-C9-C10-O15	0.00664	0.02904	–	–
Ni35-O17-H28-C5-N1	0.00843	0.03636	–	–
N1-C5-H28-O17-H33-C9-C8	0.00615	0.02345	–	–
N1-C2-C3-H25-H34-C9-C8	–	–	0.00812	0.03616
Ni35-O11-C4-C3-H25-H34-C9-C10-O15	–	–	0.00592	0.02347
Ni35-O17-H31-C8-N1	–	–	0.00917	0.03966
N1-C5-C6-H29-H24-C2	–	–	0.01335	0.05958
CCP				
Refer to Figure 4.6	0.00536	0.02511	0.00534	0.02400

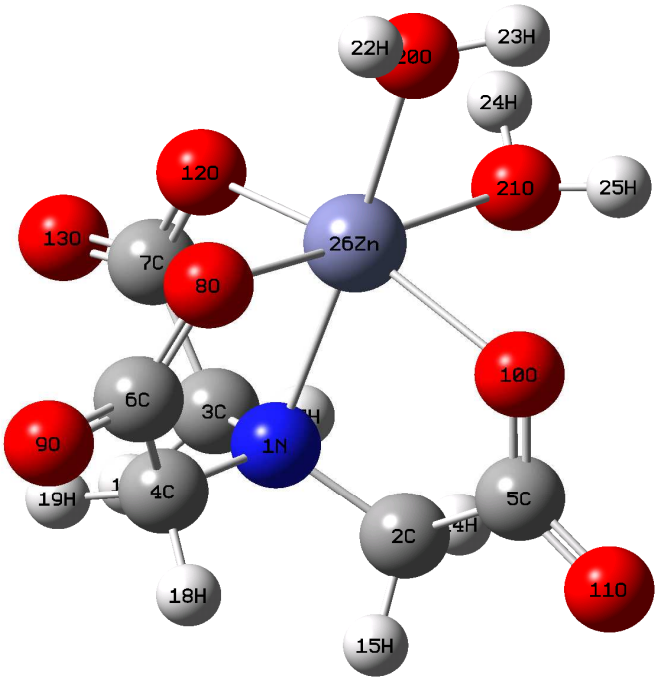
Three of the rings were formed by the $-COO^-$ containing arms, two of the rings where due to H-clashes and, interestingly, the values of ρ_{RCP} obtained for these rings were significantly larger than those obtained for the rings involving $-COO^-$ containing arms which is indicative of strong (energy stabilizing) contribution by C–H···H–C bonds to the overall stability of the

molecule. This observation is completely opposite to a previous and 'orthodox' explanation in which H-clashes, being a result of steric crowding, resulted in destabilization of a complex.[17]

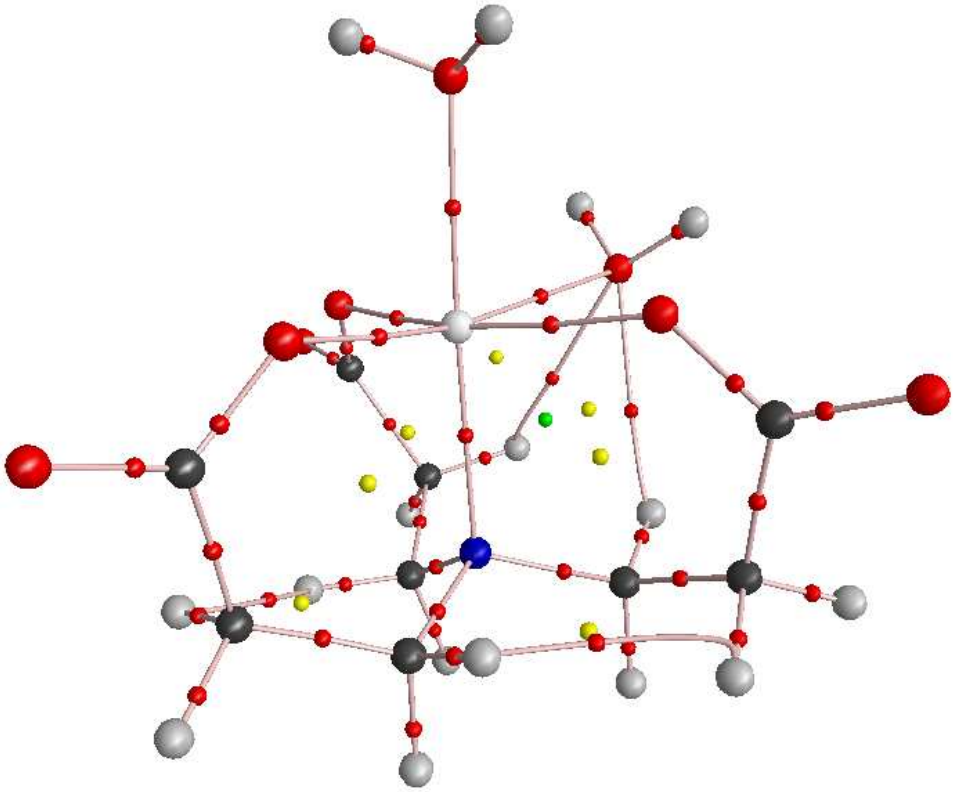
In addition, it was noticed that in the case of the rings formed by the H-clashes, there was a difference in the size of the rings between the constructed and MM/MD structures. In the case of the constructed structures there were two six membered rings formed, whereas the MM/MD structures possessed one six membered, one seven membered and one nine membered ring and as the size of the ring increases the ρ_{RCP} and $\nabla^2\rho_{RCP}$ values decrease. It was therefore only the six membered rings that possessed ρ_{RCP} values that were larger than the ρ_{RCP} values of the carboxylic acid arms. Finally the last two rings were formed due to the presence of C–H···O bonds.

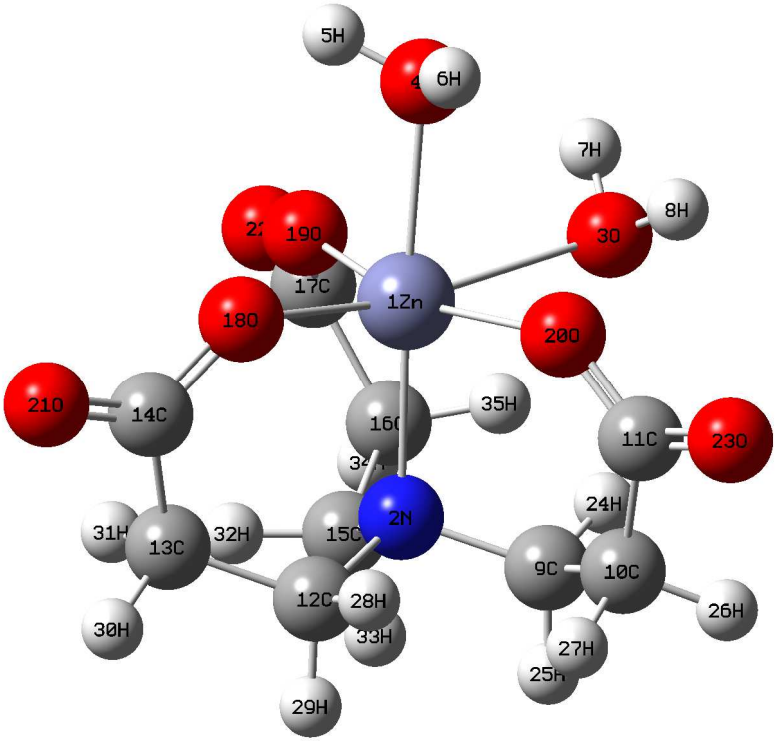
In the case of ρ_{BCP} in Ni(NTA) and Ni(NTPA), similar observations can be made as discussed for the Zn(NTA) and Zn(NTPA) complexes except the M–N ρ_{BCP} of Ni(NTA) that is larger than that of Ni(NTPA) which is the opposite to that obtained in the Zn complexes. When looking at the ρ_{RCP} of Ni(NTA) and Ni(NTPA) it is seen that RCPs of the –COO^-containing arms are twice as large in Ni(NTA) as compared to Ni(NTPA), which is exactly the same as observed for Zn(NTA) and Zn(NTPA) and this points to the fact that the Ni(NTPA) complex is less stable than the Ni(NTA) complex. What was different in Ni(NTPA) though was that in the constructed structure there was only one six membered ring formed with H-clashes, as compared to two six membered rings formed in Zn(NTPA). However, even though the number of six membered rings formed with H-clashes differed between the two complexes, the six membered ring (with H-clashes) present in Ni(NTPA) did possess a ρ_{RCP} value that was larger than that obtained for the rings involving the –COO containing arms, which has exactly the same implication as for Zn(NTPA). There were also three rings formed with C–H···O bonds as compared to two formed in Zn(NTPA). With the MM/MD Ni(NTPA) structure there was one six membered, one seven membered and one nine membered ring formed with H-clashes, which was exactly the same as the Zn(NTPA) MM/MD structure. There was also one ring formed with C–H···O bonds which again was exactly the same as that obtained for the Zn(NTPA) complex.



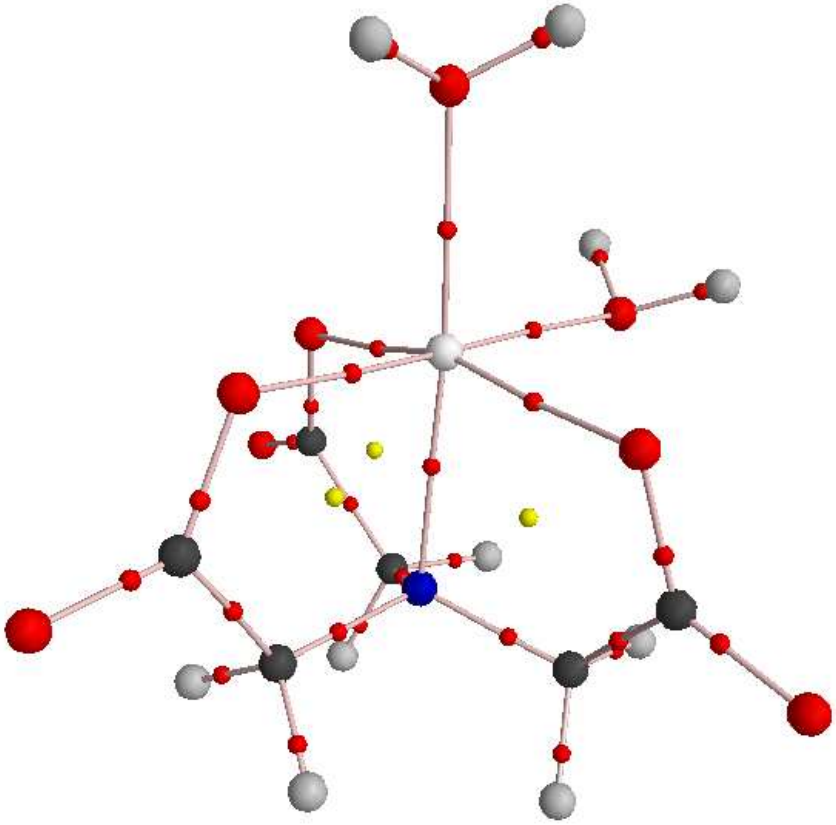

22H
200
23H
24H
120
210
25H
130
7C
26Zn
80
100
6C
3C
90
1N
19H
4C
5C
2C
18H
110
15H

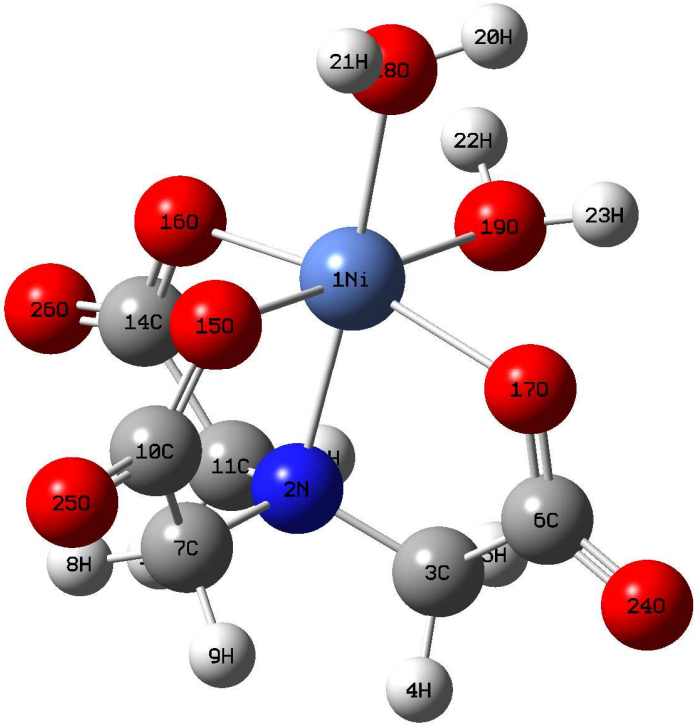
Zn(NTA)



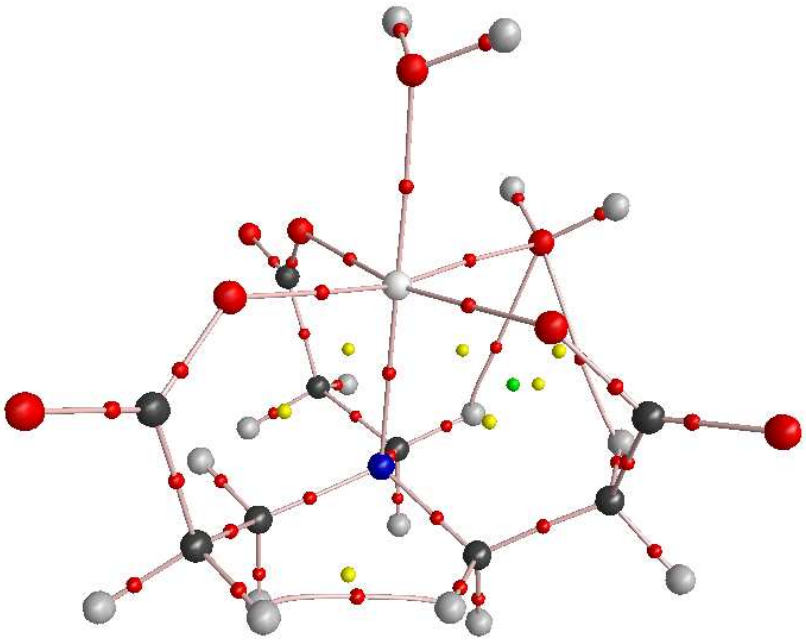

5H
6H
7H
8H
30
190
17C
1Zn
180
200
210
14C
16C
35H
11C
230
24H
2N
31H
13C
32H
15C
9C
10C
26H
12C
28H
33H
30H
27H
29H

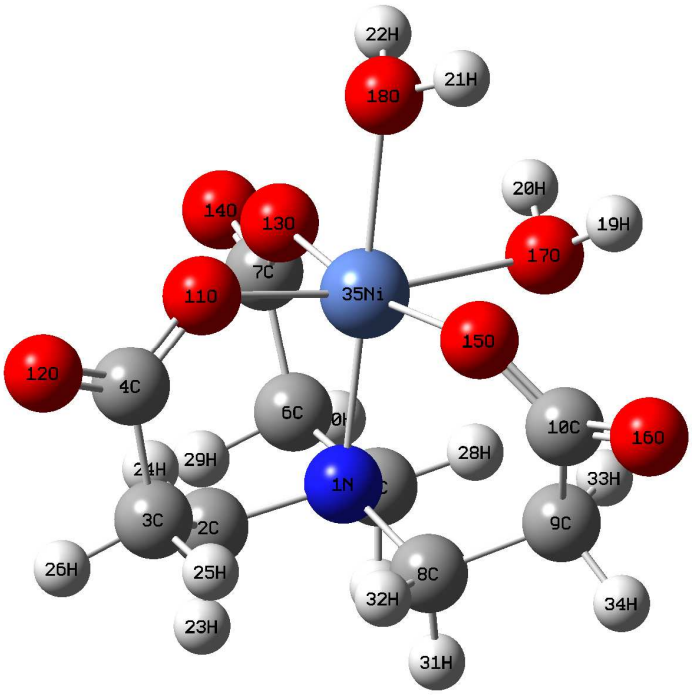
Zn(NTPA)




20H
21H
22H
160
190
23H
1Ni
260
14C
150
170
10C
11C
2N
250
6C
7C
8H
3C
240
9H
4H

Ni(NTA)





Ni(NTPA)

Figure 4.6: Molecular graphs and fully labelled solvent optimized constructed structures of Zn(NTA), Zn(NTPA), Ni(NTA) and Ni(NTPA). The bond critical points (BCPs) are denoted by red points, ring critical points (RCPs) are represented by yellow points and cage critical points (CCPs) can be seen as green points.

When the ρ_{BCP} and ρ_{RCP} of the $-COO^-$ containing arms of Zn(NTA) and Ni(NTA), as well as Zn(NTPA) and Ni(NTPA) were compared with one another it was found that the values did not differ from one another, in fact the values were almost identical between the complexes. This is probably due to the fact that the radii of the two metal ions are almost identical (as shown in Figure 1.3 of chapter 1).

Previous AIM analysis studies revealed that for some molecules the larger the ρ_{BCP} value, the shorter the bond length is observed.[42] It has been decided to perform the same kind of analysis on the metal complexes considered during these studies. Separate diagrams were constructed for Zn and Ni complexes with NTA and NTPA complexes; the relationship between ρ_{BCP} and the $M–O_L$ bond length is shown in Figure 4.7. It shall be seen that only the Zn and Ni constructed structures were used in this figure; the Zn and Ni MM/MD structures are presented in Figure C6 (Appendix C). The $M–O_{H_2O}$ bonds are not plotted in Figure 4.7 together with the $M–O_L$ bond because they are different kind of a bond and the ρ values differ largely from those observed in the $M–O_L$ bonds. Also, a separate relationship would only involve two points on the graph. From Figure 4.7 it is seen that Zn(NTA), Zn(NTPA) and Ni(NTA) complexes produced values which confirm that as ρ_{BCP} increases the bond lengths decrease. Ni(NTPA), on the other hand, possessed a trend that was completely opposite. When this was first determined it was really difficult to rationalize why this was the case, but when the MM/MD structures were analysed (Figure C6) it was found that the Ni(NTPA) complex follows the expected pattern, namely when ρ_{BCP} increases the bond length decreases. From this it was then concluded that the Ni(NTPA) complex of the constructed structure was most likely in the wrong conformation when it was optimized. This shows that even though the constructed and MM/MD structures did not differ that much conformationally, the minute difference that they do possess caused a dramatic difference in the results obtained. This emphasises that when performing computational optimization of molecules it is very important to ensure that the correct conformation is submitted as incorrect conformations can lead to incorrect results and conclusions. Notwithstanding the above, it is clear that the AIM analysis provided a lot of important information that could be used to rationalise the experimentally and computationally observed differences in stability constants.

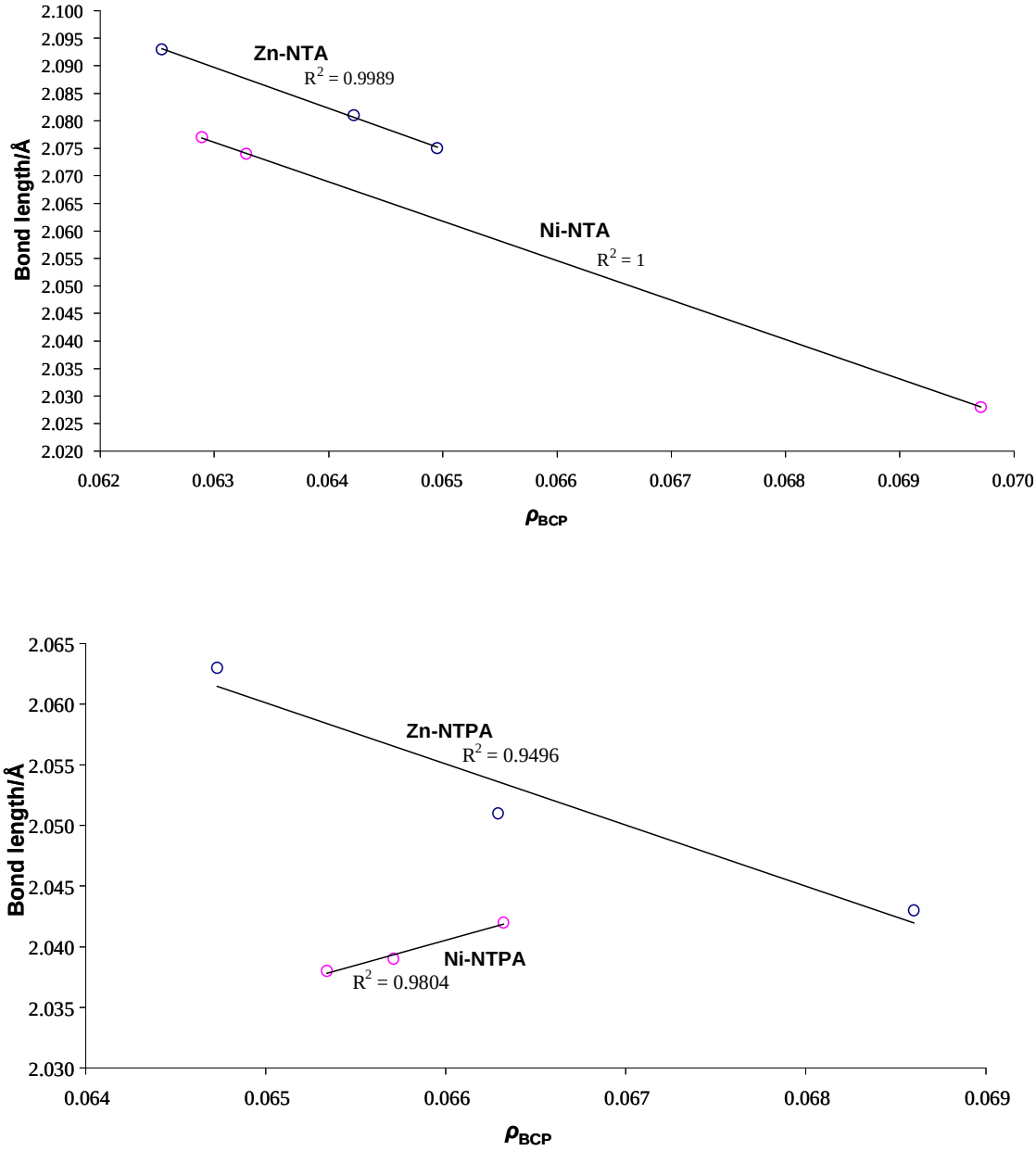


Figure 4.7: M–O_L ρ_{BCP} versus Bond length (Å) for Zn and Ni constructed complexes.

