Protonation sequence of aliphatic linear polyamines: A theoretical study

by

Adedapo Sunday Adeyinka

Supervisor: Prof. Ignacy Cukrowski

Department of Chemistry,

University of Pretoria

Submitted in partial fulfilment of the requirements for the degree

Doctor of Philosophy

in the Faculty of Natural and Agricultural Sciences

University of Pretoria

June, 2016

Declaration

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

_____

(Signature of candidate)

24th day of June 2016.

Outputs from this work

Publications

Adedapo S. Adeyinka, Bryan W. Bulling and Ignacy Cukrowski. Competition Reaction-Based Prediction of Polyamines' Stepwise Protonation Constants: a Case study Involving 1,4,7,10-tetraazadecane (2,2,2-tet). *Theor. Chem. Acc.*, 2016, 135:139, 1–17.

Adedapo S. Adeyinka and Ignacy Cukrowski. Structural-topological preferences and protonation sequence of Aliphatic polyamines: A theoretical case study of the tetramine trien, *J Mol. Mod.*, 2015, *21:162*, 1-18.

Ignacy Cukrowski, Jurgens H. de Lange, Adedapo S. Adeyinka and Paidamwoyo Mangondo. Evaluating common QTAIM and NCI interpretations of the electron density concentration through IQA interaction energies and 1D cross-sections of the electron and deformation density distributions, *Comput. Theor. Chem.*, 2015, *1053*, 60–76.

Conference Presentations:

Adedapo S. Adeyinka and Ignacy Cukrowski. Theoretical Prediction of the four stepwise protonation constants of 2-2-4-tet, *Virtual conference on Computational Chemistry*, University of Mauritius, August 2015. Oral presentation

Adedapo S. Adeyinka and Ignacy Cukrowski. The Stabilizing Effect of H•••H interactions in Conformers of Protonated Triethylenetetramine, *Proceedings of the South African Chemical Institute Inorganic Chemistry Conference*, University of KwaZulu-Natal, June 2013. (Poster)

Abstract

In this work, several computational techniques and protocols have been developed and applied in an attempt to resolve the longstanding controversy in literature as to the preferred protonation site i.e. primary (HL_p) or secondary (HL_s) nitrogen atom in singly protonated aliphatic linear polyamines. To achieve this aim, a dedicated conformational search protocol (CSP) for identifying relevant low energy conformers (LECs) of any given aliphatic linear polyamine from a large set of conformers initially generated using *MMFF*(aq) force field and Monte Carlo algorithm, was developed. Using representative LECs identified and a hybrid solvation model, a mixture of HL_p (84%) and HL_s (16%) was predicted for the first protonation of triethylenetetramine in excellent agreement with results obtained from the most consistent cluster analysis method. In addition, preliminary theoretical ^{13}C NMR–pH titration study also suggests that a mixture of both monoprotonated forms would exist in solution at thermodynamic equilibrium. Hence contrary to various opposing arguments in literature in favour of either the primary or secondary nitrogen atom, these results suggests strongly that both monoprotonated forms might be present in solution even though the species in which the primary nitrogen atom is protonated will be predominant due to its better solvation. Also, quantum topological methods have been utilized to investigate and understand factors responsible for conformational preference in aliphatic linear polyamines. In addition to NH•••N interactions which were mainly responsible for the conformational preference of polyamines, CH•••HC interactions were uncovered for the first time in low energy conformers of protonated triethylenetetramine (2,2,2-tet).

Furthermore, the CSP developed initially was refined to a 5-step *EEBGB*- conformer selection protocol which in principle could effectively and in shortest time possible identify low energy conformers for any given aliphatic linear polyamine (*E*, *B* and *G* stands for electronic-energy-, Boltzmann-distribution- and Gibbs-free-energy-based stepwise selection of conformers). This *EEBGB*-protocol (i) reduced (by 94%) the number of conformers subjected to the frequency calculations (to obtain *G*-values) from 420 MM-selected to 25 used to compute four stepwise protonation constants of triethylenetetramine and (ii) is of general-purpose as it is applicable to any flexible and poly-charged molecules. Combination of the 5-step *EEBGB*-conformer selection protocol and competition reaction methodology enabled us to theoretically predict for the first time, the four stepwise macroscopic protonation constants of trien within 0.1(-0.8) log unit of experimental values.

This work opens up the gateway for predicting protonation constants of yet-to-be synthesized aliphatic linear polyamines and a comparative determination of their usability as an anticancer drug template by medicinal chemists.

Acknowledgements

Firstly, my utmost gratitude goes to God almighty for His grace and mercy that has sustained me during the course of my postgraduate studies. I am also greatly indebted to my research supervisor Professor Ignacy Cukrowski for financial and moral support as well as academic guidance without which this project would not have seen the light of day. I am forever indebted to my wife Adejoke for all her sacrifices and labour of love which enabled me to complete my studies. Also, I want to appreciate my parents and siblings for the unflinching love and support they offered to me throughout the duration of my postgraduate studies. I am also very grateful to Pastor and Dr (Mrs.) Adegbenro for their hospitality and for showering me with love, support and prayers all through the duration of my studies. I also appreciate Dr. and Mrs. Eberechi for words of encouragement that spurred me on at various stages of this work. My appreciation also goes to my colleagues from the computational chemistry group (at the University of Pretoria) who have provided encouragement and support in one way or the other. To all my elders and brethren in the Lord who ceaselessly offered intercessions, encouragement and financial support along the way. Special thanks goes to my friends; Yinka Awofiranye for bearing the burden of this work with me at the crucial period of completion, and others too numerous to mention here who provided the much needed encouragement to finish this project. May God bless you all abundantly. Finally, I wish to express my gratitude to the University of Pretoria for financial support which enabled me to complete my studies.

Table of Contents

List of Figures

List of Schemes

List of Tables

List of abbreviations

HF–	Hatree–Fock
DFT–	Density Functional Theory
2,2,2-tet–	1,4,7,10-tetrazadodecane
Trien–	Triethylenetetramine
QTAIM–	Quantum Theory of Atoms in Molecules
NCI–	Non-Covalent Interaction
MM–	Molecular Mechanics
MO–	Molecular Orbitals
AO–	Atomic Orbital
LCAO–	Linear Combination of Atomic Orbitals
SCF–	Self Consistent Field
B3LYP–	Becke's three parameter hybrid exchange potential combined with Lee-Yang-Parr
QM/MM–	Quantum Mechanics/Molecular Mechanics
PCM–	Polarizable Continuum Model
COSMO–	Conductor-like Screening Solvation Model
ELF–	Electron Localization Function
ALPs–	Aliphatic Linear Polyamines
LECs–	Low energy conformers
MMFFaq–	Molecular Mechanics Force Field aqueous
LoT–	Level of Theory
MM/DFT–	Molecular Mechanics/Density Functional Theory
CSP–	Conformational Search Protocol
IQA–	Interacting Quantum Atoms
CRn–	Competition Reaction
TCs–	Thermodynamic Cycles
TMS–	Tetramethylsilane
ADF–	Amsterdam Density Functional

Chapter 1

Introduction

1.1. Proton Transfer

Proton transfer is the most widespread and important reaction in all branches of chemistry, including biochemistry.[1–8] This rudimentary reaction plays a fundamental role in a plethora of chemical processes and biological mechanisms[3,7] such as acid-base neutralization, flagellar rotation in bacteria, and the generation of active forms of enzymes and metalloenzymes.[2] In addition, it is a well-known fact that the proton acts as a promoter in chemical reactions occurring in solution. For example in medicinal chemistry, the ability of a drug to pass through biological membranes as well as its potential to interact with intracellular receptors is dependent on its readiness to exchange protons with other molecules within the biological environment.[9]