4.4. Conclusion

This study has provided important information as far as metal-ligand complexes are concerned. The isodesmic reactions that were used provided results (theoretically predicted stability constants) that follow the trend set by experiment, in that both the Zn^{2+} and Ni^{2+} metal complexes with NTPA are far less stable than the complexes formed with NTA. Unfortunately, the predicted stability constants were not very accurate, at least not as accurate

as values predicted for protonation constants (as discussed in previous chapters). This is probably due to the use of incorrect conformers and a full conformation analysis may provide results that are more accurate.

Structural analysis revealed that there were no significant differences present between the metal-donor atom, in $L_{(1)}$ and $L_{(2)}$ (where $L_{(1)}$ = NTPA and $L_{(2)}$ = NTA), bond lengths. On average the differences were approximately 0.020 Å. In addition it was discovered that the rule of preferential bite angles does not hold for any of the complexes studied. In general the information obtained from structural analysis is limited in explaining the difference in stability of complexes.

With NBO analysis it was discovered that the charges obtained for the different complexes did not differ much from one another and because of this they did not provide any sought after information; in order to provide an explanation in terms of charge, the charges need to be considerably different from one another. What proved to be of interest was the second order perturbation energy ($E^{(2)}$). The results obtained showed an exceptional trend that was followed for Zn(NTA) and Zn(NTPA) with only the nitrogen atom of NTPA deviating from this trend. Due to the steric crowding present in the NTPA ligand when complexed to Zn, it is not possible for the nitrogen atom to come closer to the Zn and this resulted in the deviation. However, because the nitrogen atom is further away from the Zn, a larger electron transfer has to take place in the case of NTPA – without which the Zn–N bond would be too weak to keep the metal ion and ligand together. This was an interesting observation that was made and additional studies will need to be conducted in order to determine if this is a trend that applies to other metal-ligand complexes. In the case of Ni(NTA) and Ni(NTPA) the results obtained, unfortunately, did not follow the trend set by Zn(NTA) and Zn(NTPA). A possible explanation for this is that the electron configuration of Ni differs from that of Zn. However, at this point the significance and physical meaning of why this is the case is not fully understood and in order to provide a through explanation more Ni complexes need to be studied.

The molecular fragment analysis showed that in order for a ligand (being NTA or NTPA) to reach the structural arrangement observed in a metal complex additional (strain) energy is required. Furthermore it was shown that NTPA requires more strain energy than NTA in order to complex to a metal ion. This is an aspect that correlates the strength of metal complexes with these two ligands. One very interesting observation was the correlation

between the ratio of strain energies and the ratio of experimental stability constants for the Zn and Ni complexes. In the case of the Zn complexes the strain energy ratio was 1.76 and the experimental stability constant ratio was 1.97, whereas for the Ni complexes the strain energy ratio was 1.72 and the experimental stability constant ratio was 1.98. From this it is clear that the main contribution to the decrease in the stability of Ni and Zn complexes with NTPA, when compared to NTA, comes from the strain energy of ligands involved.

With AIM analysis it was found that the ρ_{BCP} of the M–O_L bonds are almost identical when comparing Zn(NTA) with Zn(NTPA) and Ni(NTA) with Ni(NTPA). However, the ρ_{BCP} of the M–O_{H_2O} bonds of the M(NTPA) complexes were half of that in the M(NTA) complexes. This indicates that the M–O_{H_2O} bonds are also responsible for the reduced stability of M(NTPA) complexes. It was discovered that in the case of Zn(NTA) and Ni(NTA) only three RCPs were obtained (which were formed by the $–COO^-$ containing arms), whereas for Zn(NTPA) and Ni(NTPA) there were seven RCPs. The additional RCPs present in the M(NTPA) complexes are due to H-clashes (C–H···H–C bonds) as well as C–H···O bonds. Interestingly it was found that the ρ_{RCP} values obtained for six membered rings (with H-clashes) were significantly larger than the ρ_{RCP} values obtained for the $–COO^-$ containing arms of the M(NTPA) complexes. This indicated that the H-clashes that were present in the M(NTPA) complexes were in fact locally contributing to the overall stability of the molecule. This observation is completely opposite to a previous and ‘orthodox’ explanation in which H-clashes, being a result of steric crowding, resulted in destabilization of a complex. Finally it was determined that all complexes, except Ni(NTPA) constructed structure, follow an excellent trend when plotting the ρ_{BCP} versus the bond length. In the case of Ni(NTPA) constructed structure the trend is opposite to those obtained for the other complexes, which indicates that an incorrect conformer was utilized for this molecule and other conformations of this complex will need to be examined (this could result in better predictions for stability constants).

4.5. References

1. C. Orvig, M. Abrams, *Chem. Rev.* **1999**, *99*, 2201.

2. R. H. Holm, P. Kennepohl, E. I. Solomon, *Chem. Rev.* **1996**, *96*, 2239.

3. K. H. Brown, S. E. Wuehler, J. M. Peerson, *Food Nutr. Bull.* **1999**, *22*, 113.

4. M. Anke, L. Angelow, M. Glei, M. Muller, H. Illing, *Fresenius' J. Anal. Chem.* **1995**, *352*, 92.

5. N. H. Gutmann, L. Spiccia, T. W. Turney, *J. Mater. Chem.* **2000**, *10*, 1219.

6. F. J. M. Rajabalee, *J. Inorg. Nucl. Chem.* **1974**, *36*, 557.

7. R. Puchta, N. v. E. Hommes, R. Meier, R. v. Eldik, *Dalton Trans.* **2006**, 3392.

8. P. A. Petrenko, M. Gdaniec, Y. A. Simonov, V. G. Stavila, A. P. Gulea, *Russ. J. Coord. Chem.* **2004**, *30*, 813.

9. H. Thakuria, B. M. Borah, G. Das, *Eur. J. Inorg. Chem.* **2007**, *2007*, 524.

10. 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, US Department of Commerce, National Institute of Standards and Technology, 2004.

11. R. D. Hancock, L. J. Bartolotti, *Inorg. Chem.* **2005**, *44*, 7175.

12. A. E. Reed, F. Weinhold, *J. Chem. Phys.* **1983**, *78*, 4066.

13. R. F. W. Bader, 'Atoms in molecules', Claredon Press, Oxford, 1990.

14. I. Geogieva, D. Binev, N. Trendafilova, G. Bauer, *Chem. Phys.* **2003**, *286*, 205.

15. Y. X. Lu, J. W. Zou, Q. S. Yu, Y. J. Jiang, W. N. Zhao, *Chem. Phys. Lett.* **2007**, *449*, 6.

16. B. G. Oliveira, M. L. A. A. Vasconcellos, R. R. Olinda, E. B. A. Filho, *Struct. Chem.* **2009**, Article in Press.

17. A. E. Martell, R. D. Hancock, 'Metal Complexes in Aqueous Solutions', Plenum Press, New York, 1996.

18. C. F. Matta, J. Hernandez-Trujilo, T. H. Tang, R. F. W. Bader, *Chem. Eur. J.* **2003**, *9*, 1940.

19. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J.

Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Pettersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian 03, Revision D.01, Gaussian, Inc., Wallingford CT, 2004.

20. A. J. Lupinetti, V. Jonas, W. Thiel, S. H. Strauss, G. Frenking, *Chem-Eur. J.* **1999**, *5*, 2573.

21. E. Cances, B. Mennucci, J. Tomasi, *J. Chem. Phys.* **1997**, *107*, 3032.

22. M. Cossi, V. Barone, R. Cammi, J. Tomasi, *Chem. Phys. Lett.* **1996**, *255*, 327.

23. S. Miertus, E. Scrocco, J. Tomasi, *Chem. Phys.* **1981**, *55*, 117.

24. V. Barone, M. Cossi, *J. Phys. Chem. A* **1998**, *102*, 1995.

25. M. Cossi, N. Rega, G. Scalmani, V. Barone, *J. Comput. Chem.* **2003**, *24*, 669.

26. D. Gao, P. Svoronos, P. K. Wong, D. Maddalena, J. Hwang, H. Walker, *J. Phys. Chem. A* **2005**, *109*, 10776.

27. P. R. Varadwaj, I. Cukrowski, H. M. Marques, *J. Phys. Chem. A* **2008**, *112*, 10657.

28. L. L. C. Schrodinger, Schrodinger Maestro, 32nd Floor, Tower 45, 120 West Forty-Fifth Street, New York, 10036, 2003.

29. E. D. Glendening, J. K. Badenhoop, A. E. Reed, J. E. Carpenter, J. A. Bohmann, C. M. Morales, F. Weinhold, NBO 5.0., Theoretical Chemistry Institute, University of Wisconsin, Madison, 2001.

30. F. Biegler-Konig, AIM 2000 version 1.0, University of Applied Sciences, Bielefeld, Germany, 1998-2000.

31. http://www.aim2000.de/tutorial.htm

32. F. H. Allen, *Acta Cryst.* **2002**, *B58*, 380.

33. P. King, R. Clerac, W. Wernsdorfer, C. E. Anson, A. K. Powell, *Dalton Trans.* **2004**, 2670.

34. M. A. Salam, K. Aoki, *Inorg. Chim. Acta* **2000**, *311*, 15.

35. J. D. Oliver, B. L. Barnett, L. C. Strickland, *Acta Cryst., Sect. B.: Struct. Sci.* **1984**, *40*, 377.

36. M. D. Liptak, G. C. Shields, *J. Am. Chem. Soc.* **2001**, *123*, 7314.

37. M. Namazian, M. Zakery, M. R. Noorbala, M. L. Coote, *Chem. Phys. Lett.* **2008**, *451*, 163.

38. W. Sang-Aroon, V. Ruangpornvisuti, *Int. J. Quantum Chem.* **2008**, *108*, 1181.

39. M. D. Liptak, K. C. Gross, P. G. Seybold, S. Feldgus, G. C. Shields, *J. Am. Chem. Soc.* **2002**, *124*, 6421.

40. M. D. Liptak, G. C. Shields, *Int. J. Quantum Chem.* **2001**, *85*, 727.

41. R. H. Hertwig, J. Hrusak, D. Schroder, W. Koch, H. Schwarz, *Chem. Phys. Lett.* **1995**, *236*, 194.

42. P. R. Varadwaj, I. Cukrowski, H. M. Marques, *J. Mol. Struct. (Theochem)* **2009**, *902*, 21.

43. M. Namazian, H. Heidary, *J. Mol. Struct. (Theochem)* **2003**, *620*, 257.

44. M. Namazian, S. Halvani, M. R. Noorbala, *J. Mol. Struct. (Theochem)* **2004**, *711*, 13.

5. Conclusion

Due to its tremendous evolution over the years, theoretical/computational chemistry has become a large contributor to the world of chemistry. Considered a class of physical chemistry, it is now a technique used by researchers from all branches of chemistry such as organic, inorganic and analytical chemists. This dissertation illustrates this considerably well as studies conducted within incorporate aspects of solution chemistry (protonation and stability constant predictions), organic chemistry (free and protonated forms of the ligands NTA and NTPA) and inorganic chemistry (ML complexes).

Protonation constants for both NTA and NTPA were predicted, with considerable accuracy. Since all studies conducted to date involved prediction of these constants for molecules that were not highly charged, the results presented in this dissertation are the most accurate to date, and this is because a different methodology was utilized for these predictions (isodesmic reactions) as compared to those used previously (thermodynamic cycles). Unfortunately, the same accuracy was not achieved for the prediction of stability constants, but this was probably due to the conformation of complexes that was used. In order to improve on the theoretical stability constants that were obtained during these studies one would have to perform a full conformation analysis on all of the molecules studied. The theory did allow for the prediction of a trend as far as stability of the studied metal-ligand complexes were concerned. Theory complemented what was determined experimentally, hence complexes with NTA possess stability that is orders of magnitude larger than complexes formed with NTPA. This shows that even with the complexities introduced by metal ions, such as having unpaired electrons when working with anything but a d^{10} metal ion, it is possible to predict trends with the aid of theoretical applications, as was the case with the Ni^{2+} complexes were Ni^{2+} is a d^8 metal ion.

It was also discovered that by combining a number of techniques it is possible to achieve results that complement each other. As it was mentioned previously the ultimate goal of these studies was to determine why complexes with NTA are orders of magnitude more stable than complexes with NTPA, considering that there are minute structural differences present between the two ligands. Obviously this topic could not be tackled immediately as there were a number of aspects that needed to be learnt and understood before considering a means by which to try and explain why this is the case. However, the steps taken to get to the

ultimate goal were very structured and allowed for a much greater understanding and appreciation for theoretical applications. It can also be said, with great pleasure, that the goals which were set out to be achieved for this MSc study were all achieved successfully.

5.1. Future Research

Even though all of the goals of this study were achieved there are a few additional aspects that need to be considered, these include:

- Analyses of a number of different MM/MD (Schrödinger) generated conformers, as there is a likelihood that one of the conformers lies closer if not at the global energy minimum of the potential energy surface as compared to the conformers considered during these studies. If this is the case than the lower energy conformers would provide even more accurate predictions for both protonation and stability constants.
- Conformation analysis with the aid of Gaussian in order to obtain a structure that lies at the global energy minimum of the potential energy surface.
- Use of different levels of theory, such as X3LYP/6-311++G(d,p), to account for the hydrogen bonded interactions to a greater extent, as well as to better account for the negative charges present within the molecules.
- Different solvation models when performing solvent optimization of each of the molecules, such as CPCM/UAHF and CPCM/UAKS. Optimization with these models will take considerably large amounts of time though, as compared to the model considered for these studies, i.e. PCM/UA0, and there is no guarantee that these models will provide more accurate results as compared to those obtained during these studies.
- As far as isodesmic reactions are concerned, it is possible that making use of reference molecules, other than the ones considered for these studies, would provide more accurate protonation constant predictions; AIM analysis of all of the optimized ligand forms that were utilized during these studies might provide information as to why some reference molecules work better than others when predicting stepwise protonation constants.
- Studying complexes with a metal such as Cu^{2+}. This would be a challenge though as Cu^{2+} is a d^9 metal ion with varying geometries. A good approach to analyzing a metal

such as this would be to search the CSD for crystal structures with NTA and NTPA, which can then be used as input structures for all calculations that follow.

- Usage of different software packages such as ADF (Amsterdam Density Functional) to obtain quantities such as strain energies. This would reveal exactly were the complexes experience strain and could provide additional information as to why NTA complexes are more stable than NTPA complexes, however due to the unavailability of ADF at the time of this study, the results obtained within this dissertation can be considered as sufficient to explain the difference in complex stability. ADF would also provide a means to analyze metals such as lead for which Gaussian does not have parameters defined.

Appendices

Appendix A

Supplementary information related to results produced in Chapter 2 (sections 2.3.1 and 2.3.3)

Appendix B

Supplementary information related to results produced in Chapter 3 (sections 3.3.1 and 3.3.3)

Appendix C

Supplementary information related to results produced in Chapter 4 (sections 4.3.1–4.3.6)

Appendix A

Table A1: Comparison of experimental (CSD) and computed at the RB3LYP/6-311+G(d,p) level of theory selected bond distances and angles for the H_3L^* and H_3L forms of NTPA. Bond lengths and angles are in Å and °, respectively.

		Input structure and difference (Δ)[a]				
Atoms	**CSD data[b]**	**H_3L^***	**Δ**	**H_3L**	**Δ**	**δ[c]**
N-C1	1.507	1.503	0.004	1.513	-0.006	-0.010
N-C2	1.515	1.514	0.001	1.518	-0.003	-0.004
N-C3	1.509	1.507	0.002	1.536	-0.027	-0.029
C1-C4	1.517	1.53	-0.013	1.527	-0.010	0.003
C2-C5	1.511	1.526	-0.015	1.526	-0.015	0.000
C3-C6	1.507	1.522	-0.015	1.522	-0.015	0.000
C4-C7	1.494	1.514	-0.020	1.517	-0.023	-0.003
C5-C8	1.513	1.516	-0.003	1.519	-0.006	-0.003
C6-C9	1.517	1.545	-0.028	1.551	-0.034	-0.006
C7-O1	1.213	1.213	0.000	1.212	0.001	0.001
C8-O2	1.212	1.211	0.001	1.212	0.000	-0.001
C9-O3	1.257	1.245	0.012	1.257	0.000	-0.012
C7-O4	1.320	1.341	-0.021	1.335	-0.015	0.006
C8-O5	1.304	1.336	-0.032	1.335	-0.031	0.001
C9-O6	1.259	1.278	-0.019	1.258	0.001	0.020
C1-N-C2	110.2	111.7	-1.5	111.3	-1.1	0.4
C1-N-C3	110.1	111.4	-1.3	109.7	0.4	1.7
C2-N-C3	112.1	112	0.1	112.6	-0.5	-0.6
N-C1-C4	112.2	112.1	0.1	113	-0.8	-0.9
N-C2-C5	113.0	113.3	-0.3	113.1	-0.1	0.2
N-C3-C6	112.1	111.8	0.3	112.8	-0.7	-1.0
C1-C4-C7	112.5	113.5	-1.0	113.8	-1.3	-0.3
C2-C5-C8	114.7	115.2	-0.5	114.2	0.5	1.0
C3-C6-C9	112.3	111.3	1.0	112.7	-0.4	-1.4
C4-C7-O1	123.6	124.5	-0.9	123.1	0.5	1.4
C5-C8-O2	118.1	124	-5.9	122.6	-4.5	1.4
C6-C9-O3	112.6	117.9	-5.3	117.6	-5.0	0.3
C4-C7-O4	114.0	116.6	-2.6	113.6	0.4	3.0
C5-C8-O5	118.0	116.6	1.4	115.5	2.5	1.1
C6-C9-O6	113.9	111.6	2.3	113.1	0.8	-1.5

[a]) Δ = (experimental – computed) value
[b]) Average bond lengths and angles of crystal structures obtained from the CSD[66]
[c]) δ = (H_3L^* – H_3L) value

Table A2: ZPVE-corrected minimum and Gibbs free energies of the NTPA and IDA ligands obtained in solvent (water) at RB3LYP/6-311+G(d,p) level of theory and PCM/UA0 solvation model.

Species	E_{min} [a]	G_{aq} [a]
NTPA		
L^{3-}	–857.036516	–856.875606
HL^{2-}	–857.516284	–857.339621
H_2L^{-}	–857.979899	–857.79098
H_3L	–858.433776	–858.234959
H_3L^{*}	–858.441832	–858.240931
H_4L^{+}	–858.894457	–858.682587
IDA		
L^{3-}	–511.579558	–511.519032
HL^{2-}	–512.056062	–511.981098
H_2L^{-}	–512.509179	–512.423365
H_3L	–512.958694	–512.862287

[a]) In atomic unit, Hartree (1 Hartree = 627.5095 kcal/mol)

Table A3: ZPVE-corrected minimum and Gibbs free energies of the NTA ligand obtained in solvent (water) at the RB3LYP/6-311+G(d,p) level of theory and PCM/UA0 solvation model.

Species	E_{min} [a]	G_{aq} [a]
NTA		
L^{3-}	-738.936972	-738.978069
HL^{-}	-739.410772	-739.451765
H_2L^{-}	-739.852748	-739.893949
H_3L	-740.290659	-740.332226
H_4L^{+}	-740.722681	-740.764545
H_3L^{*}	-740.293694	-740.335141

[a]) In atomic unit, Hartree (1 Hartree = 627.5095 kcal/mol)

Table A4: Comparison of experimental (Exp) and calculated from isodesmic reactions protonation constants of NTPA, as log K_H, using protonation constants of the reference molecules NTA and IDA at ionic strength μ = 0 or 0.1 M and 25 °C.