Generally, the ability of a polyprotic molecule to accept proton(s) is described by its macroscopic protonation constant(s) data.[10–12] This macroscopic picture provides information about the overall ability of such molecule to bind protons at each stage of protonation.[11] However, information about the proton binding ability of individual basic sites within it can only be obtained from its microscopic protonation constants. Having obtained the microscopic protonation constants, one can then deduce its protonation sequence i.e. pattern of protonation. Protonation sequence reveals the order of protonation of basic sites in a molecule which goes a long way to determine its reactivity in living systems. Therefore understanding the protonation sequence of a molecule is crucial since several biochemical functions almost certainly depend on the charge distribution of partially protonated species.[13]

If there is a significant difference in the basicity of individual base sites of a polybasic molecule, they are protonated sequentially, with each site being essentially fully protonated before the next is protonated significantly. In some molecules however, all the basic sites are of comparable basicity. Consequently, at each stage of stepwise protonation, a number of base sites are protonated simultaneously to a significant degree. In such cases, determination of protonation sequence obviously requires an experimental measure of the extent of protonation of each base site that is independent of the extent of protonation at other basic sites. The protonation constants of each individual protonation site (i.e. microconstants) have to be accurately determined to obtain an unambiguous description of the protonation sequence. This is indispensable for understanding of the biological role of such molecules as it would reveal the exact contribution of each protonated site at a relevant pH value.

1.2. Polyamines

Polyamines are a group of naturally occurring compounds which exert a significant number of biological effects in living organisms, both plants and animals.[14–18] They are ubiquitous

polycations with a unique structural feature of regularly spaced positive charges interspaced with hydrophobic methylene bridges and occur as metabolites in all living cells.[17] They play essential roles in cell growth, division, apoptosis and death.[14–18] These compounds also modulate cell membrane functions, interact with ion channels, and stimulate or inhibit the activity of several enzymes such as protein kinase C[1]. They stabilize nucleic acids and are responsible for initiating their replication.[2] Polyamines are polycations at physiological pH ranges; therefore they have a high affinity for cellular polyanions such as fatty acids, phospholipids, nucleotides, and nucleic acids.[14] One of the main functions of the polyamines seems to be that of acting as protective or stabilizing agents by being involved in interactions with these anions. Some of the most interesting of the interactions are those involving the various forms of nucleic acids (including DNA, rRNA and tRNA).[19] As a result of these interactions, polyamines have the ability to precipitate these macromolecules at sufficiently high concentrations. The complexes are formed by ionic interactions between the cationic amine groups of the polyamine and the anionic phosphate groups of the nucleic acid. Furthermore, it is well-known that polyamines are receptors of metal ions, anions or neutral molecules in supramolecular chemistry.[2] Polyamines are bases in aqueous solution and because of that they give rise to competition between their protonation and complexation reactions.[2] For metal-ion binding, the amino groups must be deprotonated and have available electron pairs; for anion binding, high protonation degrees and thereby presence of positive charges are usually required.[2]

The interaction of the polyamine with cell membrane, which to a great extent determines their functionality, is mainly of electrostatic nature and thereby depends on the number and position of electrical charges in the protonated polyamine. Charge distribution and location is also very important for understanding polyamine interactions with DNA and RNA.[2] As enumerated above, the central role of polyamines in biochemical processes, has led to considerable interest in establishing their protonation sequence ever since they were discovered in 1678 by Antoni van Leeuwenhoek.[14,17] There has been increasing interest and intense research efforts aimed at understanding and elucidating their chemical behaviour (particularly their protonation sequence) and the mechanism responsible for the effects they have on biological processes in living systems – both plants and animals.

Hence several experimental techniques[20–28] have been applied in order to understand their protonation sequence but until now there exists in literature, a controversy on the sites of protonation particularly for the monoprotonated forms. Furthermore the difficulty encountered in isolating the singly protonated form by some physical techniques such as X-ray

crystallography has made a conclusive resolution of this challenge impossible.[22,23] This difficulty arises due to the fact that at any given pH, all protonated forms coexist in solution, and there is no easy way to isolate each of the partially protonated forms. In reality, the picture is even more complicated as a polyamine with *n* basic sites exists in solution as a collection of 2^n protonated states called microstates (or microspecies). These microstates simply represent all possible combinatorial distributions of protons to the basic sites at each stage of protonation. These different states can be descibed using a two-valued state variable s_i for each individual site *i* (i = 1,2,…, N) such that s_i =1 if the site is protonated and s_i =0 if the site is deprotonated.[10] The protonation state of all the groups within the molecule at any given stage of protonation, which is referred to as one particular microstate, can then be represented by a set of state variables $\{s_1,s_2,\ldots,s_N\}$, which could be abbreviated as $\{s_i\}$. Because of the exponential increase in the possible number of microstates for a given polyamine as the number of its basic sites increases linearly, it is virtually impossible to determine directly from experiment, the protonation sequence (microscopic equilibria) of polyamines.

1.2.1. Previous Studies of Protonation Sequence

Although numerous papers have investigated the protonation sequence of the same polyamine e.g. spermidine using the same experimental technique in several cases, the need to deduce the protonation pattern indirectly from experimental methods has led to contradicting interpretations of results obtained from the same experimental method by different research groups.[20–28] Various experimental techniques such as calorimetry, mass spectrometry, UV spectrometry, potentiometry and NMR spectroscopy have been applied to this problem till date but there is no general agreement as to the conclusions reached in different studies.

The earliest work on protonation sequence of polyamines was the one by Sudmeir and Reiley which appeared in 1964 and involved the use of ^{1}H NMR data as a function of pH to elucidate their protonation pattern.[20] These authors concluded that for *trien* the protonation of the primary nitrogen atom is somewhat preferred in the first stage of protonation. Shortly after that, two studies, one by R. Barbucci et al.[24], and the other by G.R. Hedwig and H.K.J. Powell.[25] In these studies, the authors suggested using the heat of protonation (ΔH) data obtained from calorimetry that for *trien*, there is a tautomeric equilibrium between species protonated on the primary or secondary nitrogen atoms. Their argument is based on the fact that since the enthalpy data showed that $\Delta H_1 \approx \Delta H_2$, suggesting that the first and second protonation steps involve an addition of protons onto both primary and secondary amine nitrogens. In addition, they pointed out that since ΔS_1 (entropy change of monoprotonation) for primary and secondary monoamines

as well as several aliphatic linear polyamines are of a similar magnitude (varying from 28.3 to 32.7 $Jmol^{-1}K^{-1}$), protonation in the first stepwise process occurs neither exclusively on a primary or secondary nitrogen atom.