$L_{(2)}$ = NTA		μ = 0.0 M, 25 °C (NTA)				μ = 0.1 M, 25 °C (NTA)			
Reaction	ΔG_{aq}	ΔG_1(aq)	log K_H	Exp	δ	ΔG_1(aq)	log K_H	Exp	δ
$L_{(1)}^{3-} + HL_{(2)}^{2-} = HL_{(1)}^{2-} + L_{(2)}^{3-}$	6.075	-8.023	5.88	9.49	3.61	-7.104	5.21	9.49	4.28
$L_{(1)}^{3-} + H_2L_{(2)}^{-} = HL_{(1)}^{2-} + HL_{(2)}^{2-}$	-13.699	-17.710	12.98	9.49	-3.49	-17.137	12.56	9.49	-3.07
$L_{(1)}^{3-} + H_3L_{(2)} = HL_{(1)}^{2-} + H_2L_{(2)}^{-}$	-14.322	-17.050	12.50	9.49	-3.01	-16.791	12.31	9.49	-2.82
$L_{(1)}^{3-} + H_3L_{(2)} = HL_{(1)}^{2-} + H_2L_{(2)}^{-}$	-16.151	-18.879	13.84	9.49	-4.35	-18.620	13.65	9.49	-4.16
$L_{(1)}^{3-} + H_4L_{(2)}^{+} = HL_{(1)}^{2-} + H_3L_{(2)}$	-21.718	-23.083	16.92	9.49	-7.43	-23.083	16.92	9.49	-7.43
$L_{(1)}^{3-} + H_4L_{(2)}^{+} = HL_{(1)}^{2-} + H_3L_{(2)}$	-19.890	-21.254	15.58	9.49	-6.09	-21.254	15.58	9.49	-6.09
$HL_{(1)}^{2-} + HL_{(2)}^{2-} = H_2L_{(1)}^{-} + L_{(2)}^{3-}$	14.017	-0.082	0.06	4.22	4.16	0.838	-0.61	4.22	4.83
$HL_{(1)}^{2-} + H_2L_{(2)}^{-} = H_2L_{(1)}^{-} + HL_{(2)}^{2-}$	-5.757	-9.768	7.16	4.22	-2.94	-9.195	6.74	4.22	-2.52
$HL_{(1)}^{2-} + H_3L_{(2)} = H_2L_{(1)}^{-} + H_2L_{(2)}^{-}$	-6.380	-9.108	6.68	4.22	-2.46	-8.849	6.49	4.22	-2.27
$HL_{(1)}^{2-} + H_3L_{(2)} = H_2L_{(1)}^{-} + H_2L_{(2)}^{-}$	-8.209	-10.938	8.02	4.22	-3.80	-10.678	7.83	4.22	-3.61
$HL_{(1)}^{2-} + H_4L_{(2)}^{+} = H_2L_{(1)}^{-} + H_3L_{(2)}$	-13.777	-15.141	11.10	4.22	-6.88	-15.141	11.10	4.22	-6.88
$HL_{(1)}^{2-} + H_4L_{(2)}^{+} = H_2L_{(1)}^{-} + H_3L_{(2)}$	-11.748	-13.312	9.76	4.22	-5.54	-13.312	9.76	4.22	-5.54
$H_2L_{(1)}^{-} + HL_{(2)}^{2-} = H_3L_{(1)} + L_{(2)}^{3-}$	18.648	4.549	-3.33	3.68	7.01	5.469	-4.01	3.68	7.69
$H_2L_{(1)}^{-} + H_2L_{(2)}^{-} = H_3L_{(1)} + HL_{(2)}^{2-}$	-1.126	-5.137	3.77	3.68	-0.09	-4.564	3.35	3.68	0.33
$H_2L_{(1)}^{-} + H_3L_{(2)} = H_3L_{(1)} + H_2L_{(2)}^{-}$	-3.578	-6.307	4.62	3.68	-0.94	-6.047	4.43	3.68	-0.75
$H_2L_{(1)}^{-} + H_3L_{(2)} = H_3L_{(1)} + H_2L_{(2)}^{-}$	-1.749	-4.477	3.28	3.68	0.40	-4.218	3.09	3.68	0.59
$H_2L_{(1)}^{-} + H_4L_{(2)}^{+} = H_3L_{(1)} + H_3L_{(2)}$	-7.317	-8.681	6.36	3.68	-2.68	-8.681	6.36	3.68	-2.68
$H_2L_{(1)}^{-} + H_4L_{(2)}^{+} = H_3L_{(1)} + H_3L_{(2)}$	-9.146	-10.510	7.70	3.68	-4.02	-10.510	7.70	3.68	-4.02
$H_3L_{(1)} + HL_{(2)}^{2-} = H_4L_{(1)}^{+} + L_{(2)}^{3-}$	16.358	2.260	-1.66	2.71[a]	4.37	3.179	-2.13	2.71[a]	5.04
$H_3L_{(1)} + H_2L_{(2)}^{-} = H_4L_{(1)}^{+} + HL_{(2)}^{2-}$	-3.416	-7.427	5.44	2.71[a]	-2.73	-6.854	5.02	2.71[a]	-2.31
$H_3L_{(1)} + H_3L_{(2)} = H_4L_{(1)}^{+} + H_2L_{(2)}^{-}$	-5.868	-8.596	6.30	2.71[a]	-3.59	-8.337	6.11	2.71[a]	-3.40
$H_3L_{(1)} + H_3L_{(2)} = H_4L_{(1)}^{+} + H_2L_{(2)}^{-}$	-4.039	-6.767	4.96	2.71[a]	-2.25	-6.508	4.77	2.71[a]	-2.06
$H_3L_{(1)} + H_4L_{(2)}^{+} = H_4L_{(1)}^{+} + H_3L_{(2)}$	-9.607	-10.971	8.04	2.71[a]	-5.33	-10.971	8.04	2.71[a]	-5.33
$H_3L_{(1)} + H_4L_{(2)}^{+} = H_4L_{(1)}^{+} + H_3L_{(2)}$	-11.436	-12.800	9.38	2.71[a]	-6.67	-12.800	9.38	2.71[a]	-6.67
$H_2L_{(1)}^{-} + HL_{(2)}^{2-} =$	14.900	0.802	-0.59	3.68	4.27	1.722	-1.26	3.68	4.94

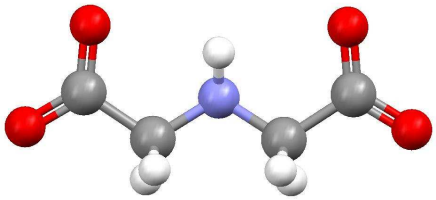
$H_3L_{(1)}^{*} + L_{(2)}^{3-}$ $H_2L_{(1)}^{-} + H_2L_{(2)}^{-} =$ $H_3L_{(1)}^{*} + HL_{(2)}^{2-}$	-4.874	-8.885	6.51	3.68	-2.83	-8.312	6.09	3.68	-2.41
$H_2L_{(1)}^{-} + H_3L_{(2)}^{*} =$ $H_3L_{(1)}^{*} + H_2L_{(2)}^{-}$	-5.496	-8.225	6.03	3.68	-2.35	-7.966	5.84	3.68	-2.16
$H_2L_{(1)}^{-} + H_3L_{(2)} =$ $H_3L_{(1)}^{*} + H_2L_{(2)}^{-}$	-7.326	-10.054	7.37	3.68	-3.69	-9.795	7.18	3.68	-3.50
$H_2L_{(1)}^{-} + H_4L_{(2)}^{+} =$ $H_3L_{(1)}^{*} + H_3L_{(2)}^{*}$	-12.893	-14.258	10.45	3.68	-6.77	-14.258	10.45	3.68	-6.77
$H_2L_{(1)}^{-} + H_4L_{(2)}^{+} =$ $H_3L_{(1)}^{*} + H_3L_{(2)}$	-11.064	-12.429	9.11	3.68	-5.43	-12.429	9.11	3.68	-5.43
$H_3L_{(1)}^{*} + HL_{(2)}^{2-} =$ $H_4L_{(1)}^{+} + L_{(2)}^{3-}$	20.105	6.007	-4.40	2.71[a]	7.11	6.927	-5.08	2.71[a]	7.79
$H_3L_{(1)}^{*} + H_2L_{(2)}^{-} =$ $H_4L_{(1)}^{+} + HL_{(2)}^{2-}$	0.331	-3.680	2.70	2.71[a]	0.01	-3.107	2.28	2.71[a]	0.43
$H_3L_{(1)}^{*} + H_3L_{(2)}^{*} =$ $H_4L_{(1)}^{+} + H_2L_{(2)}^{-}$	-0.291	-3.020	2.21	2.71[a]	0.50	-2.760	2.02	2.71[a]	0.69
$H_3L_{(1)}^{*} + H_3L_{(2)} =$ $H_4L_{(1)}^{+} + H_2L_{(2)}^{-}$	-2.120	-4.849	3.55	2.71[a]	-0.84	-4.590	3.36	2.71[a]	-0.65
$H_3L_{(1)}^{*} + H_4L_{(2)}^{+} =$ $H_4L_{(1)}^{+} + H_3L_{(2)}^{*}$	-7.688	-9.053	6.64	2.71[a]	-3.93	-9.053	6.64	2.71[a]	-3.93
$H_3L_{(1)}^{*} + H_4L_{(2)}^{+} =$ $H_4L_{(1)}^{+} + H_3L_{(2)}$	-5.859	-7.223	5.29	2.71[a]	-2.58	-7.223	5.29	2.71[a]	-2.58

$L_{(2)}$ = IDA — **μ = 0.0 M, 25 °C (IDA)** — **μ = 0.1 M, 20 °C (IDA)**

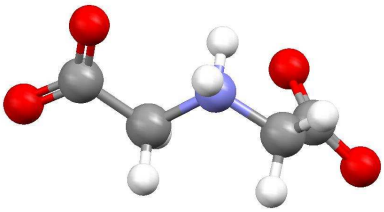
Reaction	ΔG_{aq}	ΔG_1(aq)	log K_H	Exp	δ	ΔG_1(aq)	log K_H	Exp	δ
$L_{(1)}^{3-} + HL_{(2)}^{-} =$ $HL_{(1)}^{2-} + L_{(2)}^{2-}$	-1.223	-14.579	10.69	9.49	-1.20	-13.965	10.24	9.49	-0.75
$L_{(1)}^{3-} + H_2L_{(2)} =$ $HL_{(1)}^{2-} + HL_{(2)}^{-}$	-13.647	-17.522	12.84	9.49	-3.35	-17.221	12.62	9.49	-3.13
$L_{(1)}^{3-} + H_3L_{(2)}^{+} =$ $HL_{(1)}^{2-} + H_2L_{(2)}$	-15.746	-18.270	13.39	9.49	-3.90	-18.161	13.31	9.49	-3.82
$HL_{(1)}^{2-} + HL_{(2)}^{-} =$ $H_2L_{(1)}^{-} + L_{(2)}^{2-}$	6.719	-6.637	4.87	4.22	-0.65	-6.023	4.42	4.22	-0.20
$HL_{(1)}^{2-} + H_2L_{(2)} =$ $H_2L_{(1)}^{-} + HL_{(2)}^{-}$	-5.705	-9.580	7.02	4.22	-2.80	-9.280	6.80	4.22	-2.58
$HL_{(1)}^{2-} + H_3L_{(2)}^{+} =$ $H_2L_{(1)}^{-} + H_2L_{(2)}$	-7.804	-10.328	7.57	4.22	-3.35	-10.219	7.49	4.22	-3.27
$H_2L_{(1)}^{-} + HL_{(2)}^{-} =$ $H_3L_{(1)} + L_{(2)}^{2-}$	11.350	-2.006	1.47	4.22	2.21	-1.392	1.02	3.68	2.66
$H_2L_{(1)}^{-} + H_2L_{(2)} =$ $H_3L_{(1)} + HL_{(2)}^{-}$	-1.074	-4.949	3.63	4.22	0.05	-4.649	3.41	3.68	0.27
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} =$ $H_3L_{(1)} + H_2L_{(2)}$	-3.173	-5.697	4.18	4.22	-0.50	-5.588	4.10	3.68	-0.42
$H_3L_{(1)} + HL_{(2)}^{-} =$ $H_4L_{(1)}^{+} + L_{(2)}^{2-}$	9.060	-4.296	3.15	2.71[a]	-0.44	-3.682	2.70	2.71[a]	0.01
$H_3L_{(1)} + H_2L_{(2)} =$ $H_4L_{(1)}^{+} + HL_{(2)}^{-}$	-3.364	-7.239	5.31	2.71[a]	-2.60	-6.938	5.09	2.71[a]	-2.38
$H_3L_{(1)} + H_3L_{(2)}^{+} =$ $H_4L_{(1)}^{+} + H_2L_{(2)}$	-5.463	-7.987	5.85	2.71[a]	-3.14	-7.878	5.77	2.71[a]	-3.06
$H_2L_{(1)}^{-} + HL_{(2)}^{-} =$ $H_3L_{(1)}^{*} + L_{(2)}^{2-}$	7.602	-5.754	4.22	3.68	-0.54	-5.140	3.77	3.68	-0.09
$H_2L_{(1)}^{-} + H_2L_{(2)} =$ $H_3L_{(1)}^{*} + HL_{(2)}^{-}$	-4.822	-8.696	6.37	3.68	-2.69	-8.396	6.15	3.68	-2.47
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} =$ $H_3L_{(1)}^{*} + H_2L_{(2)}$	-6.921	-9.445	6.92	3.68	-3.24	-9.336	6.84	3.68	-3.16

$H_3L_{(1)}^{*} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L_{(2)}^{2-}$	12.807	-0.549	0.40	2.71[a]	2.31	0.065	-0.05	2.71[a]	2.76
$H_3L_{(1)}^{*} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL_{(2)}^{-}$	0.383	-3.491	2.56	2.71[a]	0.15	-3.191	2.34	2.71[a]	0.37
$H_3L_{(1)}^{*} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L_{(2)}$	-1.716	-4.239	3.11	2.71[a]	-0.40	-4.130	3.03	2.71[a]	-0.32

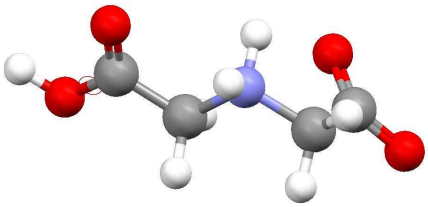
[a]) Experimental NTPA protonation constant[4,5] at μ = 0.5 M and 25 ^{o}C.
All energies are reported in kcal/mol.



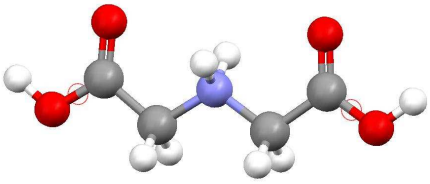
L^{3-}



HL^{2-}



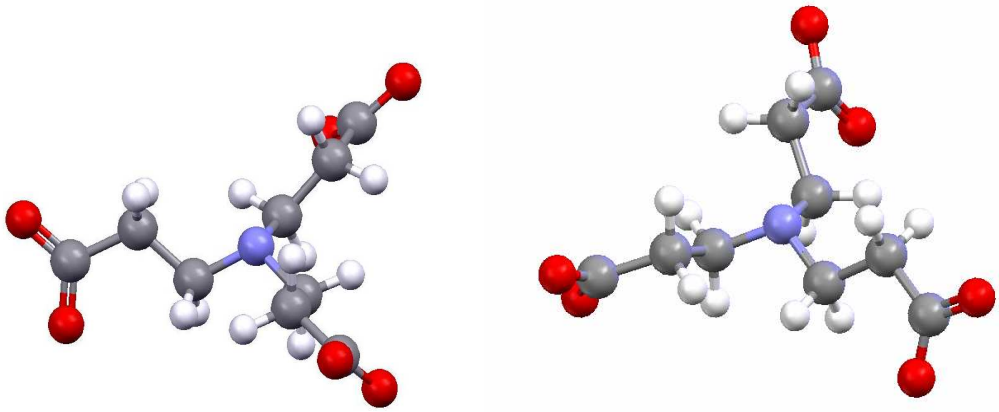
H_2L



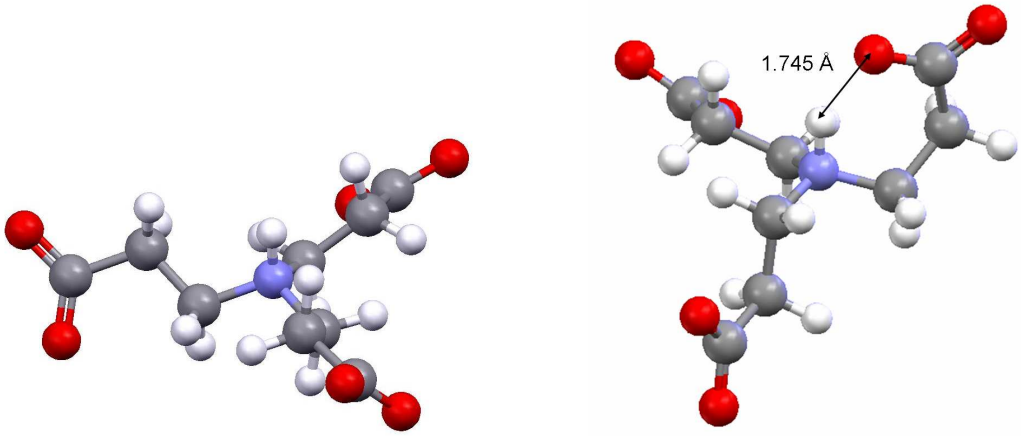
H_3L^{+}

Figure A1: Solvent (water) optimized structures of IDA at the RB3LYP/6-311+G(d,p) level of theory in combination with the solvation model PCM/UA0.

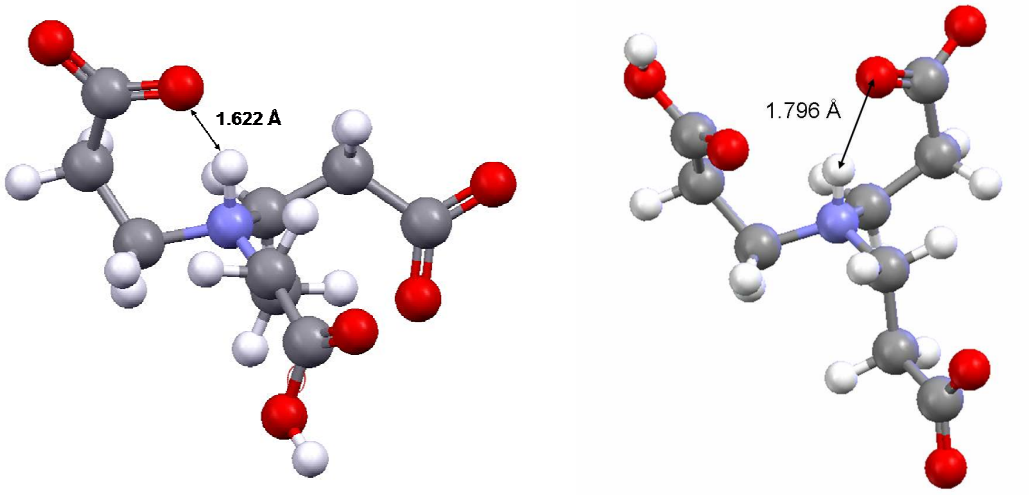
DFT-generated STRUCTURES **MM/MD-generated STRUCTURES**



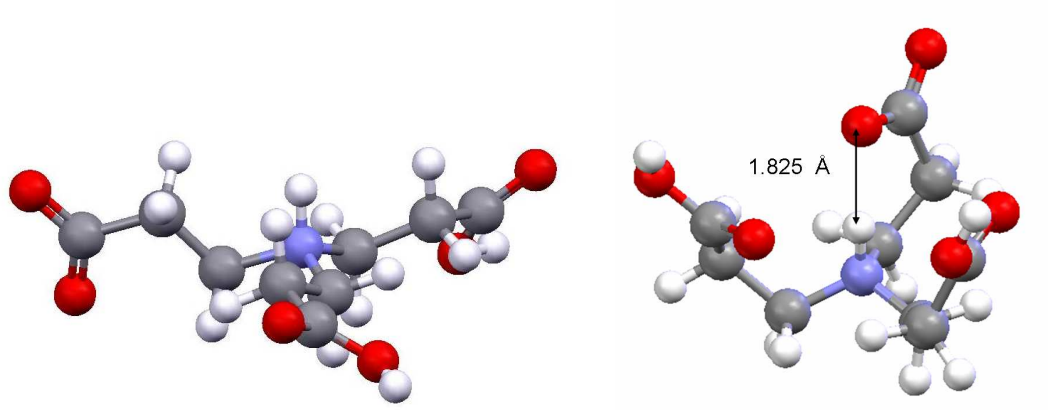
L^{3-}



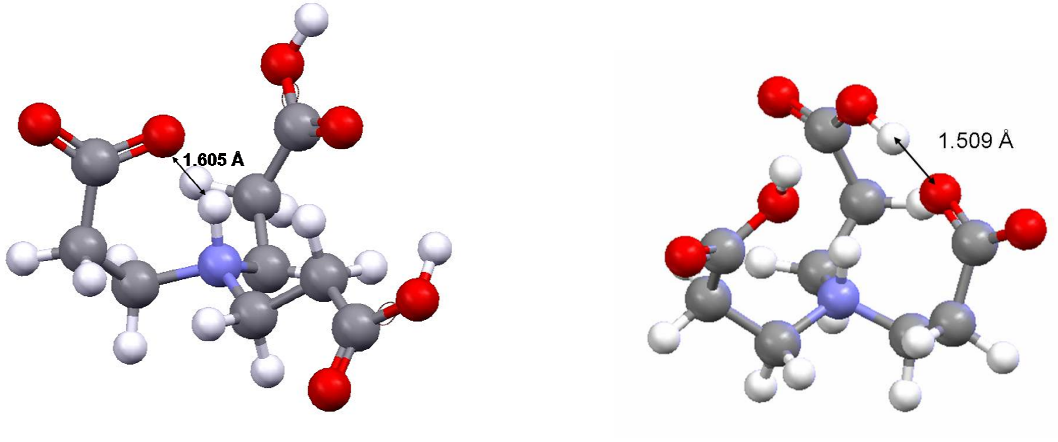
HL^{2-}



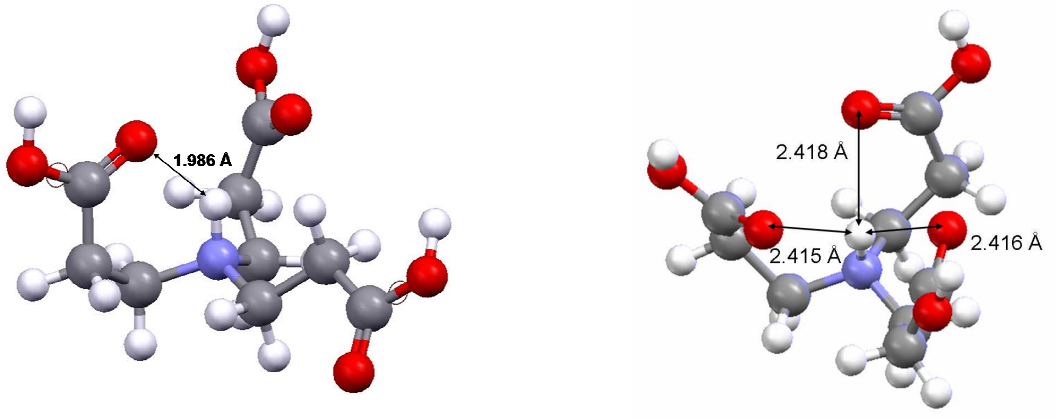
H_2L^-



H_3L



H_3L^*



H_4L^+

Figure A2: Energy-minimized in solvent structures of NTPA at the RB3LYP/6-311+G(d,p) level of theory in conjunction with the PCM/UA0 solvation model (DFT-generated structures, as seen also in Fig. 2) and lowest energy structures obtained from MM/MD-based conformational analysis performed on the DFT-optimized relevant structures.

Appendix B

Table B1: Comparison of experimental (CSD) and computed at the RB3LYP/6-311+G(d,p) level of theory in conjunction with PCM/UA0 solvation model selected bond distances and angles for the H_2L form of IDA. Bond lengths and angles are in Å and deg, respectively.

Atoms	CSD Data[a]	H_2L	Δ[b]
N-C1	1.490	1.504	–0.014
N-C3	1.481	1.488	–0.007
C1-C2	1.517	1.549	–0.032
C3-C4	1.516	1.516	0.000
C4-O1	1.206	1.210	–0.004
C2-O2	1.261	1.249	0.012
C4-O3	1.303	1.327	–0.024
C2-O4	1.241	1.255	–0.014
C1-N-C3	113.9	113.3	0.6
N-C1-C2	110.8	110.7	0.1
N-C3-C4	111.4	110.8	0.6
C3-C4-O1	121.8	123.8	–2.0
C1-C2-O2	116.2	115.2	1.0
C3-C4-O3	112.4	110.6	1.8
C1-C2-O4	118.6	115.9	2.7

[a]) Average bond lengths and angles of crystal structures obtained from the CSD
[b]) Δ = (experimental – computed) value

Table B2: Comparison of experimental (CSD) and computed at the RB3LYP/6-311+G(d,p) level of theory in conjunction with PCM/UA0 solvation model selected bond distances and angles for the H_2L^{S-c} form of MIDA. Bond lengths and angles are in Å and deg, respectively.

Atoms	CSD Data[a]	H_2L	Δ[b]
N-C1	1.493	1.502	–0.009
N-C2	1.495	1.510	–0.015
N-C3	1.484	1.496	–0.012
C2-C4	1.520	1.549	–0.029
C3-C5	1.497	1.517	–0.020
C4-O1	1.252	1.248	0.004
C5-O2	1.210	1.210	0.000
C5-O3	1.300	1.326	–0.026
C4-O4	1.247	1.255	–0.008
C1-N-C2	111.0	111.6	–0.6
C1-N-C3	112.5	111.9	0.6
C2-N-C3	111.0	110.4	0.6
N-C2-C4	110.0	111.6	–1.6
N-C3-C5	111.5	111.9	–0.4
C2-C4-O1	116.4	114.9	1.5
C3-C5-O2	123.1	123.8	–0.7
C3-C5-O3	111.4	110.5	0.9
C2-C4-O4	116.5	116.1	0.4

[a]) Average bond lengths and angles of crystal structures obtained from the CSD
[b]) Δ = (experimental – computed) value

Table B3: Comparison of experimental (CSD) and computed at the RB3LYP/6-311+G(d,p) level of theory in conjunction with PCM/UA0 solvation model selected bond distances and angles for the H_3L form of HIDA. Bond lengths and angles are in Å and deg, respectively.

Atoms	CSD[a]	H_3L	Δ[b]
N-C1	1.492	1.505	–0.013
N-C2	1.498	1.504	–0.006
N-C3	1.514	1.519	–0.005
C1-C2	1.502	1.520	–0.018
C3-C4	1.514	1.522	–0.008
C5-C6	1.501	1.521	–0.020
C4-O1	1.263	1.330	–0.067
C6-O2	1.285	1.328	–0.043
C2-O3	1.403	1.424	–0.021
C6-O4	1.230	1.206	0.024
C4-O5	1.208	1.205	0.003
C1-N-C3	112.2	111.7	0.5
C1-N-C5	111.5	111.0	0.5
C3-N-C5	111.7	112.5	–0.8
N-C1-C2	111.6	110.3	1.3
N-C3-C4	114.7	113.6	1.1
N-C5-C6	110.6	114.1	–3.5
C1-C2-O3	117.1	112.1	5.0
C3-C4-O1	115.0	112.2	2.8
C5-C6-O2	106.9	106.5	0.4
C3-C4-O5	117.2	121.7	–4.5
C5-C6-O4	119.2	121.6	–2.4

[a]) Average bond lengths and angles of two related crystal structures obtained from the CSD [71]
[b]) Δ = (experimental – computed) value

Table B4: ZPVE-corrected minimum (E_{min})and Gibbs free (G_{aq}) energies of all protonated forms of NTA and reference molecules obtained from full energy optimization in solvent (water) at the RB3LYP/6-311+G(d,p) level of theory in conjunction with the PCM/UA0 solvation model.

Species	E_{min} [a]	G_{aq} [a]
NTA		
L^{3-}	–738.936972	–738.978069
HL^{2-}	–739.410772	–739.451765
H_2L^-	–739.852748	–739.893949
H_3L	–740.290659	–740.332226
H_3L^*	–740.293694	–740.335141
H_4L^+	–740.722681	–740.764545
IDA		
L^{2-}	–511.483326	–511.519032
HL^-	–511.945458	–511.981098
H_2L	–512.387398	–512.423365
H_3L^+	–512.825900	–512.862287
MIDA		
L^{2-}	–550.762444	–550.799521
HL^-	–551.227947	–551.264704
H_2L	–551.669760	–551.706615
H_3L^+	–552.107290	–552.144445
EIDA		
L^{2-}	–590.058648	–590.097630
HL^-	–590.525520	–590.564258
H_2L	–590.965018	–591.003885
H_3L^+	–591.400337	–591.439670
PIDA		
L^{2-}	–629.354940	–629.396123
HL^-	–629.822550	–629.863327
H_2L	–630.260963	–630.302194
H_3L^+	–630.694914	–630.736248
HIDA		
L^{2-}	–665.299655	–665.340516
HL^-	–665.767828	–665.808578
H_2L	–666.207584	–666.248534
H_3L^+	–666.642422	–666.683405

[a]) In atomic unit, Hartree (1 Hartree = 627.5095 kcal mol^{-1})

Table B5: Experimental stepwise protonation constants of IDA, MIDA, EIDA, PIDA, and HIDA at μ = 0.0 and 0.1 M and 25 ^{o}C.

Reaction	Exp[a]	Exp[b]
L = IDA		
$L^{2-} + H^{+} = HL^{-}$	9.79	9.34
$HL^{-} + H^{+} = H_2L$	2.84	2.62
$H_2L + H^{+} = H_3L^{+}$	1.85	1.77
L = MIDA		
$L^{2-} + H^{+} = HL^{-}$	10.01	9.59
$HL^{-} + H^{+} = H_2L$	2.59	2.32
$H_2L + H^{+} = H_3L^{+}$	–	1.90
L = EIDA		
$L^{2-} + H^{+} = HL^{-}$	10.13	9.95
$HL^{-} + H^{+} = H_2L$	2.70	2.22
$H_2L + H^{+} = H_3L^{+}$	–	1.60
L = PIDA		
$L^{2-} + H^{+} = HL^{-}$	10.43	10.05
$HL^{-} + H^{+} = H_2L$	2.49	2.24
$H_2L + H^{+} = H_3L^{+}$	1.10	1.50
L = HIDA		
$L^{2-} + H^{+} = HL^{-}$	–	8.68
$HL^{-} + H^{+} = H_2L$	–	2.20
$H_2L + H^{+} = H_3L^{+}$	–	1.60[c]

[a]) μ = 0.0 M and 25 ^{o}C, [b]) μ = 0.1 M and 25 ^{o}C, [c]) μ = 0.1 M and 20 ^{o}C

Table B6: Comparison of experimental (Exp) and calculated stepwise protonation constants of NTA ($L_{(1)}$), as log K_H, using protonation constants of the reference molecules (a) IDA, (b) MIDA, (c) EIDA, (d) PIDA and (e) HIDA at μ = 0.0 or 0.1 M and 25 °C. All energies are reported in kcal/mol.

(a)

Reaction	ΔG_{IRn}(aq)	$\Delta G_{p_{L_{(1)}}}$(aq)	log K_H	Exp[a]	δ	$\Delta G_{p_{L_{(1)}}}$(aq)	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
$L_{(1)}^{3-} + H_2L_{(2)} = HL_{(1)}^{2-} + HL_{(2)}^{-}$	–19.722	–23.596	17.30	10.334	6.96	–23.296	17.08	9.66	7.42
$L_{(1)}^{3-} + H_3L_{(2)}^{+} = HL_{(1)}^{2-} + H_2L_{(2)}$	–21.821	–24.345	17.84	10.334	7.51	–24.236	17.76	9.66	8.10
$HL_{(1)}^{2-} + HL_{(2)}^{-} = H_2L_{(1)}^{-} + L_{(2)}^{2-}$	12.476	–0.880	0.64	2.94	–2.30	–0.266	0.19	2.52	–2.33
$HL_{(1)}^{2-} + H_2L_{(2)} = H_2L_{(1)}^{-} + HL_{(2)}^{-}$	0.052	–3.822	2.80	2.94	0.14	–3.522	2.58	2.52	0.06
$HL_{(1)}^{2-} + H_3L_{(2)}^{+} = H_2L_{(1)}^{-} + H_2L_{(2)}$	–2.047	–4.571	3.35	2.94	0.41	–4.462	3.27	2.52	0.75
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)} + L_{(2)}^{2-}$	14.928	1.572	–1.15	2.00[c]	–3.15	2.186	–1.60	1.81	–3.41
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)} + HL_{(2)}^{-}$	2.504	–1.371	1.00	2.00[c]	1.00	–1.071	0.78	1.81	–1.03
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} = H_3L_{(1)} + H_2L_{(2)}$	0.405	–2.119	1.55	2.00[c]	–0.45	–2.010	1.47	1.81	–0.34
$H_3L_{(1)} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L_{(2)}^{2-}$	18.667	5.310	–3.89	1.00[b]	–4.89	5.924	-4.34	1.00	–5.34
$H_3L_{(1)} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL_{(2)}^{-}$	6.242	2.368	–1.74	1.00[b]	–2.74	2.668	–1.96	1.00	–2.96
$H_3L_{(1)} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L_{(2)}$	4.143	1.620	–1.19	1.00[b]	–2.19	1.729	–1.27	1.00	–2.27
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)}^{*} + L_{(2)}^{2-}$	13.099	–0.257	0.19	2.00[c]	–1.81	0.356	–0.26	1.81	–2.07
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)}^{*} + HL_{(2)}^{-}$	0.675	–3.200	2.35	2.00[c]	0.35	–2.900	2.13	1.81	0.32
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} = H_3L_{(1)}^{*} + H_2L_{(2)}$	–1.424	–3.948	2.89	2.00[c]	0.89	–3.839	2.81	1.81	1.00
$H_3L_{(1)}^{*} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L_{(2)}^{2-}$	20.496	7.140	–5.23	1.00[b]	–6.23	7.754	–5.68	1.00	–6.68
$H_3L_{(1)}^{*} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL_{(2)}^{-}$	8.072	4.197	–3.08	1.00[b]	–4.08	4.497	–3.30	1.00	–4.30
$H_3L_{(1)}^{*} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L_{(2)}$	5.973	3.449	–2.53	1.00[b]	–3.53	3.558	–2.61	1.00	–3.61

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

(b)

Reaction	ΔG_{IRn}(aq)	$\Delta G_{p_{L_{(1)}}}$(aq)	log K_H	Exp[a]	δ	$\Delta G_{p_{L_{(1)}}}$(aq)	log K_H	Exp[b]	δ
$L_{(2)}$ = 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
$L_{(1)}^{3-} + H_2L_{(2)} = HL_{(1)}^{2-} + HL_{(2)}^{-}$	–19.945	–23.479	17.21	10.334	6.88	–23.110	16.94	9.66	7.28
$L_{(1)}^{3-} + H_3L_{(2)}^{+} = HL_{(1)}^{2-} + H_2L_{(2)}$	–22.506	–25.098	18.40	10.334	8.06	–25.098	18.40	9.66	8.74
$HL_{(1)}^{2-} + HL_{(2)}^{-} = H_2L_{(1)}^{-} + L_{(2)}^{2-}$	14.432	0.776	–0.57	2.94	–3.51	1.349	–0.99	2.52	–3.51
$HL_{(1)}^{2-} + H_2L_{(2)} = H_2L_{(1)}^{-} + HL_{(2)}^{-}$	–0.171	–3.705	2.72	2.94	0.22	–3.336	2.45	2.52	–0.07
$HL_{(1)}^{2-} + H_3L_{(2)}^{+} = H_2L_{(1)}^{-} + H_2L_{(2)}$	–2.732	–5.324	3.90	2.94	0.96	–5.324	3.90	2.52	1.38
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)} + L_{(2)}^{2-}$	16.884	3.228	–2.37	2.00[c]	–4.37	3.801	–2.79	1.81	–4.60
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)} + HL_{(2)}^{-}$	2.280	–1.253	0.92	2.00[c]	–1.08	–0.885	0.65	1.81	–1.16
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} = H_3L_{(1)} + H_2L_{(2)}$	–0.280	–2.873	2.11	2.00[c]	0.11	–2.873	2.11	1.81	0.30
$H_3L_{(1)} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L_{(2)}^{2-}$	20.622	6.966	–5.11	1.00[b]	–6.11	7.539	–5.53	1.00	–6.53
$H_3L_{(1)} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL_{(2)}^{-}$	6.019	2.486	–1.82	1.00[b]	–2.82	2.854	–2.09	1.00	–3.09
$H_3L_{(1)} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L_{(2)}$	3.458	0.866	–0.63	1.00[b]	–1.63	0.866	–0.63	1.00	–1.63
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)}^{*} + L_{(2)}^{2-}$	15.055	1.398	–1.03	2.00[b]	–3.03	1.971	–1.45	1.81	–3.26
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)}^{*} + HL_{(2)}^{-}$	0.451	–3.082	2.26	2.00[c]	0.26	–2.714	1.99	1.81	0.18
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} = H_3L_{(1)}^{*} + H_2L_{(2)}$	–2.110	–4.702	3.45	2.00[c]	1.45	–4.702	3.45	1.81	1.64
$H_3L_{(1)}^{*} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L_{(2)}^{2-}$	22.452	8.795	–6.45	1.00[b]	–7.45	9.368	–6.87	1.00	–7.87
$H_3L_{(1)}^{*} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL_{(2)}^{-}$	7.848	4.315	–3.16	1.00[b]	–4.16	4.683	–3.43	1.00	–4.43
$H_3L_{(1)}^{*} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L_{(2)}$	5.287	2.695	–1.98	1.00[b]	–2.98	2.695	–1.98	1.00	–2.98

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

(c)

Reaction	ΔG_{IRn}(aq)	$\Delta G_{p_{L_{(1)}}}$(aq)	log K_H	Exp[a]	δ	$\Delta G_{p_{L_{(1)}}}$(aq)	log K_H	Exp[b]	δ
$L_{(2)}$ = 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
$L_{(1)}^{3-} + H_2L_{(2)} = HL_{(1)}^{2-} + HL_{(2)}^{-}$	–21.379	–25.062	18.37	10.334	8.04	–24.407	17.89	9.66	8.23
$L_{(1)}^{3-} + H_3L_{(2)}^{+} = HL_{(1)}^{2-} + H_2L_{(2)}$	–23.790	–25.972	19.04	10.334	8.70	–25.972	19.04	9.66	9.38
$HL_{(1)}^{2-} + HL_{(2)}^{-} = H_2L_{(1)}^{-} + L_{(2)}^{2-}$	15.339	1.519	–1.11	2.94	–4.05	1.765	–1.29	2.52	–3.81
$HL_{(1)}^{2-} + H_2L_{(2)} = H_2L_{(1)}^{-} + HL_{(2)}^{-}$	–1.605	–5.288	3.88	2.94	0.94	–4.633	3.40	2.52	0.88
$HL_{(1)}^{2-} + H_3L_{(2)}^{+} = H_2L_{(1)}^{-} + H_2L_{(2)}$	–4.015	–6.198	4.54	2.94	1.60	–6.198	4.54	2.52	2.02
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)} + L_{(2)}^{2-}$	17.791	3.971	–2.91	2.00[c]	–4.91	4.216	–3.09	1.81	–4.90
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)} + HL_{(2)}^{-}$	0.847	–2.836	2.08	2.00[c]	0.08	–2.182	1.60	1.81	–0.21
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} = H_3L_{(1)} + H_2L_{(2)}$	–1.564	–3.747	2.75	2.00[c]	0.75	–3.747	2.75	1.81	0.94
$H_3L_{(1)} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L_{(2)}^{2-}$	21.529	7.709	–5.65	1.00[b]	–6.65	7.955	–5.83	1.00	–6.83
$H_3L_{(1)} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL_{(2)}^{-}$	4.586	0.902	–0.66	1.00[b]	–1.66	1.557	–1.14	1.00	–2.14
$H_3L_{(1)} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L_{(2)}$	2.175	–0.008	0.01	1.00[b]	–0.99	–0.008	0.01	1.00	–0.99
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)}^{*} + L_{(2)}^{2-}$	15.961	2.141	–1.57	2.00[c]	–3.57	2.387	–1.75	1.81	–3.56
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)}^{*} + HL_{(2)}^{-}$	–0.982	–4.666	3.42	2.00[c]	1.42	–4.011	2.94	1.81	1.13
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} = H_3L_{(1)}^{*} + H_2L_{(2)}$	–3.393	–5.576	4.09	2.00[c]	2.09	–5.576	4.09	1.81	2.28
$H_3L_{(1)}^{*} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L_{(2)}^{2-}$	23.358	9.539	–6.99	1.00[b]	–7.99	9.784	–7.17	1.00	–8.17
$H_3L_{(1)}^{*} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL_{(2)}^{-}$	6.415	2.732	–2.00	1.00[b]	–3.00	3.386	–2.48	1.00	–3.48
$H_3L_{(1)}^{*} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L_{(2)}$	4.004	1.821	–1.34	1.00[b]	–2.34	1.821	–1.34	1.00	–2.34

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

(d)

Reaction	ΔG_{IRn}(aq)	$\Delta G_{p_{L_{(1)}}}$(aq)	log K_H	Exp[a]	δ	$\Delta G_{p_{L_{(1)}}}$(aq)	log K_H	Exp[b]	δ
$L_{(2)}$ = 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
$L_{(1)}^{3-} + H_2L_{(2)} = HL_{(1)}^{2-} + HL_{(2)}^{-}$	–21.856	–25.253	18.51	10.334	8.18	–24.911	18.26	9.66	8.60
$L_{(1)}^{3-} + H_3L_{(2)}^{+} = HL_{(1)}^{2-} + H_2L_{(2)}$	–24.876	–26.376	19.33	10.334	9.00	–27.290	20.00	9.66	10.34
$HL_{(1)}^{2-} + HL_{(2)}^{-} = H_2L_{(1)}^{-} + L_{(2)}^{2-}$	15.700	1.471	–1.08	2.94	–4.02	1.990	–1.46	2.52	–3.98
$HL_{(1)}^{2-} + H_2L_{(2)} = H_2L_{(1)}^{-} + HL_{(2)}^{-}$	–2.081	–5.478	4.02	2.94	1.08	–5.137	3.77	2.52	1.25
$HL_{(1)}^{2-} + H_3L_{(2)}^{+} = H_2L_{(1)}^{-} + H_2L_{(2)}$	–5.102	–6.602	4.84	2.94	1.90	–7.516	5.51	2.52	2.99
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)} + L_{(2)}^{2-}$	18.152	3.923	–2.88	2.00[c]	–4.88	4.441	–3.26	1.81	–5.07
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)} + HL_{(2)}^{-}$	0.370	–3.027	2.22	2.00[c]	0.22	–2.686	1.97	1.81	0.16
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} = H_3L_{(1)} + H_2L_{(2)}$	–2.650	–4.151	3.04	2.00[c]	1.04	–5.065	3.71	1.81	1.90
$H_3L_{(1)} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L_{(2)}^{2-}$	21.891	7.661	–5.62	1.00[b]	–6.62	8.180	–6.00	1.00	–7.00
$H_3L_{(1)} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL_{(2)}^{-}$	4.109	0.712	–0.52	1.00[b]	–1.52	1.053	–0.77	1.00	–1.77
$H_3L_{(1)} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L_{(2)}$	1.089	–0.412	0.30	1.00[b]	–0.70	-1.326	0.97	1.00	–0.03
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)}^{*} + L_{(2)}^{2-}$	16.323	2.094	–1.53	2.00[c]	–3.53	2.612	–1.91	1.81	–3.72
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)}^{*} + HL_{(2)}^{-}$	–1.459	–4.856	3.56	2.00[c]	1.56	–4.515	3.31	1.81	1.50
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} = H_3L_{(1)}^{*} + H_2L_{(2)}$	–4.479	–5.980	4.38	2.00[c]	2.38	–6.894	5.05	1.81	3.24
$H_3L_{(1)}^{*} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L_{(2)}^{2-}$	23.720	9.491	–6.96	1.00[b]	–7.96	10.009	–7.34	1.00	–8.34
$H_3L_{(1)}^{*} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL_{(2)}^{-}$	5.938	2.541	–1.86	1.00[b]	–2.86	2.882	–2.11	1.00	–3.11
$H_3L_{(1)}^{*} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L_{(2)}$	2.918	1.417	–1.04	1.00[b]	–2.04	0.503	–0.37	1.00	–1.37