Following these earlier studies, a number of papers[13,24,27,29] particularly focused on elucidating the protonation sequence of spermidine using various experimental methods but there was also no agreement as to their conclusions about its protonation pattern. While Kimberly and Goldstein[27] concluded from a ^{13}C NMR–pH titration that the nitrogen atom attached to the longer chain is protonated for a greater percentage of the time in the first step (i.e. N_a and N_b is 32%, while N_c is protonated 43% of the time , see Figure 1), Anichini et al.[29] on the other hand concluded that the nitrogen atom which is protonated in the first step is a primary amine, a conclusion which was also supported by Delfini and co-workers who studied the protonation sequence of spermidine using ^{13}C–NMR and ^{1}H–NMR.[24] In agreement with Kimberly and Goldstein, another study by David Aikens et al. also proposed that the extent of protonation at the three basic sites of spermidine (N_a, N_b and N_c) is 28, 31 and 40% respectively.[13]

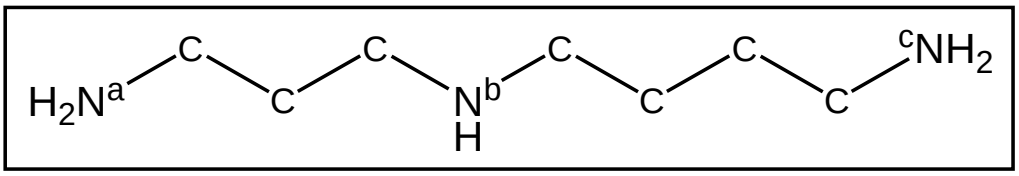


Figure.1.1. Structure of Spermidine showing the labelling of nitrogen atoms

It should be noted at this point however that studies involving the use of NMR spectroscopy utilized nonlinear curve fitting procedure to estimate microscopic protonation constants from which protonation sequence is determined. This involved the fitting of as much as fifteen adjustable parameters in some cases.[30] Since it is well-known that curve fitting methods lose much of their value when the number of adjustable parameters is this large, this raises a question as to the validity of these studies. The implication of this is that although the derived constants may exhibit an excellent fit statistically, their physical significance becomes questionable. More recently, Hague and Moreton[22,23] have demonstrated that ^{13}C chemical shifts of linear aliphatic polyamines, where the nitrogens are separated by two or three methylene groups can all be expressed as a linear combination of two amine shift parameters (π and π^*), which depend on the nature of the adjacent amino groups and the electrical charge of the amino group and their distance from the carbon atom i.e. α, β or γ. This can be expressed as

$$\delta_{calc} = \pi_1(\text{or } \pi_1^*) + \pi_2(\text{or } \pi_2^*) \qquad (1.1)$$

With the aid of this equation, the predicted chemical shifts for all the carbon atoms given a particular protonation sequence can be calculated. Since the experimental shifts observed from the ^{13}C NMR spectra are a weighted average of the chemical shifts of the different microstates contributing to a given protonation state, the closer the calculated chemical shift of a microstate is to experimental value, the greater will be its percentage contribution to such a protonation state. However, the microconstants obtained from these parameters could not pinpoint the preferred microstate out of the possible ones in certain cases and hence an unambiguous choice of the protonation sequence could not be made.[23]

Finally and most recently, Borkovec et al. have resorted to a cluster expansion technique in order to derive microconstants and consequently protonation sequence of polyamines from NMR studies.[10–12] These techniques enable one to parameterize the large number of microconstants in terms of much smaller number of cluster parameters. These parameters consist of one microconstant per basic site and a set of interaction parameters, which account for the influence of neighbouring groups. Furthermore, these parameters may satisfy simple group additivity relationships, and their number can be substantially reduced by taking into account molecular symmetries. Using this technique, they noted that for spermidine, the degree of protonation of the various basic sites on addition of a proton to it is 54.9%, 31.6% and 13.5% for N_a, N_b and N_c respectively. (See figure 1), i.e. the dominant singly protonated species is the one where the propyl-amine group is protonated, but the other two species also occur at substantial concentrations. For *trien*, they suggested that the primary and secondary nitrogen atoms are both protonated in the ratio 86% to 14% respectively.

1.3. Aim of this Study

It is clear from our brief discussion of previous works that although various experimental techniques such as mass spectrometry, calorimetry, potentiometry, and NMR titration procedures have been utilized in order to understand the protonation sequence of aliphatic polyamines, clearly there is still a lack of consensus in this field as to which of the nitrogen atoms (primary or secondary) is protonated in the first stage of protonation for a given polyamine.

In recent years, significant advances have been made in computational chemistry due to the increase in computational power and the development of numerous softwares for implementing electronic structure theory algorithms.[31] A number of ab-initio methods such as HF, MP2 and DFT methods can now be used to investigate chemical structure and reactivity problems such as the determination of protonation sequence of aliphatic polyamines within reasonable time and affordable computational resources.

Theoretical investigations of such problems should be of interest to chemists because (i) One can gain mechanistic insights from theoretical/computational modelling which is otherwise not accessible via experiment. (e.g. protonation states such as conformational or tautomeric forms that are inaccessible experimentally can be assessed theoretically for useful information (ii) this creates an avenue for predicting the protonation sequence of yet to be synthesized molecules (iii) it eliminates the difficulties related to the experimental study of the protonation reactions of many biomolecules due to solubility and stability issues.[7,8] Hence one would have envisaged that numerous theoretical studies on this unresolved challenge would have appeared in literature in recent years due to the availability of cheap state of art quantum chemistry methods. Unfortunately, this is not the case as there is little or no computational studies focused on protonation behaviour of aliphatic linear polyamines. This might be due to their extreme flexibility and almost infinite number of conformational states in which they could exist, hence making a theoretical investigation of their chemistry a difficult and almost impossible task.

In view of all these challenges, a major step forward in this direction would be the development of a dedicated protocol for characterizing low energy conformations of any given aliphatic polyamine. This is not an easy task given their extreme conformational and tautomeric diversity. Hence this study involved firstly, the development and testing of a computational protocol to characterize the low energy conformations of any given aliphatic polyamine so as to be able to carry out computational studies on it. This protocol is applicable (with a slight modification) to any given polyamine. In the second section, the conformational search protocol developed was combined with a competition based reaction methodology to predict the four stepwise protonation constants of triethylenetetramine (2,2,2-tet or *trien*) with excellent accuracy. Protonation constant is a very crucial indicator of the potential of any molecule as a drug since this determines how easily the molecule can move across cell membranes in biological system. Therefore being able to theoretically predict the stepwise protonation constants of polyamines is very important for a preliminary evaluation of their bioactivity prior to embarking on expensive synthetic process in the laboratory.

Finally, a preliminary theoretical investigation of the protonation sequence of *trien* by computing theoretical NMR shifts and then analysing the shifts in order to gain insight as to its proton distribution pattern was carried out. Furthermore, an attempt was made to ascertain whether either the primary or secondary nitrogen atom or a mixture of both would be formed in solution during the first stage of protonation. Results obtained in this work have been compared to that of studies involving a nonlinear regression analysis of experimental NMR shift

measurements. Results of this investigation were discussed in the light of the controversy that exists in literature regarding the initial site of protonation in aliphatic polyamines.

1.4. References

[1] A. Bencini, A.Bianchi, E. García-España, M. Micheloni, J. A. Ramirez, *Coord. Chem. Rev.*, **1999**, *188*, 97–156.

[2] M. T. Albeda, J.C. Frías, E. García-España, *Encyclopaedia of Supramolecular Chemistry*, **2007,** *1*:*1*,1–37.

[3] R. Casasnovas, J. Ortega-Castro, J. Frau, J. Donoso, F. Munoz, *Int. J. Quantum Chem.*, **2014**, 114, 1350–1363.

[4] J. Ho, *Aust. J. Chem.*, **2014**, 67, 1441–1460.

[5] J. Ho, M. L. Coote, *Comput. Mol. Sci.*, **2011**, 1, 649–660, doi:10.1002/WCMSW.43.

[6] M. Namazian, S. Halvani, *J.Chem. Thermodyn.*, **2006**, *38*, 1495–1502.

[7] K. K. Govender, I. Cukrowski, *J. Phys. Chem. A.*, **2009**, *113*, 3639–3647.

[8] K. K. Govender, I. Cukrowski, *J. Phys. Chem. A.*, **2010**, *114*, 1868–1878.

[9] D. T. Manallack, *Perspectives in Medicinal Chemistry*, **2007**, *1*, 25–38.

[10] M. Borkovec, G. J. M. Koper, *Anal. Chem.*, **2000**, *72*, 3272–3279.

[11] M. Borkovec, G. J. M. Koper, *J. Phys. Chem.*, **1994**, *98*, 6038–6045.