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

(e)

Reaction	ΔG_{IRn}(aq)	$\Delta G_{P_{L_{(1)}}}$(aq)	log K_H	Exp[a]	δ	$\Delta G_{P_{L_{(1)}}}$(aq)	log K_H	Exp[b]	δ
$L_{(2)}$ = HIDA									
$L_{(1)}^{3-} + HL_{(2)}^{-} = HL_{(1)}^{2-} + L_{(2)}^{2-}$	–3.535	–15.377	11.27	10.334	0.94	–15.377	11.27	9.66	1.61
$L_{(1)}^{3-} + H_2L_{(2)} = HL_{(1)}^{2-} + HL_{(2)}^{-}$	–21.172	–24.174	17.72	10.334	7.39	–24.174	17.72	9.66	8.06
$L_{(1)}^{3-} + H_3L_{(2)}^{+} = HL_{(1)}^{2-} + H_2L_{(2)}$	–24.363	–26.546	19.46	10.334	9.12	–26.546	19.46	9.66	9.80
$HL_{(1)}^{2-} + HL_{(2)}^{-} = H_2L_{(1)}^{-} + L_{(2)}^{2-}$	16.239	4.397	–3.22	2.94	–6.16	4.397	–3.22	2.52	–5.74
$HL_{(1)}^{2-} + H_2L_{(2)} = H_2L_{(1)}^{-} + HL_{(2)}^{-}$	–1.398	–4.399	3.22	2.94	0.28	–4.399	3.22	2.52	0.70
$HL_{(1)}^{2-} + H_3L_{(2)}^{+} = H_2L_{(1)}^{-} + H_2L_{(2)}$	–4.589	–6.772	4.96	2.94	2.02	–6.772	4.96	2.52	2.44
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)} + L_{(2)}^{2-}$	18.690	6.849	–5.02	2.00[c]	–7.02	6.849	–5.02	1.81	–6.83
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)} + HL_{(2)}^{-}$	1.054	–1.948	1.43	2.00[c]	–0.57	–1.948	1.43	1.81	–0.38
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} = H_3L_{(1)} + H_2L_{(2)}$	–2.137	–4.320	3.17	2.00[c]	1.17	–4.320	3.17	1.81	1.36
$H_3L_{(1)} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L_{(2)}^{2-}$	22.429	10.587	–7.76	1.00[b]	–8.76	10.587	–7.76	1.00	–8.76
$H_3L_{(1)} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL_{(2)}^{-}$	4.792	1.791	–1.31	1.00[b]	–2.31	1.791	–1.31	1.00	–2.31
$H_3L_{(1)} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L_{(2)}$	1.601	–0.581	0.43	1.00[b]	–0.57	–0.581	0.43	1.00	–0.57
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)}^{*} + L_{(2)}^{2-}$	16.861	5.019	–3.68	2.00[c]	–5.68	5.019	–3.68	1.81	–5.49
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)}^{*} + HL_{(2)}^{-}$	–0.776	–3.777	2.77	2.00[c]	0.77	–3.777	2.77	1.81	0.96
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} = H_3L_{(1)}^{*} + H_2L_{(2)}$	–3.966	–6.149	4.51	2.00[c]	2.51	–6.149	4.51	1.81	2.70
$H_3L_{(1)}^{*} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L_{(2)}^{2-}$	24.258	12.417	–9.10	2.00[c]	–11.1	12.417	–9.10	1.81	–10.9
$H_3L_{(1)}^{*} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL_{(2)}^{-}$	6.621	3.620	–2.65	1.00[b]	–3.65	3.620	–2.65	1.00	–3.65
$H_3L_{(1)}^{*} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L_{(2)}$	3.431	1.248	–0.91	1.00[b]	–1.91	1.248	–0.91	1.00	–1.91

[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 B7: Part (a). Minimum energies of MM/MD conformers in solvent (E_{C1}-E_{C5}) and energies obtained from MM-based SPC performed on IDA, MIDA, EIDA and PIDA structures seen in Figures S4-S8. Part (b). DFT-calculated solvent-optimized energies (E_{min} = ZPVE-corrected energy) of structures seen in Figures S4-S8 and lowest energy MM/MD-generated C-1 conformers.

(a)

L = IDA	E_{SPC}	E_{C-1}	δE[a] kJ/mol	δE[a] kcal/mol	E_{C-2}	E_{C-3}	E_{C-4}	E_{C-5}
L^{2-}	–663.08	–676.62	13.53	3.23	–676.52	–676.52	–657.53	–657.53
HL^{-}	–1002.94	–1034.18	31.24	7.47	–1034.17	–1031.99	–	–
H_2L	–759.59	–794.46	34.88	8.34	–794.46	–793.52	–793.51	–792.31
H_3L^{+}	–483.78	–518.66	33.88	8.10	–517.64	–517.38	–514.21	–514.20
L = MIDA								
L^{2-}	–518.64	–535.72	17.08	4.08	–535.54	–535.53	–534.43	–534.42
HL^{-}	–879.42	–904.62	25.20	6.02	–903.12	–903.12	–886.29	–886.27
H_2L	–631.93	–668.97	37.04	8.85	–668.47	–668.01	–665.93	–664.53
H_3L^{+}	–349.66	–396.19	46.53	11.12	–396.19	–394.64	–392.10	–392.09
L = EIDA								
L^{2-}	–540.92	–566.19	25.27	6.04	–564.90	–564.90	–562.62	–562.62
HL^{-}	–850.31	–886.48	36.17	8.65	–886.48	–885.78	–885.77	–884.22
H_2L	–608.17	–656.82	48.65	11.63	–656.03	–655.35	–655.25	–654.89
H_3L^{+}	–327.29	–387.44	60.16	14.38	–387.44	–387.36	–387.35	–384.49
L = PIDA								
L^{2-}	–533.78	–559.38	25.59	6.12	–558.28	–558.28	–555.84	–555.84
HL^{-}	–837.97	–869.39	31.42	7.51	–869.39	–868.62	–868.62	–868.51
H_2L	–605.16	–643.47	38.31	9.16	–642.96	–642.66	–642.65	–642.10
H_3L^{+}	–332.29	–378.86	46.56	11.13	–378.84	–377.78	–377.78	–375.61

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

(b)

	Structures seen in Figures B4-B8		C-1			
L = IDA	E_{min} **(Hartree)**	G_{aq} **(Hartree)**	E_{min} **(Hartree)**	G_{aq} **(Hartree)**	**δG[b] Hartree**	**δG[b] kcal/mol**
L^{2-}	–511.483326	*–511.519032*	–511.483246	–511.518986	–0.000046	–0.03
HL^{-}	–511.945458	–511.981098	–511.945929	*–511.982547*	0.001449	0.91
H_2L	–512.387398	*–512.423365*	–512.385598	–512.422578	–0.000787	–0.49
H_3L^{+}	–512.825900	*–512.862287*	–512.821788	–512.858264	–0.004023	–2.52
L = MIDA						
L^{2-}	–550.762444	*–550.799521*	–550.762444	–550.799492	–0.000029	–0.02
HL^{-}	–551.227947	–551.264704	–551.227622	*–551.264875*	0.000171	0.11
H_2L	–551.669760	*–551.706615*	–551.666562	–551.704179	–0.002436	–1.53
H_3L^{+}	–552.107290	*–552.144445*	–552.102379	–552.139587	–0.004858	–3.05
L = EIDA						
L^{2-}	–590.058648	*–590.097630*	–590.056350	–590.095538	–0.002092	–1.31
HL^{-}	–590.525520	*–590.564258*	–590.524932	–590.563994	–0.000264	–0.17
H_2L	–590.965018	*–591.003885*	–590.964244	–591.003469	–0.000416	–0.26
H_3L^{+}	–591.400337	*–591.439670*	–591.397985	–591.437169	–0.002501	–1.57
L = PIDA						
L^{2-}	–629.354940	*–629.396123*	–629.352250	–629.393521	–0.002602	–1.63
HL^{-}	–629.822550	*–629.863327*	–629.821420	–629.862576	–0.000751	–0.47
H_2L	–630.260963	*–630.302194*	–630.260785	–630.301969	–0.000225	–0.14
H_3L^{+}	–630.694914	–630.736248	–630.697061	*–630.738256*	0.002008	1.26

[b]) $\delta G_{aq} = G_{aq}$ (structure from Fig. S4-8) – G_{aq} (C-1)

Table B8. Comparison of experimental (Exp) and calculated stepwise protonation constants of NTA, as log K_H, using lowest energy structures from Tables 6 and 7 (main paper) and Table S7 (seen in Italic) with protonation constants of the reference molecules (a) IDA, (b) MIDA, (c) EIDA, (d) PIDA and (e) HIDA at μ = 0.0 and 0.1 M and 25 °C.

(a)

Reaction	log K_H	Exp[a]	δ	log K_H	Exp[b]	δ
$L_{(2)}$ = IDA						
$L_{(1)}^{3-} + HL_{(2)}^{-} = HL_{(1)}^{2-} + L^{2-}$	14.56	10.334	–4.22	14.11	9.66	–4.45
$L_{(1)}^{3-} + H_2L_{(2)} = HL_{(1)}^{2-} + HL^{-}$	18.05	10.334	–7.71	17.83	9.66	–8.17
$L_{(1)}^{3-} + H_3L_{(2)}^{+} = HL_{(1)}^{2-} + H_2L$	17.93	10.334	–7.59	17.85	9.66	–8.19
$HL_{(1)}^{2-} + HL_{(2)}^{-} = H_2L_{(1)}^{-} + L^{2-}$	–0.02	2.94	2.96	–0.47	2.52	2.99
$HL_{(1)}^{2-} + H_2L_{(2)} = H_2L_{(1)}^{-} + HL^{-}$	3.47	2.94	–0.53	3.25	2.52	–0.73
$HL_{(1)}^{2-} + H_3L_{(2)}^{+} = H_2L_{(1)}^{-} + H_2L$	3.35	2.94	–0.41	3.27	2.52	–0.75
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)} + L^{2-}$	–1.90	2.00[c]	3.90	–2.35	1.81	4.16
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)} + HL^{-}$	1.59	2.00[c]	0.41	2.73	1.81	–0.92
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} = H_3L_{(1)} + H_2L$	1.47	2.00[c]	0.53	1.39	1.81	0.42
$H_3L_{(1)} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L^{2-}$	–4.56	1.00[b]	5.56	–5.01	1.00	6.01
$H_3L_{(1)} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL^{-}$	–1.07	1.00[b]	2.07	–1.29	1.00	2.29
$H_3L_{(1)} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L$	–1.19	1.00[b]	2.19	–1.27	1.00	2.27
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)}^{*} + L^{2-}$	–0.56	2.00[c]	2.56	–1.01	1.81	2.82
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)}^{*} + HL^{-}$	2.93	2.00[c]	–0.93	2.71	1.81	–0.90
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} = H_3L_{(1)}^{*} + H_2L$	2.81	2.00[c]	–0.81	2.73	1.81	–0.92
$H_3L_{(1)}^{*} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L^{2-}$	–5.90	1.00[b]	6.90	–6.35	1.00	7.35
$H_3L_{(1)}^{*} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL^{-}$	–2.41	1.00[b]	3.41	–2.63	1.00	3.63
$H_3L_{(1)}^{*} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L$	–2.53	1.00[b]	3.53	–2.61	1.00	3.61

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

(b)

Reaction	log K_H	Exp[a]	δ	log K_H	Exp[b]	δ
$L_{(2)}$ = MIDA						
$L_{(1)}^{3-} + HL_{(2)}^{-} = HL_{(1)}^{2-} + L^{2-}$	13.93	10.334	–3.60	13.51	9.66	–3.85
$L_{(1)}^{3-} + H_2L_{(2)} = HL_{(1)}^{2-} + HL^{-}$	17.37	10.334	–7.04	17.10	9.66	–7.44
$L_{(1)}^{3-} + H_3L_{(2)}^{+} = HL_{(1)}^{2-} + H_2L$	18.48	10.334	–8.15	18.48	9.66	–8.82
$HL_{(1)}^{2-} + HL_{(2)}^{-} = H_2L_{(1)}^{-} + L^{2-}$	–0.65	2.94	3.59	–1.07	2.52	3.59
$HL_{(1)}^{2-} + H_2L_{(2)} = H_2L_{(1)}^{-} + HL^{-}$	2.79	2.94	0.15	2.52	2.52	0.00
$HL_{(1)}^{2-} + H_3L_{(2)}^{+} = H_2L_{(1)}^{-} + H_2L$	3.90	2.94	–0.96	3.90	2.52	–1.38
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)} + L^{2-}$	–2.53	2.00[c]	4.53	–2.95	1.81	4.76
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)} + HL^{-}$	0.91	2.00[c]	1.09	0.64	1.81	1.17
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} = H_3L_{(1)} + H_2L$	2.02	2.00[c]	–0.02	2.02	1.81	–0.21
$H_3L_{(1)} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L^{2-}$	–5.18	1.00[b]	6.18	–5.60	1.00	6.60
$H_3L_{(1)} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL^{-}$	–1.74	1.00[b]	2.74	–2.01	1.00	3.01
$H_3L_{(1)} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L$	–0.63	1.00[b]	1.63	–0.63	1.00	1.63
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)}^{*} + L^{2-}$	–1.19	2.00[c]	3.19	–1.61	1.81	3.42
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)}^{*} + HL^{-}$	2.26	2.00[c]	–0.26	1.99	1.81	–0.18
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} = H_3L_{(1)}^{*} + H_2L$	3.36	2.00[c]	–1.36	3.36	1.81	–1.55
$H_3L_{(1)}^{*} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L^{2-}$	–6.53	1.00[b]	7.53	–6.95	1.00	7.95
$H_3L_{(1)}^{*} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL^{-}$	–3.08	1.00[b]	4.08	–3.35	1.00	4.35
$H_3L_{(1)}^{*} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L$	–1.98	1.00[b]	2.98	–1.98	1.00	2.98

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

(c)

Reaction	log K_H	Exp[a]	δ	log K_H	Exp[b]	δ
$L_{(2)}$ = EIDA						
$L_{(1)}^{3-} + HL_{(2)}^{-} = HL_{(1)}^{2-} + L^{2-}$	13.46	10.334	–3.13	13.28	9.66	–3.62
$L_{(1)}^{3-} + H_2L_{(2)} = HL_{(1)}^{2-} + HL^{-}$	18.45	10.334	–8.12	17.97	9.66	–8.31
$L_{(1)}^{3-} + H_3L_{(2)}^{+} = HL_{(1)}^{2-} + H_2L$	19.12	10.334	–8.79	19.12	9.66	–9.46
$HL_{(1)}^{2-} + HL_{(2)}^{-} = H_2L_{(1)}^{-} + L^{2-}$	–1.11	2.94	4.05	–1.29	2.52	3.81
$HL_{(1)}^{2-} + H_2L_{(2)} = H_2L_{(1)}^{-} + HL^{-}$	3.87	2.94	–0.93	3.39	2.52	–0.87
$HL_{(1)}^{2-} + H_3L_{(2)}^{+} = H_2L_{(1)}^{-} + H_2L$	4.54	2.94	–1.60	4.54	2.52	–2.02
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)} + L^{2-}$	–2.99	2.00[c]	4.99	–3.17	1.81	4.98
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)} + HL^{-}$	2.00	2.00[c]	0.00	1.52	1.81	0.29
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} = H_3L_{(1)} + H_2L$	2.66	2.00[c]	–0.66	2.66	1.81	–0.85
$H_3L_{(1)} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L^{2-}$	–5.65	1.00[b]	6.65	–5.83	1.00	6.83
$H_3L_{(1)} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL^{-}$	–0.66	1.00[b]	1.66	–1.14	1.00	2.14
$H_3L_{(1)} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L$	0.01	1.00[b]	0.99	0.01	1.00	0.99
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)}^{*} + L^{2-}$	–1.65	2.00[c]	3.65	–1.83	1.81	3.64
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)}^{*} + HL^{-}$	3.34	2.00[c]	–1.34	2.86	1.81	–1.05
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} = H_3L_{(1)}^{*} + H_2L$	4.00	2.00[c]	–2.00	4.00	1.81	–2.19
$H_3L_{(1)}^{*} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L^{2-}$	–6.99	1.00[b]	7.99	–7.17	1.00	8.17
$H_3L_{(1)}^{*} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL^{-}$	–2.00	1.00[b]	3.00	–2.48	1.00	3.48
$H_3L_{(1)}^{*} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L$	–1.34	1.00[b]	2.34	–1.34	1.00	2.34

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

(d)

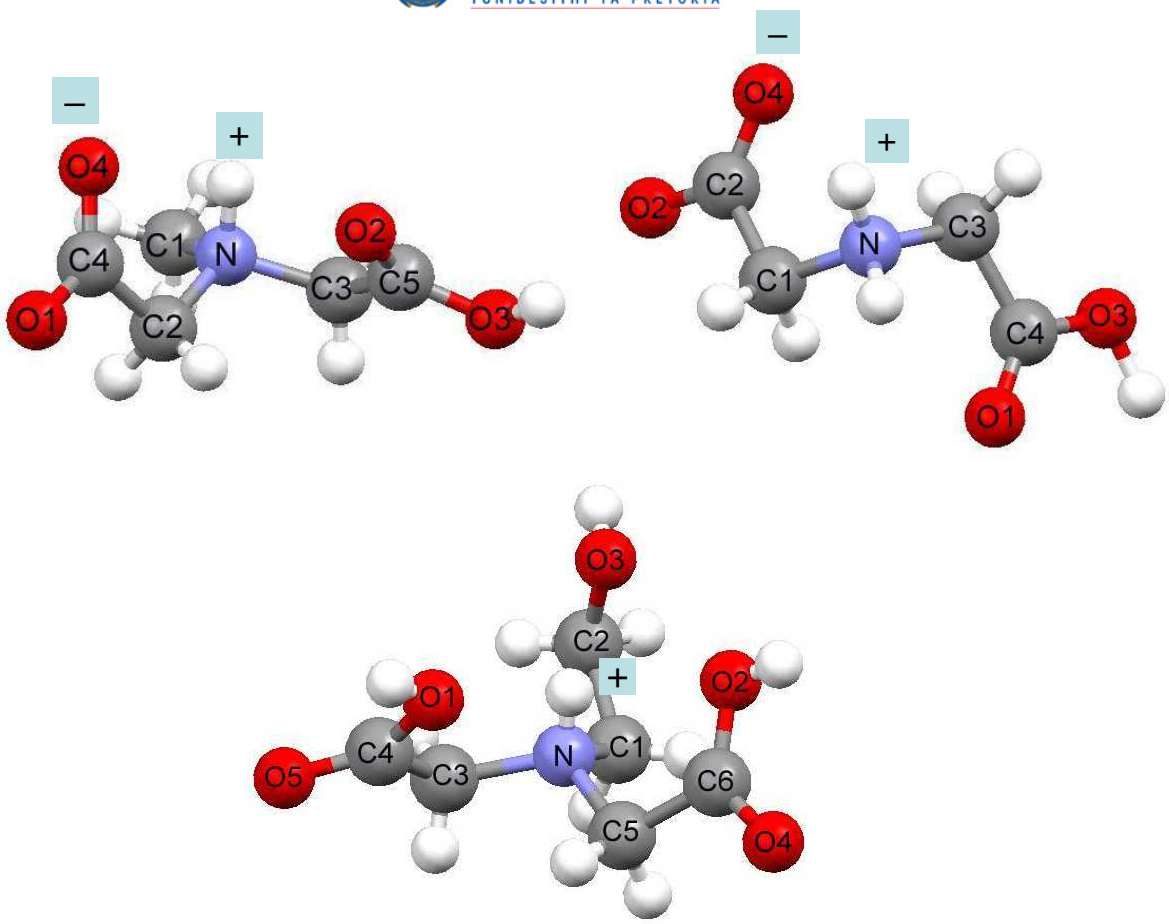
Reaction	log K_H	Exp[a]	δ	log K_H	Exp[b]	δ
$L_{(2)}$ = PIDA						
$L_{(1)}^{3-} + HL_{(2)}^{-} = HL_{(1)}^{2-} + L^{2-}$	13.50	10.334	–3.17	13.12	9.66	–3.46
$L_{(1)}^{3-} + H_2L_{(2)} = HL_{(1)}^{2-} + HL^{-}$	18.59	10.334	–8.26	18.34	9.66	–8.68
$L_{(1)}^{3-} + H_3L_{(2)}^{+} = HL_{(1)}^{2-} + H_2L$	18.49	10.334	–8.16	19.16	9.66	–9.50
$HL_{(1)}^{2-} + HL_{(2)}^{-} = H_2L_{(1)}^{-} + L^{2-}$	–1.08	2.94	4.02	–1.46	2.52	3.98
$HL_{(1)}^{2-} + H_2L_{(2)} = H_2L_{(1)}^{-} + HL^{-}$	4.01	2.94	–1.07	3.76	2.52	–1.24
$HL_{(1)}^{2-} + H_3L_{(2)}^{+} = H_2L_{(1)}^{-} + H_2L$	3.91	2.94	–0.97	4.58	2.52	–2.06
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)} + L^{2-}$	–2.96	2.00[c]	4.96	–3.34	1.81	5.15
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)} + HL^{-}$	2.14	2.00[c]	–0.14	1.89	1.81	–0.08
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} = H_3L_{(1)} + H_2L$	2.04	2.00[c]	–0.04	2.71	1.81	–0.90
$H_3L_{(1)} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L^{2-}$	–5.62	1.00[b]	6.62	–6.00	1.00	7.00
$H_3L_{(1)} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL^{-}$	–0.52	1.00[b]	1.52	–0.77	1.00	1.77
$H_3L_{(1)} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L$	–0.62	1.00[b]	1.62	0.05	1.00	0.95
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)}^{*} + L^{2-}$	–1.62	2.00[c]	3.62	–2.00	1.81	3.81
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)}^{*} + HL^{-}$	3.48	2.00[c]	–1.48	3.23	1.81	–1.42
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} = H_3L_{(1)}^{*} + H_2L$	3.38	2.00[c]	–1.38	4.05	1.81	–2.24
$H_3L_{(1)}^{*} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L^{2-}$	–6.96	1.00[b]	7.96	–7.34	1.00	8.34
$H_3L_{(1)}^{*} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL^{-}$	–1.86	1.00[b]	2.86	–2.11	1.00	3.11
$H_3L_{(1)}^{*} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L$	–1.96	1.00[b]	2.96	–1.29	1.00	2.29