[12] M. Borkovec, D. Cakara, G. J. M. Koper, *J. Phys. Chem. B.*, **2012**, *116*, 4300–4309.

[13] F. Onasch, D. Aikens, S. Bunce, H. Schwartz, D. Nairn, C. Hurwitz, *Biophys. Chem.*, **1984**, *19*, 245–253.

[14] B. Ganem, *Acc. Chem. Res.*, **1982**, *15*, 290–298.

[15] M. D. Bratek-Wiewiórowska, M. Alejska, M. Figlerowicz, J. Barciszewski, M. Wiewiórowski, *Pure & Appl. Chem.*, **1987**, *59*, 407–414.

[16] T. Thomas, T. J. Thomas, *Cell. Mol. Life Sci.*, **2001**, *58*, 244–258.

[17] A. Gugliucci, *Clinica Chimica Acta.*, **2004**, *344*, 23–35.

[18] E. Agostinelli, M. P. M. Marques, R. Calheiros, F. P. S. C. Gil, G. Tempera, N. Viceconte, V. Battaglia, S. Grancara, A. Toninello, *Amino Acids* , **2010,** *38*, 393–403.

[19] M.M. Kimberly, J. H. Goldstein, *Anal. Chem*., **1981**, *53*, 789–793.

[20] J. L. Sudmeir, C. N. Reiley, *Anal. Chem*., **1964**, 1698–1706.

[21] P. Paoletti, M. Ciampolini, A. Vacca, *J. Phys. Chem*., **1963**, *67*, 1065–1067.

[22] D. N. Hague, A. D. Moreton, *J. Chem. Soc. Perkin Trans*., **1994**, *2*, 265– 270.

[23] S. P. Dagnall, D. N. Hague, M. E. McAdam, *J. Chem. Soc. Perkin Trans*., **1984**, *2*, 1111–1114.

[24] M. Delfini, A. L. Segre, F. Conti, R. Barbucci, V. Barone, P. Ferruti, *J. Chem. Soc. Perkin Trans*., **1980**, *2*, 900–903.

[25] G. R. Hedwig, H. K. J. Powell, *J.C.S. Dalton*., **1973**, 793–797.

[26] R. Barbucci, L. Fabbrizzi, P. Paoletti, *J.C.S. Dalton*., **1972**, 745–749.

[27] M. M. Kimberly, J. H. Goldstein, *Anal. Chem*., **1981**, *53*, 789–793.

[28] I. Cukrowski C. F. Matta, *Comput. Theoret. Chem*., **2011**, *966*, 213–219.

[29] A. Anichini, L. Fabrizzi, R. Barbucci, A. Mastorianni, *J.C.S Dalton*., **1977**, 2224–2228.

[30] D. Aikens, S. Bunce, F. Onasch, R. Parker III, C Hurwitz, S Clemans, *Biophys. Chem*., **1983**, *17*, 67–74.

[31] E. G. Lewars, in *Computational Chemistry: Introduction to the theory and Applications of Molecular and Quantum Mechanics*. Second Edition, 2011. Springer Dordrecht.

Chapter 2

Computational methods

2.1. Introduction

In order to accomplish the goal of this thesis, a number of computational chemistry methods have been utilized which can be broadly classified into three groups; (i) force field methods such molecular mechanics as implemented in the Merck Molecular Force Field (MMFF) which was used to carry out conformational analysis (ii) electronic structure methods such as ab initio quantum chemistry methods and density functional theory based methods which were used to obtain the optimized geometry, electron density, and electronic energy component of Gibbs free energy for molecules investigated in this study and (iii) statistical mechanics methods used to compute contributions due to thermal motions of nuclei to the Gibbs free energies (in gas and solution phase) . Furthermore, topological methods such as the popular Quantum theory of atomis in molecules (QTAIM) and Non-covalent interaction index (NCI) were used as computational tools to obtain chemically meaningful information from electron density and to analyse intra- and intermolecular interactions that influence conformational preference of polyamines. Hence the aim of this chapter is to give a brief theoretical background of these methods and in some cases specific information about their practical implementation in order to ensure reproducibility of this work.

Computational chemistry involves modelling all aspects of chemistry by calculation instead of carrying out experiments with reagents in the laboratory.[1,2] It can be described simply as performing chemistry using computers instead of chemicals.[3] It encompasses a broad range of topics including but not limited to cheminformatics , statistical mechanics, molecular mechanics, semi-empirical methods and *Ab–initio* quantum chemistry. Computational chemistry is fast becoming indispensable for chemistry research due to the following reasons:

- It provides an avenue for modelling a molecular system before synthesizing it in the laboratory. Theoretical models are often accurate enough to eliminate 90% of possible compounds as being suitable for a particular application.[4] This information is invaluable because synthesizing a single compound could require months of labour and raw materials and generate toxic waste. This could be avoided by carrying out a preliminary investigation using computational methods.

- It enables us to obtain properties of molecules which are tedious and costly to determine using experimental methods. For example it has been pointed out the cost of evaluating heat of formation for a molecule with computational methods is four times more cost efficient than experimental approaches for results of comparable accuracy.[5]

- The dramatic increase in computer speed and the design of efficient quantum chemical algorithms. This has made calculations on medium to moderate chemical systems feasible on a personal computer.[1]

However, despite its huge success, computational chemistry cannot replace experimental studies. It only plays an important role in aiding chemists to gain insight into chemical structure and reactivity. More importantly, it also enables chemists to explore new or unknown chemistry.[3] Computational chemistry has become a vital tool in the arsenal of chemists (including non-specialists) used in much the same way as infrared or NMR spectroscopy have become versatile tools to investigate the structure of molecules by experimental chemists[1], therefore care is taken here not to delve too deep into the theoretical background and derivations of the various methods. The discussion presented in this chapter were mainly derived from two comprehensive and invaluable textbooks on computational and quantum chemistry, Computational Chemistry; Introduction to the theory and Applications of Molecular Quantum Mechanics by Errol G. Lewars[1] and Exploring Chemistry with Electronic Structure methods by Eileen Frisch.[6] Minor contributions were also taken from the textbooks by Koch and Holthausen[7] and Cramer.[8]

2.2. Molecular Mechanics

Molecular mechanics (MM) uses the laws of classical physics to predict the structures and properties of molecules. It is based on the idea that a molecule can be modelled theoretically as a collection of balls (representing atoms) connected by springs.[1] The energy of the molecule is then described as a function of its resistance toward bond stretching, bond bending and steric crowding. This function can then be optimized to find bond lengths, angles and dihedrals corresponding to the minimum energy geometry. Various molecular mechanics method exists, each distinguished by its particular force field. A force field consists of a set of parameters incorporated into a mathematical expression showing how the potential energy of a molecule varies with respect to the locations of its constituent atoms. The term “force field” originated from a well-known fact that the negative of the first derivative of potential energy of a particle with respect to displacement along a given direction is the force on the particle. Specifically, each force field is characterized by a series of atom types which defines the characteristics and behaviour of an element within a specific chemical environment. Factors such as hybridization, charge, and the types of atoms to which an element is bonded determines its chemical environment according to MM model.[6]

To develop force field, a set of parameters are used to fit equations and atom types to

experimental and or computational data obtained from ab intio computations. The parameter sets define force constants, which are values used in the equations to relate the characteristics of atoms to molecular energy and structural data such as bond lengths and angles. Starting from normal bond lengths and the angles between them, and how much energy it takes to stretch and bend the bonds, we can calculate the energy of any given guess geometry of a particular molecule. Hence, this initial guess geometry can be varied until the lowest energy structure is found i.e. geometry optimization. Molecular mechanics is fast hence fairly large biomolecules e.g. amino acids and polyamines can be optimized in few seconds on a good minicomputer. This speed is achievable because molecular mechanics theory does not account explicitly for the electrons in a molecular system. Its computations are simply based on nuclear interactions. On the other hand, since MM neglects electronic motion it cannot provide insight on electronic properties like charge distributions or chemical reactivity. However, electronic effects are implicitly included in force fields through parameterization. Due to the formulation strategy, molecular mechanics calculations have the following inherent limitations[6]:

(i) A given force field will only achieve good results for a select class of molecules, similar in structure to those for which it was parameterized. No force field can be generally used for all molecular systems of interest.

(ii) MM cannot be used solely to investigate chemical problems where electronic effects are predominant e.g. bond formation or bond breaking. In addition, molecular properties which depend on subtle electronic details cannot be investigated using MM methods.

MM differs mainly from electronic structure calculations in that the concept of a bond is central to its description of a molecule unlike electronic structure calculations which depends on relative positions of constituent atomic nuclei, number of electrons, and multiplicity. MM is the most widely-used method for computing the geometries and energies of large biomolecules such as proteins and nucleic acids.

Generally, in MM theory, the potential energy of a molecule is expressed as:

$$E = \sum_{bonds} E_{\text{stretch}} + \sum_{angles} E_{bend} + \sum_{dihedrals} E_{torsion} + \sum_{pairs} E_{nonbond} \qquad (2.1)$$

where the various energetic contributions are attributed to bond stretching, angle bending, torsional motion, and interactions between nonbonded atoms or groups respectively. These contributions are then summed up are over all bonds, all the angles defined by three atoms, all the dihedral angles defined by four atoms and all pairs of significant nonbonded interactions. The mathematical forms of these terms and the parameters in them are unique for each particular force field.

Specifically, the MMFF force field developed by Merck research laboratories [9] was used in this thesis to discover the low energy conformers of polyamines because it was specially parameterized with the aim of precisely reproducing conformational energies and describing nonbonded interactions. Some of the distinct features of MMFF in comparison to other force fields, are (i) It uses a unique functional form to describe vDW interactions, (ii) its parameterization is mainly based on a large amount of high quality computational data obtained from *ab initio* electronic structure methods and (iii) it was parameterized to handle an unusually wide variety of organic molecules. The MMFFaq force field is a slightly modified version which incorporates energetic correction for the effect of aqueous solvent. It should be noted however that this correction does not affect the molecular structure in any way.[10] MM method has been used extensively in this work to locate energy minima on the conformational energy surface of aliphatic linear polyamines.

2.2.1 Conformational Search Methods

Conformational search methods consist of various computational/theoretical techniques and strategies used to explore conformational space in order to characterize a representative sample of all energetically favourable conformations of a molecule. This is a crucial aspect of any computational chemistry research work such as the one in this thesis where interpretation and understanding of experimental results e.g. NMR spectroscopic data is being sought from a fundamental perspective.[11] All conformational search methods reported in literature fall under three broad categories;

- Systematic method: This involves a thorough exploration of all areas of conformational space by a gradual variation of the torsion angles for rotatable bonds in a molecule. Typically, the torsion angles are varied in step sizes of 360◦/n, where n could be 3, 6 or 12 giving a total of 2^n possible conformers. Hence the number of conformers generated increases as *n* increases. Starting geometries are produced by using a combination of selected values for some or all rotatable torsion angles distributed throughout conformational space. Therefore in principle, systematic search methods are able to characterize all sterically possible conformations of a molecule hence providing guaranteed coverage of all regions of space. Systematic methods are generally preferred to stochastic methods due to their inherent ability to comb through all energy regions of conformational space. Stochastic methods on the other hand are not path dependent and have the tendency of being terminated prematurely in a local minima within the conformational space. The only disadvantage of systematic methods is that the number of

possible conformations increases exponentially as the number of rotatable bonds increase, this is why random methods are automatically used in molecular modelling softwares when the number of rotatable bonds is greater than five.

- Stochastic or random methods: This is the exact opposite of systematic search. Here the conformational search algorithm is fashioned in such a way that it can move from one region of the conformational space to a completely unconnected new region in a single step. These methods walk randomly through various regions of conformational space in order to identify energetically favourable conformations. Likewise they employ a random translation algorithm to generate starting geometries. However, if they are unconstrained and allowed to run for a sufficiently long period of time random methods will cover all regions of space. It should be noted however that these methods do not search conformational space in a completely random manner. Rather they proceed with stable conformers and limit their conformational space explorations to variations in selected internal coordinates (i.e. torsion angle) or to small (< 3.5 Å) Cartesian displacements of atoms. Random searches takes advantage of all symmetry elements to simplify conformational search while systematic methods make it difficult to take advantage of all symmetry elements and tend to approach all conformational problems as if they were the more complex (but more general) unsymmetrical ones. Common examples of stochastic methods are:

  (i) Monte Carlo (MC) Method: This method involves using a single starting geometry repeatedly to discover various low energy conformers that vary structurally compared to starting structures. It is designed to have a bias for stable conformations hence some MC algorithms ignore high energy conformers since they have negligible contributions to conformer distribution. This kind of approach is referred to as importance sampling. The parameter coordinates are changed at each step so as to achieve maximum effectiveness. Monte Carlo method is suitable for large molecules with complex interconversion pathway. Due to unavailability of a natural end point; the conformational search process continues until either a predefined number of iterations have been attempted and/or until no new conformations are generated.[10]

  (ii) Simulated Annealing Method: This approach is a variation of the Monte Carlo method in which the probability of locating the global minimum is increased using a thermodynamic process. It involves running the Monte Carlo Algorithm

in phases at different temperatures. To begin, the molecule is assumed to exist in a high temperature system. Its bonds are randomly rotated or rings bent until a preferential (minimum energy) geometry is attained. Due to the initial high temperature of the system, the molecule has significant energy and is flexible enough to move from a low to high energy conformation. This is crucial due to the often different structure of the global minimum compared with the initial conformation. As more conformers are explored, the temperature of the system decreases, making the molecules less inclined to move out of low energy conformations, thus forcing it to narrow down its search to other minima in this region. In essence, this method only samples a small part of the conformation space meticulously. Similar to the Monte Carlo method, the simulated annealing method uses Boltzmann probability criteria to accept or reject new conformers. Several modifications have been made to avoid a situation where the search gets stuck in a particular minima.[12]

(iii) Molecular Dynamics Method: This involves the uses of Newtonian laws of mechanics to generate representative set of conformations from a starting geometry of a given molecule. Starting from an initial geometrical position and given velocity, atoms of a molecule are then allowed to evolve in time. This provides a trajectory that defines how the positions and velocities of the particles of the system varies with time. After a brief period of motion usually 10^{-5} seconds, the position and velocity of each atom is determined. Availability of the initial and final atomic positions and velocities then facilitates the calculation of forces acting on the molecule. Furthermore one can calculate the acceleration and direction of each atom. To obtain various conformations, the molecule is "heated" implying that changes in the bond length and torsion angles occur. This process can be repeated automatically to give any number of feasible structures. Molecular Dynamics has the added advantage of not only providing information about the various possible conformations but also the interconversion path of the conformers discovered.[13]

- Heuristic or artificial intelligence based methods: Although these methods are in principle stochastic in operation, they are unique in the way they probe conformational space hence their classification as another type of conformational search method. These methods identify a representative set of conformations by only exploring a small fraction of conformational configuration space using customised algorithms. Heuristic methods

can be further divided into step and non-step methods. Non-step methods generate a series of system configurations that are independent of each other. Step methods on the other hand generate a complete system configuration in a stepwise manner by either using configurations of molecular fragments or the previous configuration.[14] Examples of heuristic methods include Genetic algorithms, evolutionary algorithm and artificial neural networks.