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

(e)

Reaction	log K_H	Exp[a]	δ	log K_H	Exp[b]	δ
$L_{(2)}$ = HIDA						
$L_{(1)}^{3-} + HL_{(2)}^{-} = HL_{(1)}^{2-} + L^{2-}$	13.39	10.334	–3.06	13.39	9.66	–3.73
$L_{(1)}^{3-} + H_2L_{(2)} = HL_{(1)}^{2-} + HL^{-}$	17.80	10.334	–7.47	17.80	9.66	–8.14
$L_{(1)}^{3-} + H_3L_{(2)}^{+} = HL_{(1)}^{2-} + H_2L$	19.54	10.334	–9.21	19.54	9.66	–9.88
$HL_{(1)}^{2-} + HL_{(2)}^{-} = H_2L_{(1)}^{-} + L^{2-}$	–1.19	2.94	4.13	–1.19	2.52	3.71
$HL_{(1)}^{2-} + H_2L_{(2)} = H_2L_{(1)}^{-} + HL^{-}$	3.22	2.94	–0.28	3.22	2.52	–0.70
$HL_{(1)}^{2-} + H_3L_{(2)}^{+} = H_2L_{(1)}^{-} + H_2L$	4.96	2.94	–2.02	4.96	2.52	–2.44
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)} + L^{2-}$	–3.07	2.00[c]	5.07	–3.07	1.81	4.88
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)} + HL^{-}$	1.35	2.00[c]	0.65	1.35	1.81	0.46
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} = H_3L_{(1)} + H_2L$	3.08	2.00[c]	–1.08	3.08	1.81	–1.27
$H_3L_{(1)} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L^{2-}$	–5.73	1.00[b]	6.73	–5.73	1.00	6.73
$H_3L_{(1)} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL^{-}$	–1.31	1.00[b]	2.31	–1.31	1.00	2.31
$H_3L_{(1)} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L$	0.42	1.00[b]	0.58	0.42	1.00	0.58
$H_2L_{(1)}^{-} + HL_{(2)}^{-} = H_3L_{(1)}^{*} + L^{2-}$	–1.73	2.00[c]	3.73	–1.73	1.81	3.54
$H_2L_{(1)}^{-} + H_2L_{(2)} = H_3L_{(1)}^{*} + HL^{-}$	2.69	2.00[c]	–0.69	2.69	1.81	–0.88
$H_2L_{(1)}^{-} + H_3L_{(2)}^{+} = H_3L_{(1)}^{*} + H_2L$	4.42	2.00[c]	–2.42	4.42	1.81	–2.61
$H_3L_{(1)}^{*} + HL_{(2)}^{-} = H_4L_{(1)}^{+} + L^{2-}$	–7.07	1.00[b]	8.07	–7.07	1.00	8.07
$H_3L_{(1)}^{*} + H_2L_{(2)} = H_4L_{(1)}^{+} + HL^{-}$	–2.65	1.00[b]	3.65	–2.65	1.00	3.65
$H_3L_{(1)}^{*} + H_3L_{(2)}^{+} = H_4L_{(1)}^{+} + H_2L$	–0.92	1.00[b]	1.92	–0.92	1.00	1.92

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



HIDA (H_3L^+)

Figure B1: Fully labelled reported crystal structures of the H_2L forms of IDA, MIDA, and H_3L^+ form of HIDA.

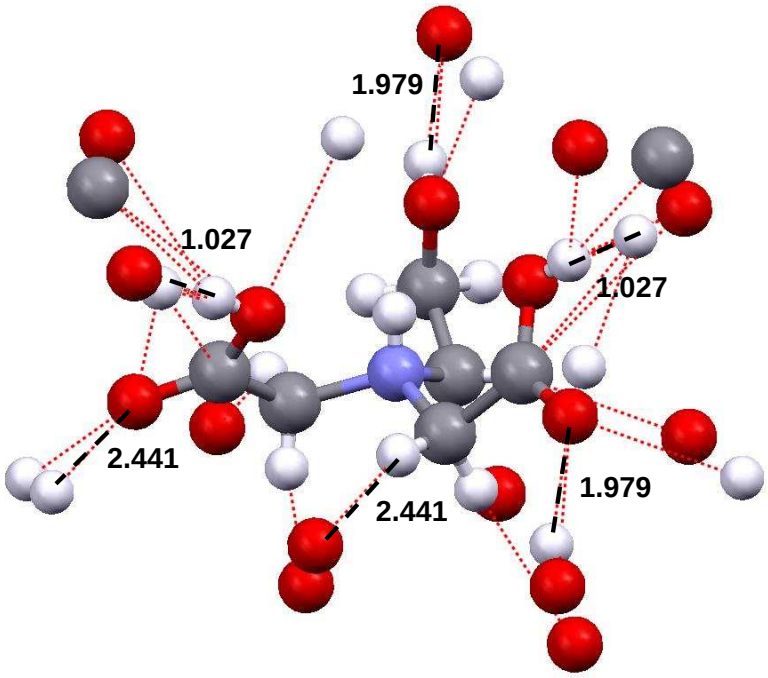
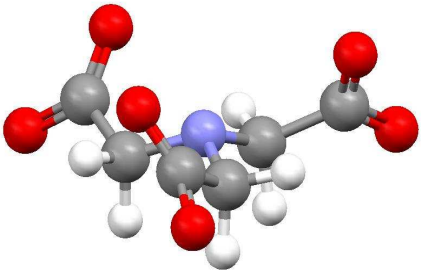
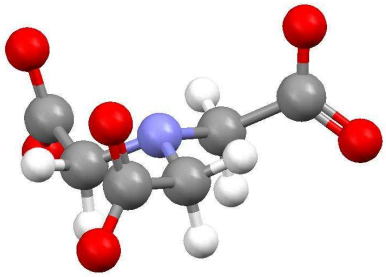


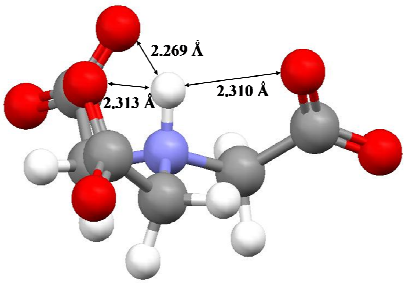
Figure B2: Crystal structure [86] of HIDA showing intermolecular short contacts (dotted red lines). For selected contacts (marked with thick dashed lines) distances are provided (in Angstrom) for illustration purposes.

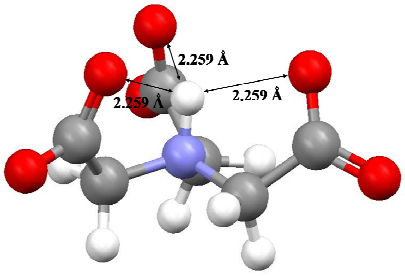
Self-constructed input structures **MM/MD-generated input structures**



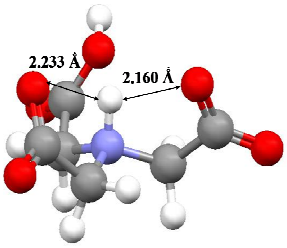


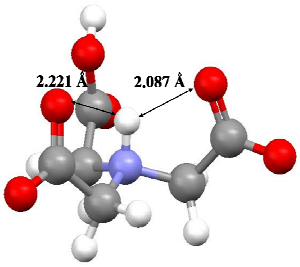
L^{3-}





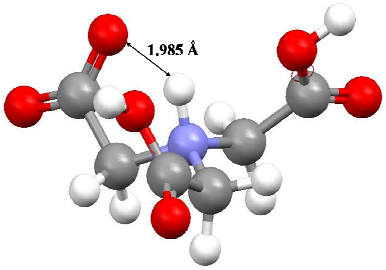
HL^{2-}

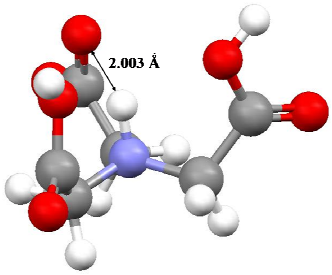




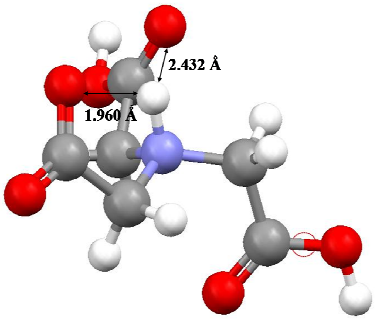
H_2L^-

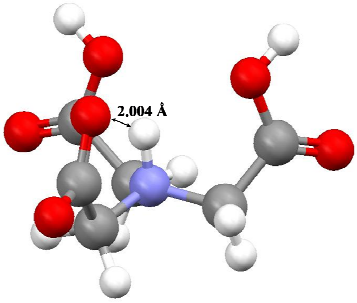
B20



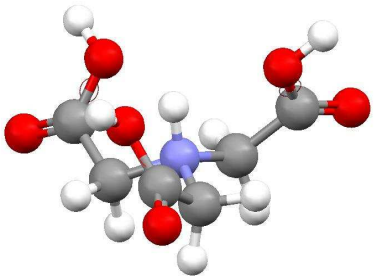


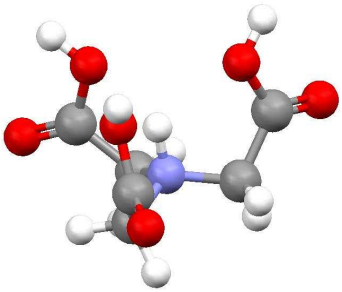
H_3L





H_3L^*

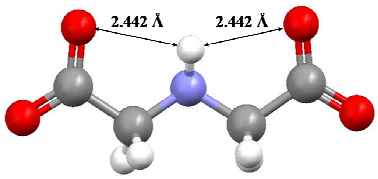


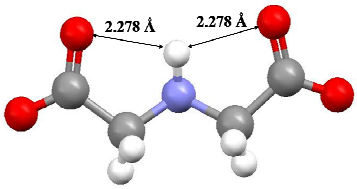


H_4L^+

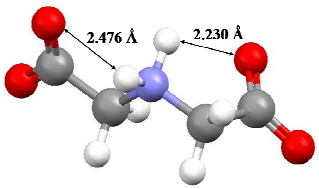
Figure B3: Energy-minimized in solvent, at the RB3LYP/6-311+G(d,p) level of theory in conjunction with the PCM/UA0 solvation model, all protonated forms of NTA.

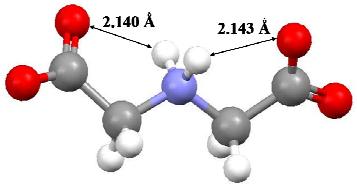
Self-constructed input structures **MM/MD-generated input structures**



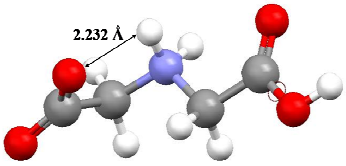


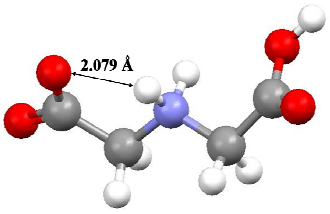
L^{2-}





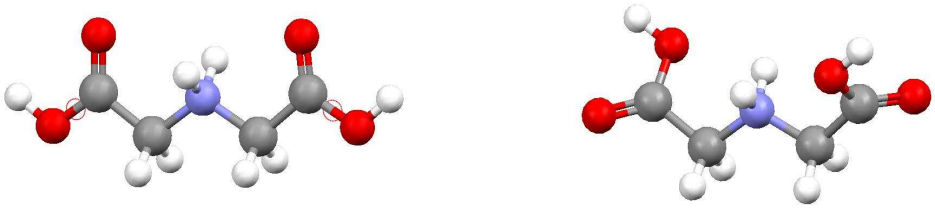
HL^{-}





H_2L

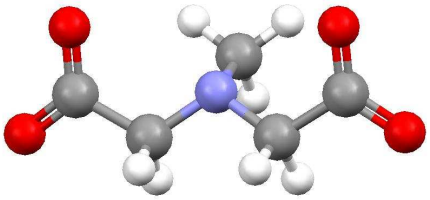
B22

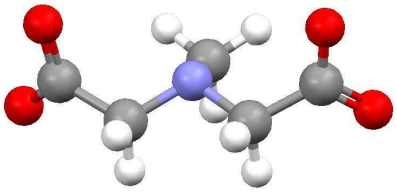


H_3L^+

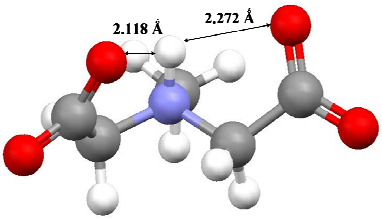
Figure B4: Energy-minimized in solvent, at the RB3LYP/6-311+G(d,p) level of theory in conjunction with the PCM/UA0 solvation model, all protonated forms of IDA.

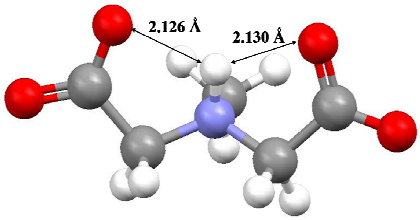
Self-constructed input structures **MM/MD-generated input structures**



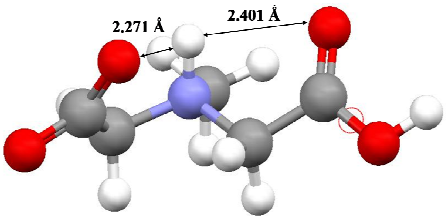


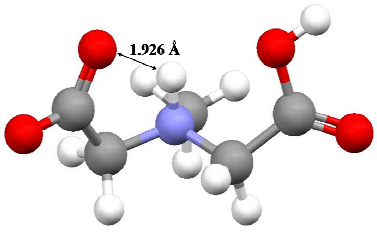
L^{2-}





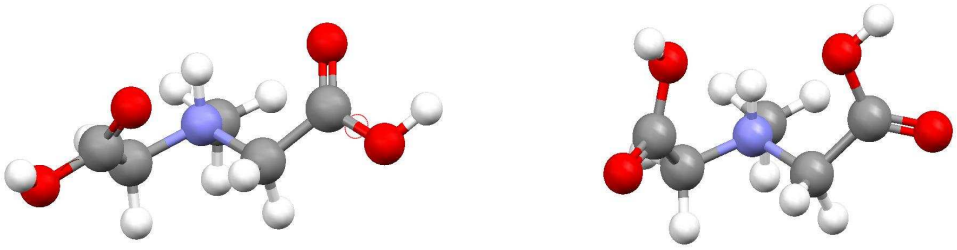
HL^{-}





H_2L

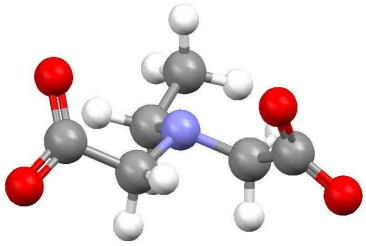
B24

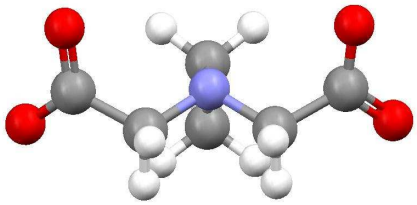


H_3L^+

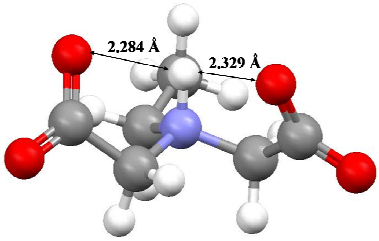
Figure B5: Energy-minimized in solvent, at the RB3LYP/6-311+G(d,p) level of theory in conjunction with the PCM/UA0 solvation model, all protonated forms of MIDA.

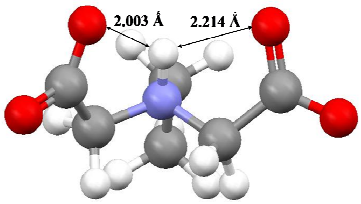
Self-constructed input structures **MM/MD-generated input structures**



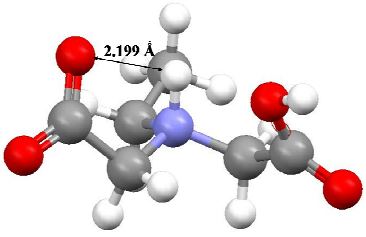


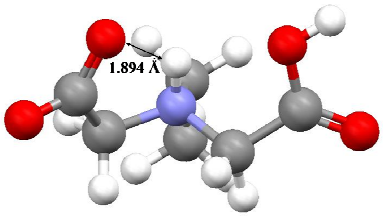
L^{2-}





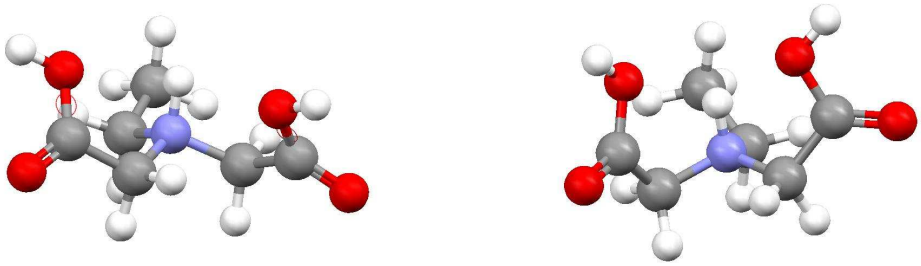
HL^{-}





H_2L

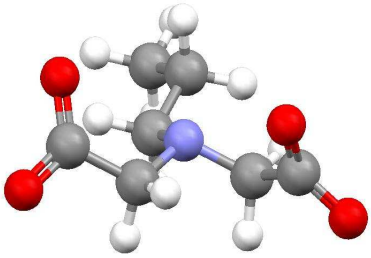
B26

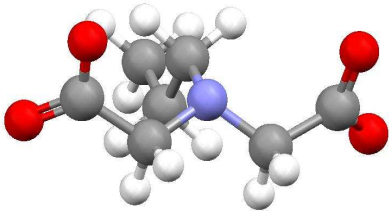


H_3L^+

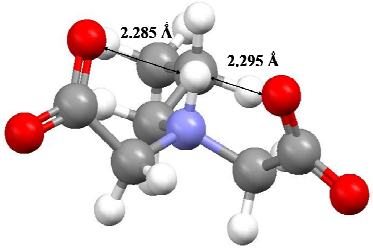
Figure B6: Energy-minimized in solvent, at the RB3LYP/6-311+G(d,p) level of theory in conjunction with the PCM/UA0 solvation model, all protonated forms of EIDA.

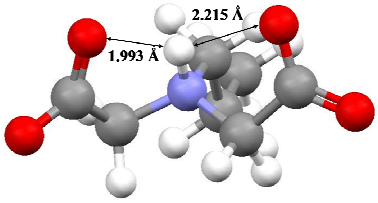
Self-constructed input structures **MM/MD-generated input structures**



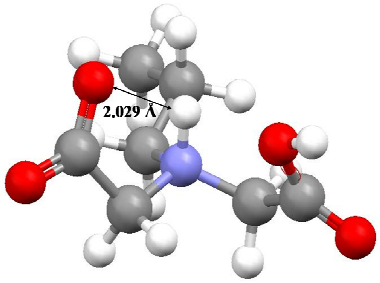


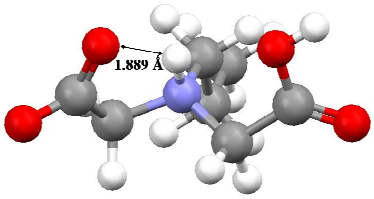
L^{2-}





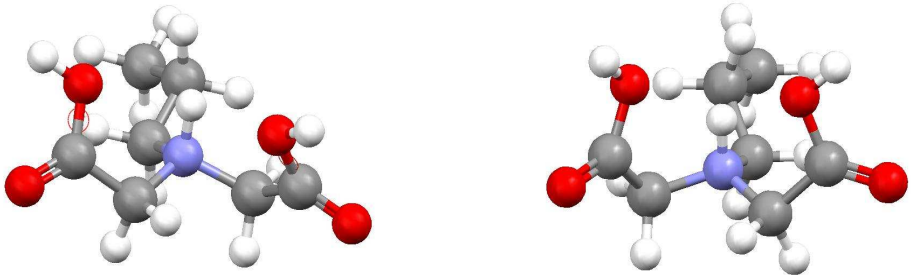
HL^{-}





H_2L

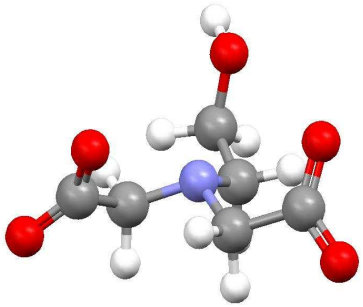
B28

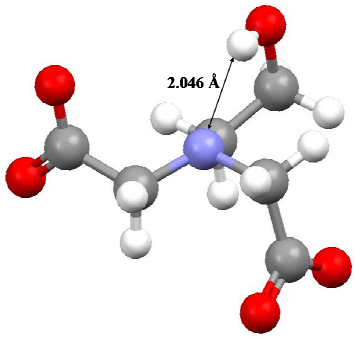


H_3L^+

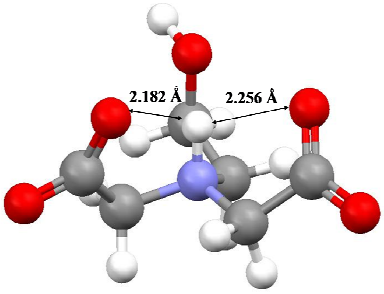
Figure B7: Energy-minimized in solvent, at the RB3LYP/6-311+G(d,p) level of theory in conjunction with the PCM/UA0 solvation model, all protonated forms of PIDA.