2.3. Electronic Structure Methods

Electronic structure methods consist of a broad group of computational methods which apply the principle of quantum mechanics to molecular problems. They are used to predict molecular structures as well as physiochemical properties of molecules. In some cases, these methods can even be used to investigate the outcome of a chemical reaction. They all aim at solving the Schrodinger equation for a molecule by applying varying levels of approximations. The Schrodinger equation is one of the fundamental equations of modern physics and describes, among other things, the behaviour of electrons in a molecule. The time-independent, non-relativistic Schrodinger wave equation is a fundamental postulate of quantum mechanics and it is the theoretical foundation of all electronic structure methods:

$$\hat{H}\Psi_i(x_1, x_2, \ldots, x_N, R_1, R_2, \ldots, R_M) = E_i\Psi_i(x_1, x_2, \ldots, x_N, R_1, R_2, \ldots, R_M) \quad (2.2)$$

where Ψ_i is not only a function of N electrons and M nuclei, but also depends on the spatial coordinates of each of these fundamental particles. E_i is the energy (scalar number) corresponding to a specific Ψ_i and $\hat{H}$ is the overall Hamilton operator,

$\hat{H}$ comprises of all the terms that contribute to the energy of a system:

$$\hat{H} = \hat{T} + \hat{V} \quad (2.3)$$

$\hat{T}$ is the kinetic energy operator which can be further decomposed into kinetic energy due to the electronic motion ($\hat{T}_e$) and that due to nuclear motion $\hat{T}_n$. Similarly $\hat{V}$ is the potential energy operator and is the sum of nuclear-nuclear repulsion $\hat{V}_{nn}$, nuclear-electron repulsion $\hat{V}_{ne}$ and electron-electron repulsion $\hat{V}_{ee}$, therefore

$$\hat{H} = \hat{T}_e + \hat{T}_N + \hat{V}_{eN} + \hat{V}_{ee} + \hat{V}_{NN} \quad (2.4)$$

$$H = -\frac{1}{2}\sum_{i=1}\nabla^2_i - \frac{1}{2}\sum_A \frac{1}{M_A}\nabla^2_i - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{1A}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A Z_B}{R_{AB}} \quad (2.5)$$

In the equation above, r_{1A}, r_{ij} and R_{AB} are the distance between a specified electronic coordinate

and nuclear coordinate, the distance between two electronic coordinates and that between two nuclear coordinates, in that order. Z_A represents the nuclear charge for nucleus A whilst M_A gives the mass of the nucleus A. Lastly, ∇^2 refers to the Laplacian operator in Cartesian coordinates and is given by:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{2.6}$$

Unfortunately equation (2.5) is intractable due to the correlated movement of each particle with all other ones. To simplify this situation, the Born-Oppenheimer approximation is invoked. This is based on the assumption that the positions of nuclei may be assumed to be fixed since the motions of electrons in a molecule is much faster than that of the nuclei. In essence, this allows separation of the nuclear and electronic terms in the Schrodinger equation hence enabling it to be solved for fixed positions of the nuclei, and the electronic energy can be calculated at various internuclear distances. It is important to note at this point that the Born-Oppenheimer approximation is central to all electronic structure calculations as it allows the potential energy profile of the molecule to be conveniently studied.

Electronic structure theories differ from another in the level of mathematical approximation applied in order to obtain a meaningful solution to the electronic Schrodinger equation. The solution of the equation provides the wavefunction, Ψ, which in principle tells us all that can be known about the particular chemical system. In addition, it describes the behaviour of electrons in atoms and molecules as well as enables us to compute their associated energies, E. It is well-known, however, that exact analytical solutions can only be found for simple systems such as the particle in a box, harmonic oscillator, rigid rotor, and the hydrogen atom or hydrogen-like ions. This challenge is aptly summarized by Dirac[15] in his famous quotation of 1929 where he noted that "*The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble*". Hence the general problem of electronic structure theory is concerned with how to obtain approximate and realistic solutions to the Schrodinger equation using various mathematical approaches and assumptions unique to each electronic structure method. For a particular research problem, the choice of electronic structure method applied should be guided by the level of accuracy required to obtain sensible chemical information from the system being investigated. In general, electronic structure theory methods can be classified broadly into ab initio, density functional theory and semiempirical methods some of which we will be discussing in the following sections.

2.4. Ab initio methods

2.4.1. Hatree-Fock Method

The goal of *Ab initio* methods is to obtain direct solutions i.e. energy and other related properties of molecules from the Schrodinger equation using only mathematically tested approximations. These methods distinguish themselves from other computational methods in that they are based solely on established laws of quantum mechanics. However as mentioned earlier, the exact solution of the Schrodinger equation is only practical for any but the smallest chemical systems. Thus approximations are used; the less serious these are, the "higher" the level of the ab initio calculation is said to be. Regardless of its level, an *Ab initio* calculation is based only on the established laws of nature (i.e. quantum mechanics) and is in this sense "from first principles". *Ab initio* methods solve the Schrodinger wave equation for a molecule in order to obtain its wavefunction and corresponding energy.

The HF method is the cornerstone of all traditional wave function based *Ab initio* electronic structure methods. It involves the simplest albeit physically sound approximation and is synonymous with the molecular orbital approximation. Here it is assumed that each electron only experiences the mean field of all other electrons in the system. Hence the probability of finding a given electron at a particular position in space is independent of the positions of the other electrons. It eliminates the need for the evaluation of complex two electron integrals in the original Schrodinger equation resulting from the interdependent or correlated motion of electrons in a molecule. In this theory, the many electron wave function of a molecule is written as an antisymmetrized product of N one-electron wave functions $\chi_i(x_i)$.where N is the number of electrons in the molecule. This product is usually referred to as a Slater determinant, Φ_{SD}:

$$\Psi o \approx \Phi_{SD} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\vec{x}_1) & \chi_2(\vec{x}_1) & \cdots & \chi_N(\vec{x}_1) \\ \chi_1(\vec{x}_2) & \chi_2(\vec{x}_2) & & \chi_N(\vec{x}_2) \\ \vdots & \vdots & \vdots & \vdots \\ \chi_1(\vec{x}_N) & \chi_2(\vec{x}_N) & \cdots & \chi_N(\vec{x}_N) \end{vmatrix} \tag{2.7}$$

This determinant can be abbreviated in such a way that only the diagonal elements are displayed:

$$\Phi_{SD} = \frac{1}{\sqrt{N!}} \det\left\{ \chi_1(\vec{x}_1)\ \chi_1(\vec{x}_1)\ \cdots \chi_1(\vec{x}_1) \right\} \tag{2.8}$$

These individual one-electron wave functions $\chi_i(\vec{x}_i)$ are called molecular orbitals (MO) and each is also a product of a spatial orbital $\phi(r)$ and either of the spin functions, $\alpha(s)$ or $\beta(s)$ hence they are commonly referred to as spin orbitals.

$\chi_i(\vec{x}_i) = \varphi(r)\sigma(s)$, where $\sigma = \alpha, \beta$.

One crucial property of these spin orbitals is that they are orthonormal since both the spin functions and the spin orbitals themselves are chosen to be orthonormal for computational efficiency i.e.

$$\int \psi_i \psi_j = \left\langle \psi_i \middle| \psi_j \right\rangle = \delta_{ij} \tag{2.9}$$

where δ_{ij} is the kronecker delta (which is equal to 1 if $i = j$, and zero otherwise). The antisymmetric nature of the exact wave function is preserved by the slater determinant due to the fact that any determinant changes sign if two columns or rows are interchanged. In essence, the HF wavefunction is an antisymmetric wave function expressed in terms of the one-electron molecular orbitals. Furthermore, each of these molecular orbitals can be represented as a linear combination of atomic orbitals i.e.

$$\psi_i(r_i) = \sum_\mu C_{\mu i}\chi_\mu(r_i) \tag{2.10}$$

Where χ_μ are atomic orbitals or basis functions and $C_{\mu i}$ are MO coefficients.

If the wave function is normalized, the expectation value of the energy is given by:

$$E = \left\langle \psi \middle| \hat{H} \middle| \psi \right\rangle \tag{2.11}$$

In case of the HF wavefunction, the expectation value of the energy is given by:

$$E_{HF} = \sum_i H_i + \frac{1}{2}\sum_{ij}(j_{ij} - k_{ij}) \tag{2.12}$$

H_i collects all the one-electron terms arising from the kinetic energy of the electrons and the nuclear attraction energy.

J_{ij} involves two-electron terms associated with the coulomb repulsion between the electrons and

K_{ij} involves two-electron terms associated with the exchange of electronic coordinates.

Since our wavefunction is given in terms of a determinant and our MOs are written as a Linear combination of atomic orbitals (LCAO), the one-electron parts of the energy can be expressed as:

$$H_i = \langle \psi_i | \hat{h} | \psi_i \rangle = \sum_{\mu\nu} C_{\mu i} C_{\nu i} \langle \chi_\mu | \hat{h} | \chi_\nu \rangle \quad (2.13)$$

Similarly, we can express the J_{ij} and K_{ij} matrices in terms of the MO coefficients, $C_{\mu i}$.

To proceed from this point, the variational theorem is then utilized to determine the MO coefficients. This theorem states that the energy determined from any approximate wave function will always be greater than the energy for the exact wave function. The implication of this is that the energy of the exact wavefunction will serve as a lower bound to that obtained from the approximate one, hence $C_{\mu i}$ can be simply varied until the total energy of the system is minimized with the constraint that the spin orbitals remain orthonormal. In practise, a set of guess molecular orbitals is used to construct a trial wave function from which one could obtain an expectation value of the energy. This expectation value of the energy is then minimized iteratively via a procedure referred to as the self-consistent field until the difference between the energies obtained from two successive minimizations obtained fall within an acceptable threshold.

The implication of the assumptions made in the HF method is that, electrons are allowed to approach each other a little too closely. Consequently HF calculations overestimate electron–electron repulsion and so predict higher electronic energies than the correct ones, even with huge basis sets. Also the difference between the energy obtained from the HF method (using an infinite basis set) and that of the exact energy (i.e. exact eigenvalue of the Hamiltonian) for the actual *N*–electron wave function is the well–known correlation energy. This is the energy due to the correlated motion of electrons. All other *Ab initio* techniques such as Møller–Plesset perturbation theory or density functional theory that we discuss subsequently seek for a means of accounting for this energy.

2.4.2. Basis Sets

A basis set is a set of mathematical functions (basis functions) which is combined linearly to yield another function suitable for representing a molecular orbital. Selection of an appropriate basis set is an important aspect of any electronic structure calculations as these calculations depend on the assumption that molecular wave functions can be represented as a linear combination of atomic orbitals. This dependence can be linked to the intrinsic ability of computers to perform better when solving algebraic sets of equations as compared to differential ones.[16] In order to take advantage of this innate ability of computers, HF equations are set up in a linear algebra form by expanding the unknown molecular orbital functions in terms of a given,

fixed set of functions whose number is finite. It is also crucial to select the functions in such a way that as their numbers increases, we would obtain increasingly accurate approximations of the molecular orbitals and there will be fewer restrictions on the wave function obtained from such molecular orbitals.

These functions are called the atomic orbital (AO) basis, because they are usually but necessarily centred on the atom. AO basis functions are Gaussian functions, or linear combinations of Gaussians because this form enables all required matrix elements to be evaluated analytically. Approximating molecular orbitals (MO) as linear combinations of basis functions is usually referred to as the linear combination of atomic orbital (LCAO) method. It should also be noted at this point that in reality, basis functions are not necessarily conventional atomic orbitals. They can be any set of mathematical functions provided that they can be conveniently manipulated. In addition, they should give useful representations of molecular orbitals when combined linearly. In essence, electron distribution around an atom can be described using several basis functions and a linear combination of these atomic basis functions yields the electron distribution in the molecule as a whole.

An atomic orbital can be represented in various ways: (i) Hydrogen like functions based on solutions of the Schrödinger equation for the hydrogen atom (ii) polynomial functions with adjustable parameters, (iii) Slater functions and (iv) Gaussian functions. The Slater and Gaussian functions are the ones commonly implemented in computational chemistry softwares because they can be easily manipulated mathematically. Furthermore, even though Slater functions are better approximations to atomic wave functions, the prohibitive computational time required for the evaluation of certain two-electron integrals with it makes Gaussian functions preferable in most commercial programs. Practically, this is achieved by using several Gaussian functions to approximate a Slater function. A single Gaussian is a poor approximation to the almost ideal description of an atomic wave function (Hydrogen like functions) that the Slater orbital provides hence the need to combine several Gaussians. Typically an *Ab-initio* basis functions consists of a set of primitive Gaussians combined together with a set of contraction coefficients i.e. a single basis function is made up of one or more Gaussian functions. For example an s-type basis function can be expressed as:

$$\Psi_{\mu}(r) = \sum_{i}^{N} d_{i\mu} \exp(-\alpha_{i\mu} f_{\mu}^{2} r^{2}) \qquad (2.14)$$

Where *N* is the number of Gaussian functions that makes up the basis function, and is referred to as the degree of contraction of the basis function.[17] The coefficients d_i are called contraction

coefficients, and they indicate the weight of each Gaussian contribution to the basis function. The quantities $\alpha_{i\mu}$ are called the exponents and they define the width of the Gaussian; a large value corresponds to a sharp Gaussian whereas a small value corresponds to a broad Gaussian. The parameters f_μ are called the scale factors for the basis functions although they are often set equal to one. Optimum values for all the coefficients are obtained by fitting this equation to experimentally observed atomic properties such as ionization energy.

Numerous standardized basis sets are widely available under various and seemingly obscure acronyms. They are generally named according to the number of AO per valence atomic orbital used in building them. The smallest basis set implemented in standard ab initio calculations by commercial programs is of the minimal type and is referred to as STO–3G meaning “Slater-type orbitals approximated by three Gaussians. Here the 1s orbital and each of the 2s and 2p orbitals are represented by three Gaussians. This description of an atomic orbital can be further improved upon by using two 1s functions for hydrogen (different α values) and two 2s and two 2p functions for atoms in the second row of the periodic table such as carbon. These are known as split valence basis sets because the valence orbitals are represented by two sets of functions but the core orbitals are represented by a single set of functions. For example, a 3–21G basis set implemented for carbon will consists of a combination of three Gaussians for 1s, two Gaussians each for 2s and 2p plus one Gaussian each for 2s’ and 2p’. The notation used to describe basis sets here can be denoted generally as *i-jk*, where *i* stands for the number of Gaussians representing each core basis functions and *j* and *k* are the numbers of Gaussians representing the split–valence basis functions.

In order to improve the accuracy of electronic structure calculations further, one can use the triple-split-valence basis sets, the most commonly used one being 6-311G. This is constituted from 6 Gaussians for the core basis functions and the valence functions are further split into three sets, comprising three, one and one Gaussians, respectively.