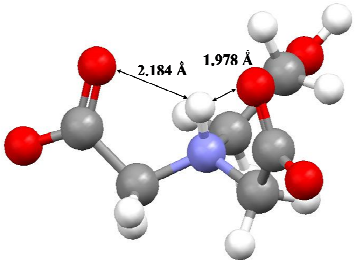
Self-constructed input structures **MM/MD-generated input structures**



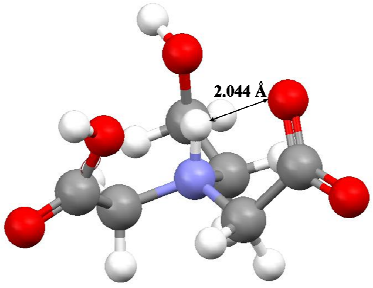


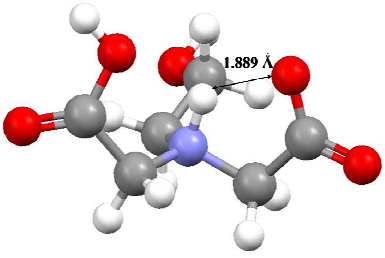
L^{2-}





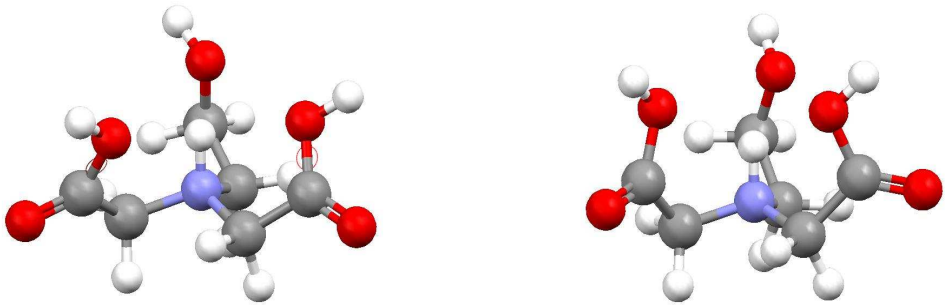
HL^{-}





H_2L

B30



H_3L^+

Figure B8: Energy-minimized in solvent, at the RB3LYP/6-311+G(d,p) level of theory in conjunction with the PCM/UA0 solvation model, all protonated forms of HIDA.

Appendix C

Table C1: Comparison of experimental (CSD) and computed at the RB3LYP/6-311+G(d,p) level of theory and PCM-UA0 solvation model selected bond distances and angles for the Ni(NTA) complex. Bond lengths and angles are in Å and °, respectively.

Atoms	CSD data[b]	Input structure	Difference (Δ)[a]
Ni-N	2.079	2.103	–0.024
Ni-O1	2.058	2.077	–0.019
Ni-O2	2.042	2.028	0.014
Ni-O3	2.084	2.074	0.010
Ni-O4	2.022	2.116	–0.094
Ni-O5	2.087	2.157	–0.070
N-Ni-O4	174.4	174.2	0.2
O1-Ni-O3	162.1	161.8	0.3
O2-Ni-O5	178.4	178.8	–0.4
N-Ni-O1	82.1	82.4	–0.3
N-Ni-O2	96.1	95.9	0.2
N-Ni-O3	84.6	85.2	–0.6
N-Ni-O5	81.0	81.6	–0.6
O1-Ni-O2	88.9	89.1	–0.2
O1-Ni-O4	90.4	98.7	–8.3
O1-Ni-O5	96.2	94.3	1.9
O2-Ni-O3	101.1	98.1	3.0
O2-Ni-O4	88.0	86.5	1.5
O3-Ni-O4	91.3	86.4	4.9
O3-Ni-O5	93.0	93.0	0.0
O4-Ni-O5	87.9	89.8	–1.9

[a]) Δ = (experimental – computed) value
[b]) Average bond lengths and angles of crystal structures obtained from the CSD

Table C2: Comparison of experimental (CSD) and computed at the RB3LYP/6-311+G(d,p) level of theory and PCM-UA0 solvation model selected bond distances and angles for the Ni(NTPA) complex. Bond lengths and angles are in Å and °, respectively.

Atoms	CSD data[b]	Input structure	Difference (Δ)[a]
Ni-N	2.066	2.128	–0.062
Ni-O1	2.095	2.039	0.056
Ni-O2	2.072	2.038	0.034
Ni-O3	2.007	2.042	–0.035
Ni-O4	2.044	2.183	–0.139
Ni-O5	2.067	2.256	–0.189
N-Ni-O4	170.1	178.8	–8.7
O1-Ni-O3	171.8	170.2	1.6
O2-Ni-O5	172.6	168.4	4.2
N-Ni-O1	95.2	95.7	–0.5
N-Ni-O2	93.2	93.9	–0.7
N-Ni-O3	92.7	92.3	0.4
N-Ni-O5	91.9	94.3	–2.4
O1-Ni-O2	92.2	86.9	5.3
O1-Ni-O4	94.0	94.3	–0.3
O1-Ni-O5	80.7	84.3	–3.6
O2-Ni-O3	80.2	83.5	–3.3
O2-Ni-O4	89.4	86.6	2.8
O3-Ni-O4	88.8	86.0	2.8
O3-Ni-O5	86.2	86.3	–0.1
O4-Ni-O5	94.8	94.8	0.0

[a]) Δ = (experimental – computed) value
[b]) Average bond lengths and angles of crystal structures obtained from the CSD

Table C3: ZPVE-corrected minimum and Gibbs free energies of NTA, NTPA, Zn(NTA), Zn(NTPA), Ni(NTA) and Ni(NTPA) MM/MD structures obtained in solvent (H_2O) at the RB3LYP/6-311+G(d,p) level of theory and PCM-UA0 solvation model

Species	E_{min}	G_{aq}
NTA^{3-}	–738.936972	–738.978069
$NTPA^{3-}$	–856.828229	–856.877193
Zn(NTA)	–2671.017279	–2671.068919
Zn(NTPA)	–2788.895663	–2788.948049
Ni(NTA)	–2399.983998	–2400.034016
Ni(NTPA)	–2517.862869	–2517.914537

All energies are reported in atomic units, Hartree (1 Hartree = 627.5095 kcal/mol)

Table C4: Comparison of experimental (Exp) and calculated stability constants of Zn(NTPA) and Ni(NTPA) MM/MD metal complexes, using stability constants of the reference molecule NTA.

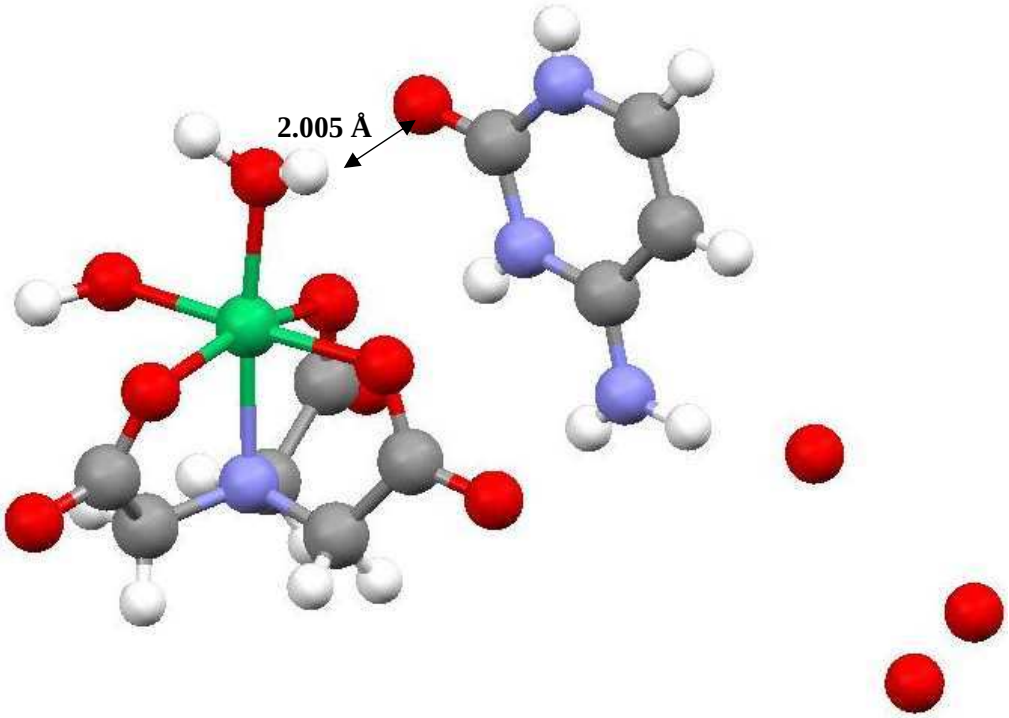
Reaction	Exp	ΔG_{aq}	ΔG_1(aq)	log K_1	δ
$[ZnL_{(2)}(H_2O)_2]^- + L_{(1)}^{3-} = [ZnL_{(1)}(H_2O)_2]^- + L_{(2)}^{3-}$	5.3	12.546	–1.712	1.25	4.05
$[NiL_{(2)}(H_2O)_2]^- + L_{(1)}^{3-} = [NiL_{(1)}(H_2O)_2]^- + L_{(2)}^{3-}$	5.8	11.675	–4.031	2.95	2.85

All energies are reported in kcal/mol
Solvation model used was PCM/UA0

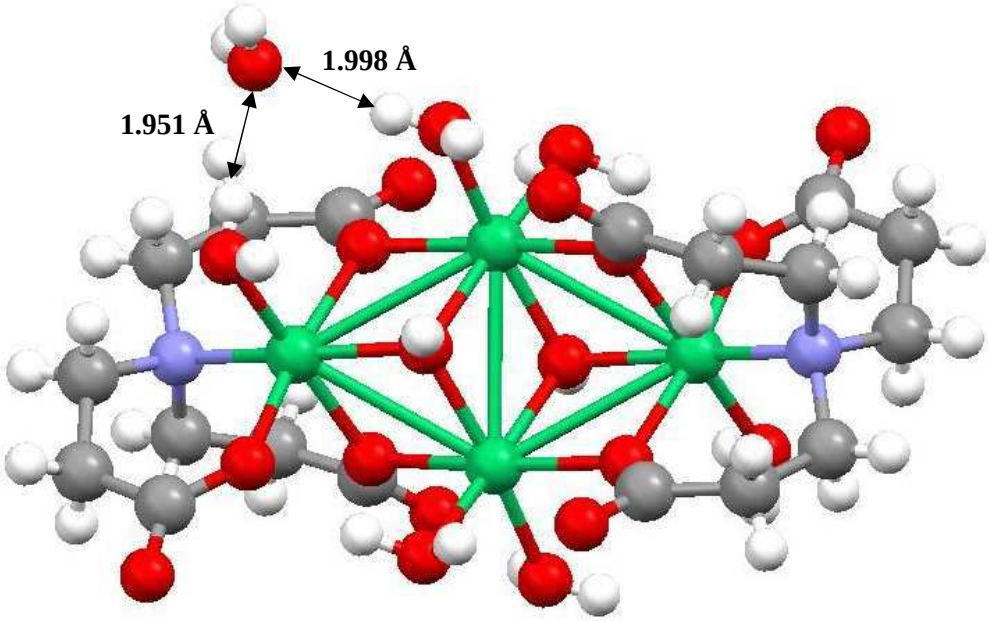
Table C5: Comparison of NBO and Bader charges obtained for the Zn^{2+} and Ni^{2+} complexes of NTA and NTPA, respectively.

	Constructed structure						MM/MD structures					
	NTA		NTPA				NTA		NTPA			
Atoms	**NBO**	**Bader**	**NBO**	**Bader**	**Δ[a]**	**Δ[b]**	**NBO**	**Bader**	**NBO**	**Bader**	**Δ[a]**	**Δ[b]**
						M = Zn^{2+}						
Zn	1.675	1.344	1.677	1.343	–0.002	0.001	1.673	1.338	1.676	1.350	–0.003	–0.012
N	–0.657	–0.941	–0.667	–0.888	0.010	–0.053	–0.658	–0.931	–0.672	–0.877	0.014	–0.054
O1	–0.876	–1.206	–0.898	–1.203	0.022	–0.003	–0.877	–1.200	–0.896	–1.207	0.019	0.007
O2	–0.875	–1.196	–0.902	–1.202	0.027	0.006	–0.875	–1.197	–0.898	–1.202	0.023	0.005
O3	–0.876	–1.197	–0.896	–1.201	0.020	0.004	–0.877	–1.192	–0.893	–1.192	0.016	0.000
O4	–1.009	–1.095	–0.984	–1.088	–0.025	–0.007	–1.008	–1.086	–0.994	–1.079	–0.014	–0.007
O5	–1.000	–1.087	–0.994	–1.068	–0.006	–0.019	–0.996	–1.090	–0.989	–1.076	–0.007	–0.014
						M = Ni^{2+}						
Ni	1.432	1.264	1.440	1.305	–0.008	–0.041	1.428	1.264	1.434	1.290	–0.006	–0.026
N	–0.613	–0.835	–0.614	–0.813	0.001	–0.022	–0.614	–0.836	–0.621	–0.817	0.007	–0.019
O1	–0.833	–1.162	–0.854	–1.157	0.021	–0.005	–0.834	–1.155	–0.850	–1.171	0.016	0.016
O2	–0.826	–1.160	–0.849	–1.157	0.023	–0.003	–0.822	–1.157	–0.849	–1.172	0.027	0.015
O3	–0.834	–1.155	–0.832	–1.173	–0.002	0.018	–0.836	–1.150	–0.843	–1.164	0.007	0.014
O4	–0.979	–1.070	–0.969	–1.044	–0.010	–0.026	–0.976	–1.048	–0.968	–1.032	–0.008	–0.016
O5	–0.980	–1.063	–0.976	–1.054	–0.004	–0.009	–0.977	–1.062	–0.972	–1.057	–0.005	–0.005

[a]) Δ = (NTA(NBO) – NTPA(NBO))
[b]) Δ = (NTA(Bader) – NTPA(Bader))

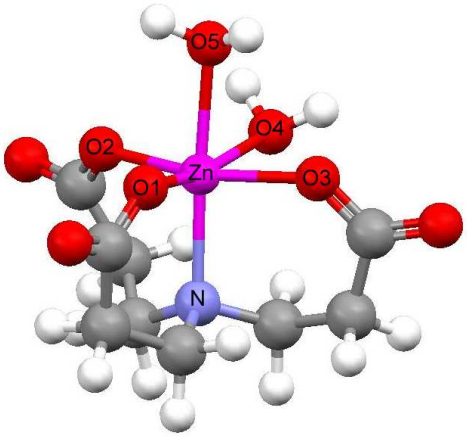


Ni(NTA)

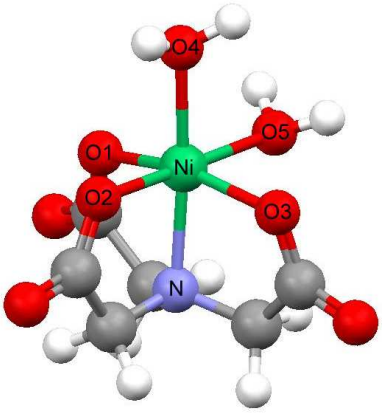


Ni(NTPA)

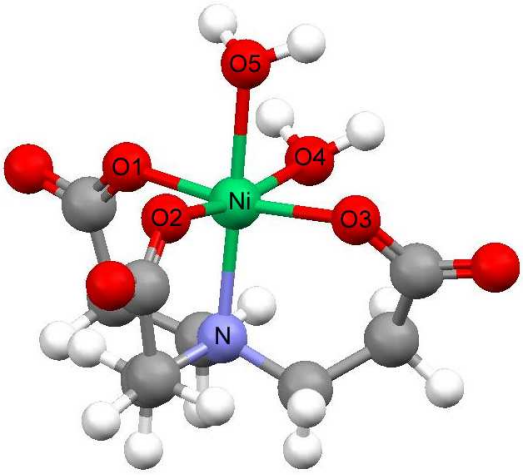
Figure C1: Crystallographic structures of Ni(NTA) and Ni(NTPA) obtained from the Cambridge Crystallographic Database.



Zn(NTPA)

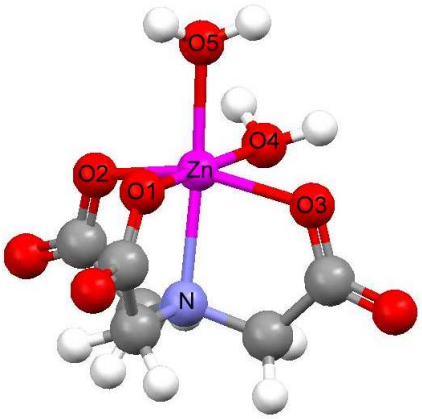


Ni(NTA)

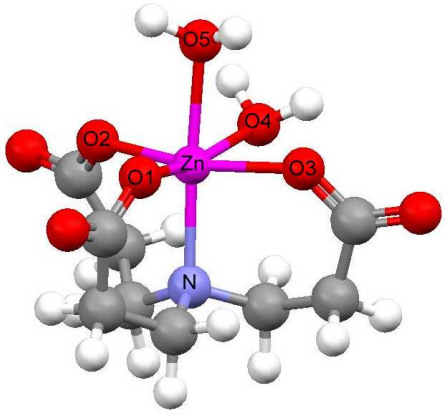


Ni(NTPA)

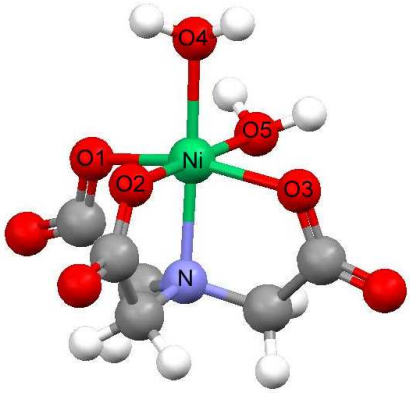
Figure C2: Partially labelled, solvent optimized at the RB3LYP/6-311+G(d,p) level of theory in combination with the PCM/UA0 solvation model, constructed structures of Zn(NTPA), Ni(NTA) and Ni(NTPA) metal complexes.



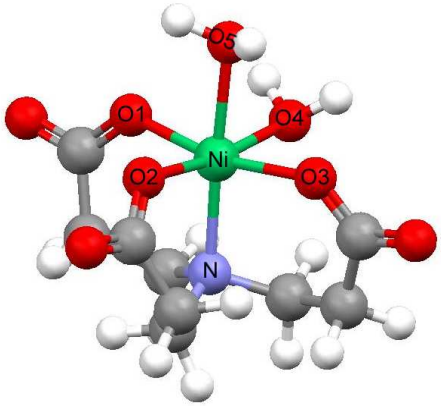
Zn(NTA)



Zn(NTPA)

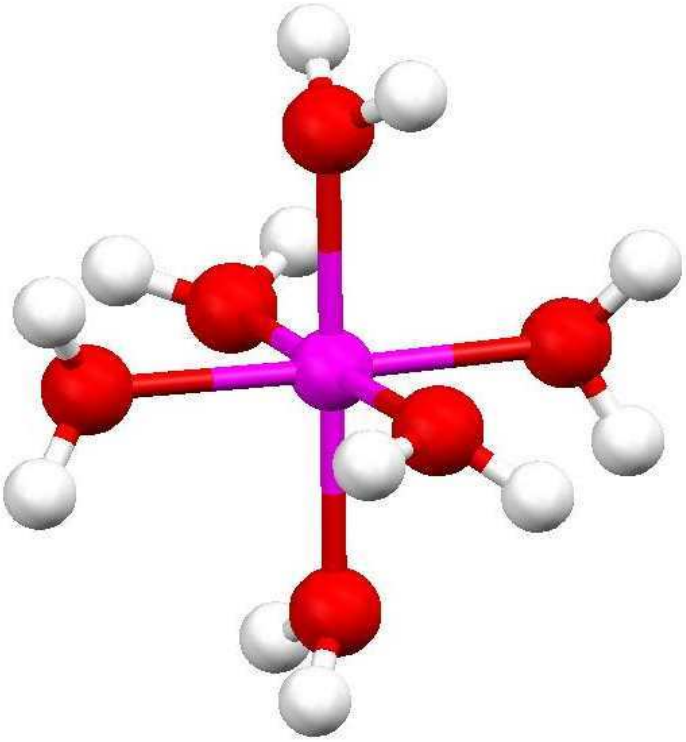


Ni(NTA)

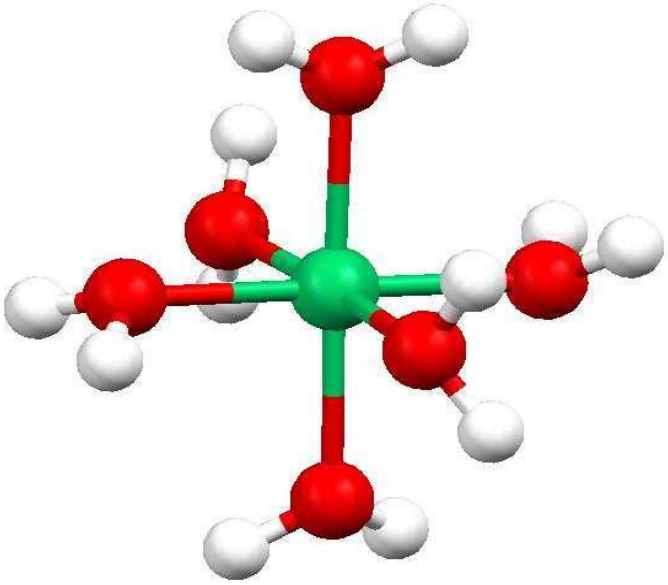


Ni(NTPA)

Figure C3: Partially labelled, solvent optimized at the RB3LYP/6-311+G(d,p) level of theory in combination with the PCM/UA0 solvation model, MM/MD-generated structures of Zn(NTA), Zn(NTPA), Ni(NTA) and Ni(NTPA).

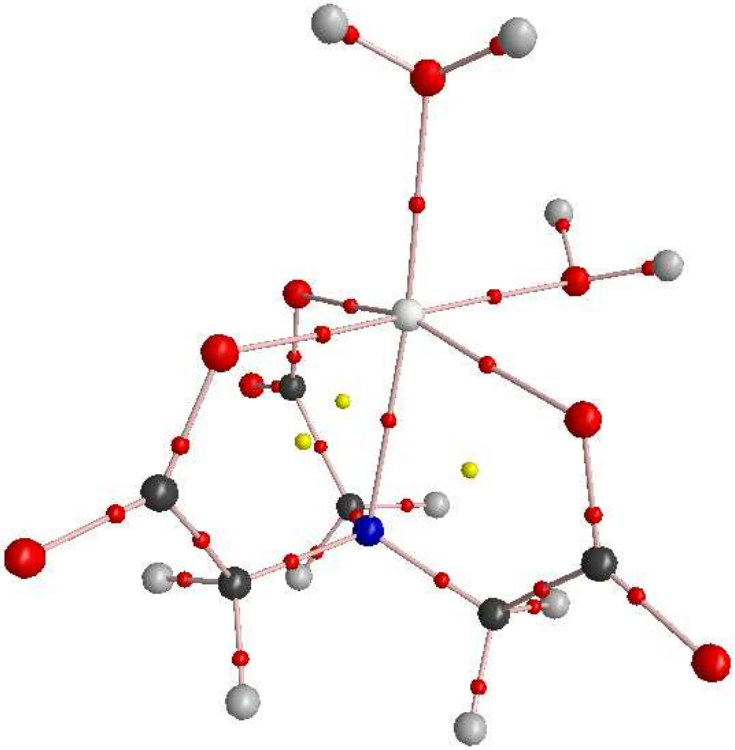


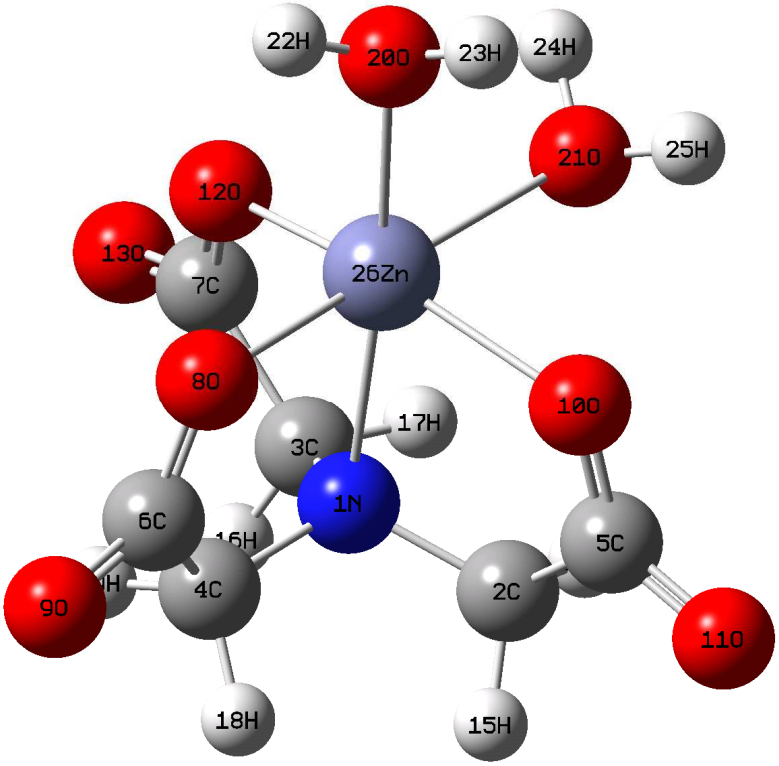
$Zn(H_2O)_6$



$Ni(H_2O)_6$

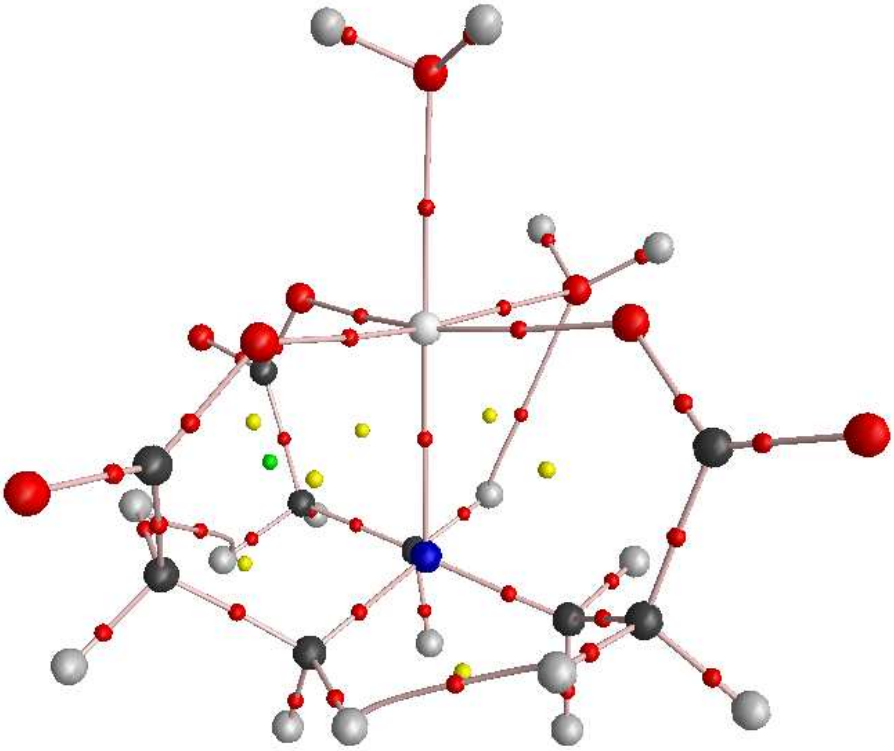
Figure C4: Structures of $Zn(H_2O)_6$ and $Ni(H_2O)_6$ that were optimized in solvent at the RB3LYP/6-311+G(d,p) level of theory in combination with the PCM/UA0 solvation model.

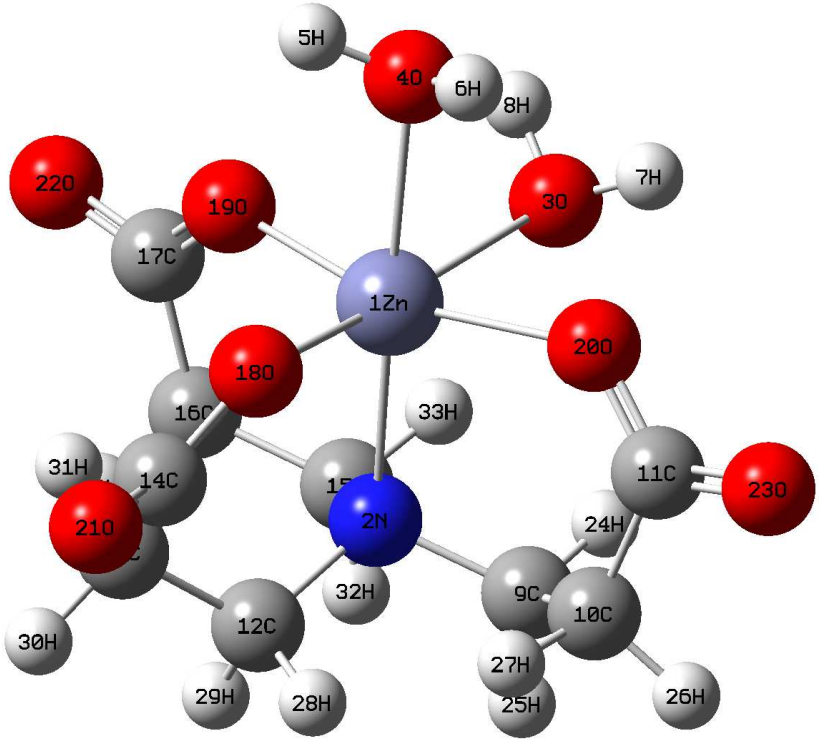



22H
20O
23H
24H
21O
25H
12O
13O
7C
26Zn
8O
10O
17H
3C
1N
6C
16H
5C
4C
2C
9O
11O
18H
15H

Zn(NTA)

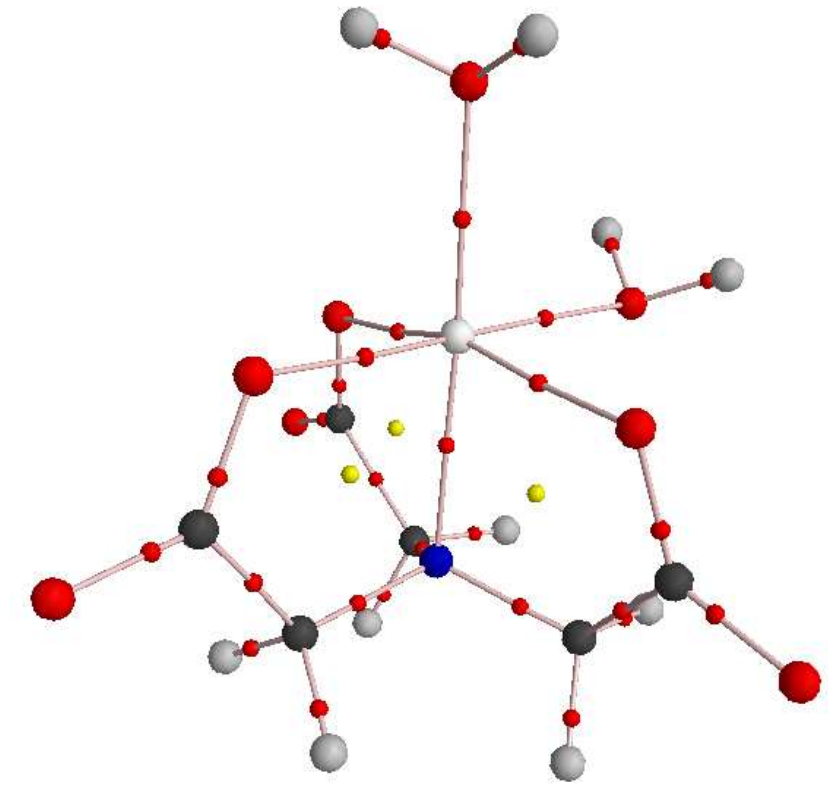
C10

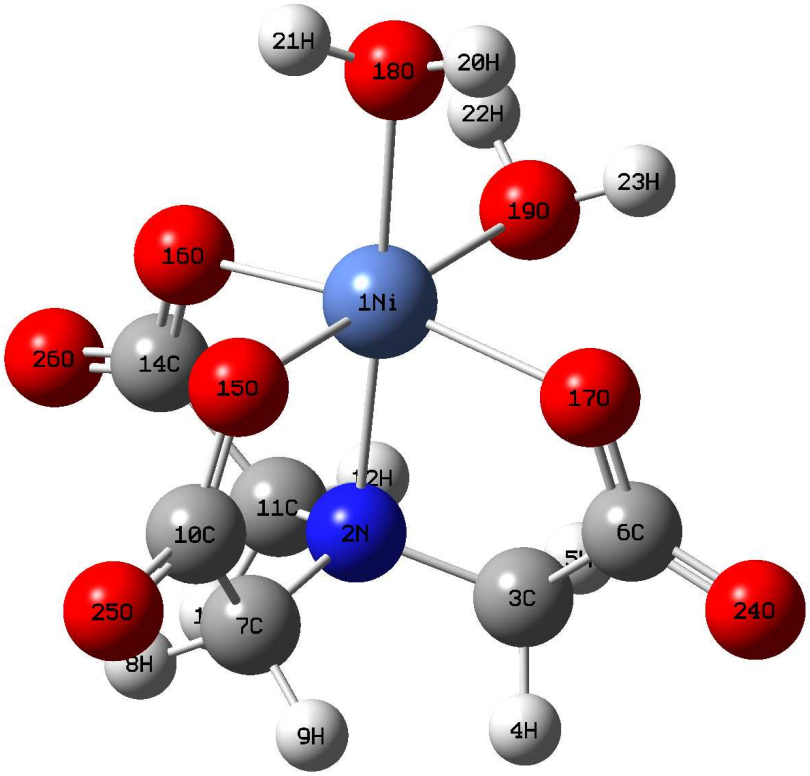



5H
40
6H
8H
220
190
30
7H
17C
1Zn
200
180
16C
33H
31H
14C
11C
230
210
2N
24H
32H
9C
10C
30H
12C
27H
29H
28H
25H
26H

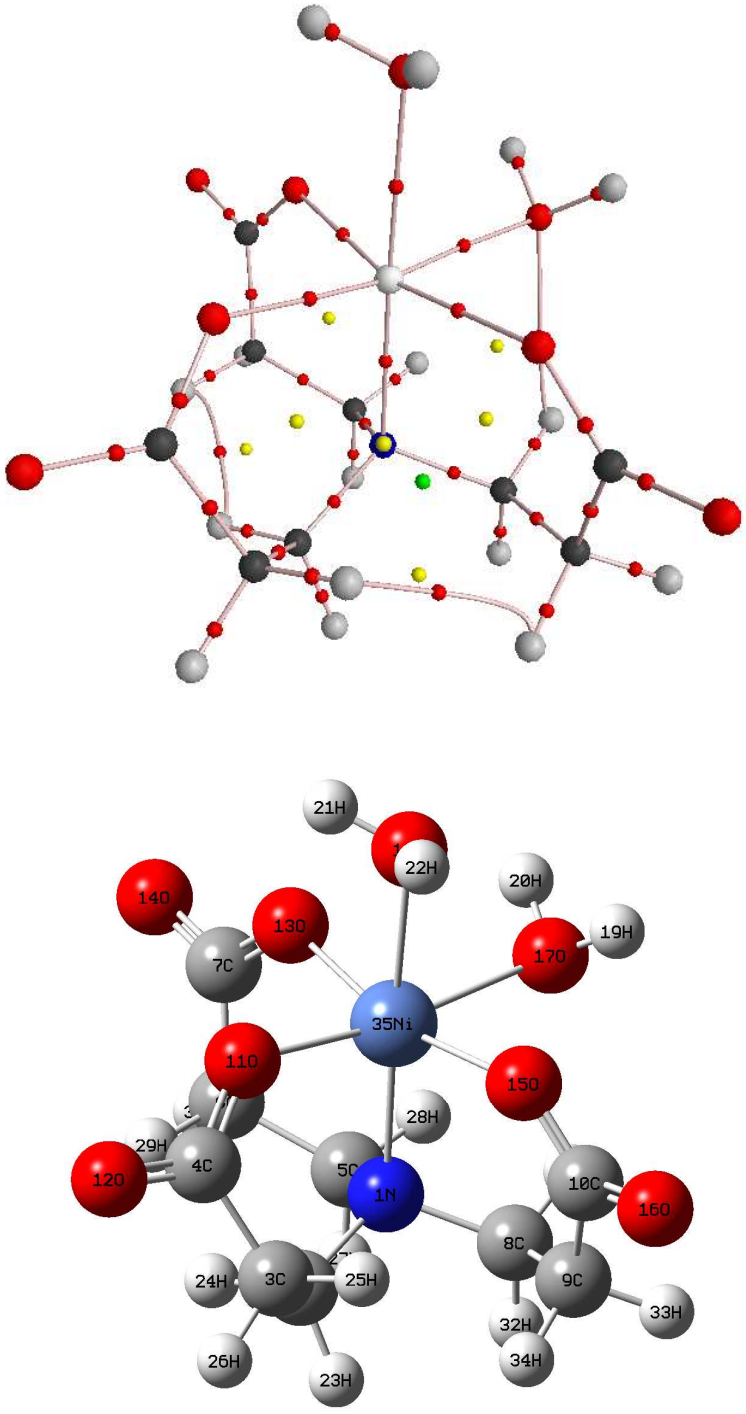
Zn(NTPA)

C11




21H
18O
20H
22H
23H
19O
16O
1Ni
26O
14C
15O
17O
12H
11C
10C
2N
6C
25O
7C
3C
24O
8H
9H
4H

Ni(NTA)



Ni(NTPA)

Figure C5: Molecular graphs and fully labelled solvent optimized MM/MD structures of Zn(NTA), Zn(NTPA), Ni(NTA) and Ni(NTPA). The bond critical points (BCPs) are denoted by red points, ring critical points (RCPs) are represented by yellow points and cage critical points (CCPs) can be seen as green points.

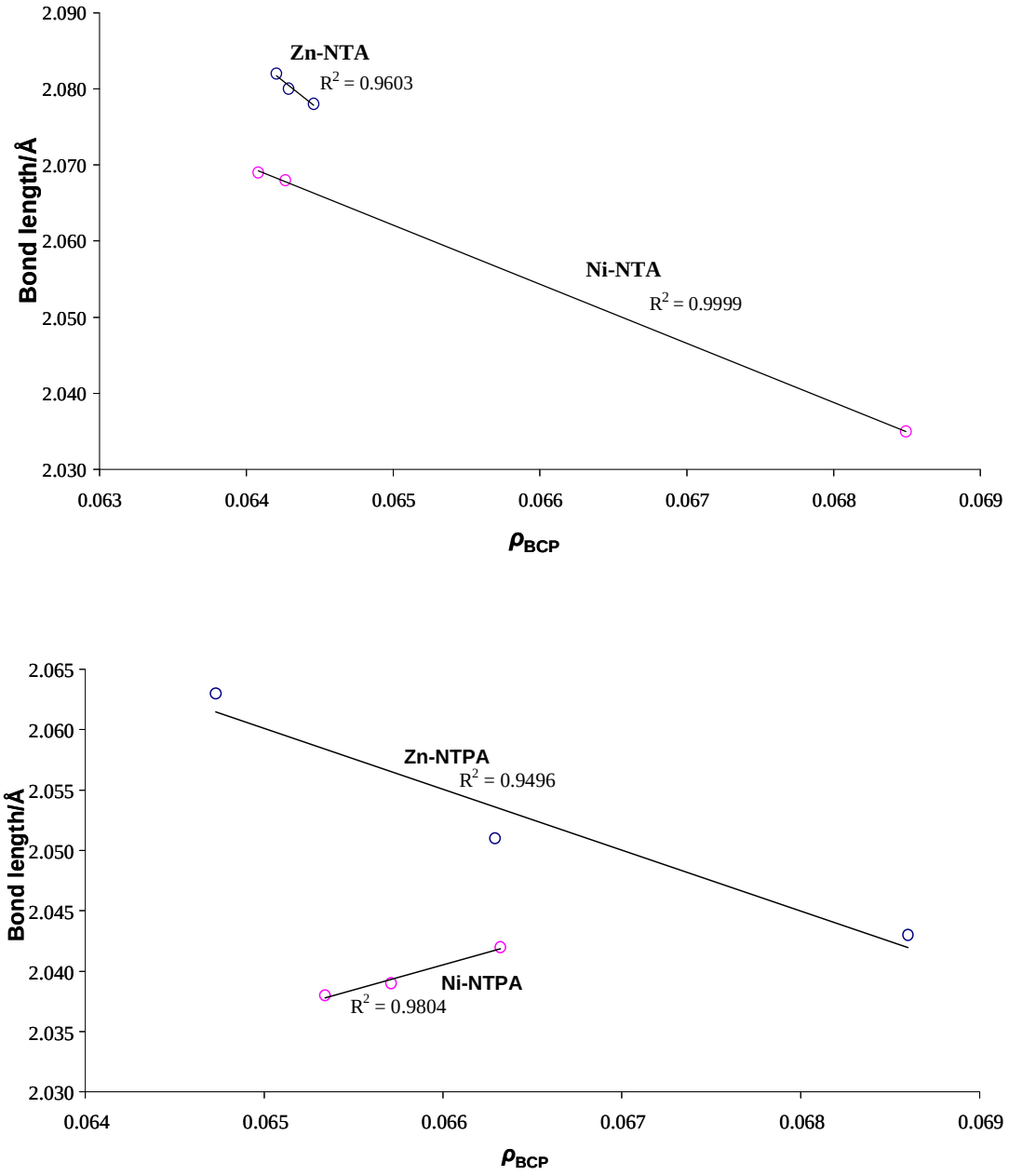


Figure C6: M-O_L Rho (ρ) versus Bond length (Å) for Zn and Ni MM/MD complexes.