A totally different alternative to the split-valence basis sets is the multiple zeta basis sets the most common of which is the double zeta(DZ) and triple zeta (TZ) basis sets which split all the orbitals into either two or three sets of functions, where the term “zeta” refers to the exponents. Similarly, the quadruple zeta (QZ) basis sets split all orbitals into four sets of functions. A crucial difference between multiple zeta and split valence basis function is that different α coefficients are used for s and p orbitals. It is important to note at this junction that the essence of using multiple basis functions per atomic orbital is to allow for the expansion or contraction of the size of an orbital e.g. along a bond axis or perpendicular to a bond axis respectively.

To achieve better accuracy than that possible from a triple-split-valence basis set, one would often need to add functions corresponding to orbitals which have a higher angular momentum than those that are occupied. For example, p functions may be added to hydrogen, d functions to second row elements and so forth, and f functions to transition metals. These are usually indicated by adding an asterisk to the basis set or specifying p,d functions, and so forth, for example, 6-31G* or 6-31G(d) adds d functions while 6-31G** or 6-31G(d,p) adds d functions to second row elements and p functions to the hydrogen atom.

In the case of anions and atoms with lone pair of electrons where there is a possibility that some electron density exists far from the nuclei, this can be accounted for by adding diffuse functions, which are broad Gaussian functions (small α coefficients) that are not readily calculated for isolated atoms, but are chosen by well-established rule of thumb. These functions are represented in the basis set notation by a + or ++ sign. It has been shown that there is no need to add polarisation and diffuse functions to H atoms so for example, 6-31+G adds diffuse functions while 6-31++G adds diffuse functions to second row elements and hydrogen. When electron correlation is to be accounted for, “correlation consistent” basis sets: cc-pVNZ, where N=D (double), T(triple), Q(quadruple) and 5(quintuple) zeta should be utilized. These basis sets also contain polarization functions and diffuse functions can be added by using the prefix “AUG”-.

As a consequence of their large number of electrons and the fact that larger basis sets are not available for heavier atoms such as transition metals, calculations on them can be time-consuming. In order to mitigate this, the effective core potential method is normally used in their calculations. This method only considers their valence electrons explicitly and the core electrons are only included as an effective core potential. There are a number of ECP methods with the most popular one being the Los Alamos ECP. This is usually combined with double-zeta functions for the valence electrons in the “LanL2DZ” basis set. This basis set is commonly employed in calculations involving transition metal compounds.

Most quantum chemistry softwares contain an extensive range of basis sets in addition to allowing the use of user-defined basis sets. To select a basis set for carrying out a particular calculation, the general approach taken is dictated by the nature of the problem at hand. Basis set should be chosen in such a way that the biggest (and therefore the best) basis set available for the atoms contained within the molecule of interest within an affordable timescale to obtain a chemically meaningful result is selected. The issue of an affordable time scale is subjective as it depends on a number of factors including but not limited to the timeframe within which a

researcher hopes to obtain a meaningful solution to the problem at hand. Hence it is not surprising that in practice, the split-valence basis sets 6-31G and 6-311G are the most widely used despite the fact that they are inferior to the double and triple zeta basis sets respectively. Moreover the zeta basis sets have the advantage that the addition of diffuse functions is not necessary since these basis sets contain Gaussian functions with small α values.

2.4.3 Chemical Accuracy and Electron Correlation

As pointed out earlier, the Hatree-Fock method does not account explicitly for the correlated motion of electrons hence there is a difference between HF molecular energies and the best estimates of exact nonrelativistic energies. This energy difference is referred to as the correlation energy. There are two components to the correlation energy[1,7,8];

(i) Dynamic correlation which has to do with the movement of electrons as they try to avoid crossing one another's path. This plays a crucial role in bond breaking processes.

(i) Static correlation which arises due to the deficiencies in the single determinant HF wave function. It is important in systems low lying excited states and elongated bonds.

Inability of any electronic structure method to account for this correlation energy might hinder it from achieving the 1kcal chemical accuracy target of such calculations. Hence, there is need to improve upon the HF description of molecular processes if we are to achieve chemical accuracy when modelling physiochemical properties. This can only be achieved by an accurate description of electron correlation. There exist numerous methods to compute the correlation energy, each with their own strengths and weaknesses as described in the following sections.

2.4.4 Moller-Plesset perturbation theory

In perturbation theory, the Hamiltonian can be represented by the addition of a small perturbation to one for which solutions are known in order to solve the Schrodinger equation.[3] Essentially, the Hamiltonian is divided into two parts:

$$\hat{H} = H_o + \lambda \hat{V} \qquad (2.15)$$

The perturbation, $\lambda \hat{V}$, is assumed to be small. The wave function and the energy are then expanded as a power series in λ. i.e.

$$\psi_\lambda = \psi_o + \lambda\ \psi_1 + \lambda^2 \psi_2 + \ldots + \lambda^n \psi_n \tag{2.16}$$

$$E_\lambda = E_o + \lambda\ E_1 + \lambda^2 E_2 + \ldots + \lambda^n E_n \tag{2.17}$$

Where ψ_o and E_o are the HF wave function and energy respectively. Also *n* represent the order of perturbation correction applied to the initial HF wave function. It is instructive to note that the MP1 energy is the same as the HF energy. MP methods are classified according to the point at which the power series is truncated i.e. MP*n* is obtained by truncating the expansion at order λ^n. By taking the perturbation to higher orders, the solutions to the Schrodinger equation can be systematically improved. Usually, the perturbation can be of second order (MP2), third order (MP3) or even higher orders (MP4 or MP5). This method is computationally expensive particularly for higher orders and even for small molecules as several gigabytes of disk space for temporary storage of integrals are required. Fortunately, this is compensated for by improved accuracy of results obtained. MP2 is the simplest treatment of electron correlation possible. Therefore, it is the most commonly used theoretical model chemistry of all MP methods. It is capable of recovering about 80-90% of the correlation energy per electron pair on average. However, we cannot achieve chemical accuracy with MP2 unless additional improvement due to cancellation of errors is inherent in the calculation. Although the perturbation series can be carried to higher order, it is generally preferable to seek a self-consistent treatment of electron pair correlations via other methods as the series may diverge for large orders.

2.4.5 Configuration Interaction Methods

The HF wave function is usually referred to as a single determinant method since it involves a single determinant that is made up of the product of occupied molecular orbitals. Configuration interaction methods use a linear combination of various configurations of the HF wavefunction to obtain a better approximation of the molecular wave function. The various wave function configurations are obtained by exciting one or more electrons from occupied to virtual orbitals of the HF molecular orbitals. It should be noted that from solving the HF equations, we have a set of occupied MO's, comprising the HF determinant, and a set of unoccupied MO's. Keeping in mind the condition that the exact many–Fermion wave function describing a molecule must be antisymmetric just like the HF determinant, we can represent this wave function as a mixture or linear combination of determinants. If we mix together all possible determinants (i.e. all possible orbital occupancies) and variationally determine the mixing coefficients, we will obtain the exact solution to the Schrodinger equation within a given AO basis. This is the full configuration

interaction (FCI) solution and it gives (at least in principle) the exact description of electron correlation within a chosen basis set.[16] Unfortunately, FCI is computationally unfeasible for many molecular systems in chemistry; hence we must seek approximations such as CISD, CCSD and CCSD(T) methods.

Beginning from the solution to the HF equations, the MP2, CCSD, and CCSD(T) methods form a hierarchy of size-consistent correlation treatments that introduce successively more refined (and computationally demanding) treatments of electron correlation.[16] The inherent disadvantage of these methods is that while they yield chemical accuracy for the potential energy surface in the vicinity of equilibrium geometries, they cannot in general describe reaction coordinates for bond breaking to the same level of accuracy. This is because the starting point, the HF determinant, is not suitable for modelling bond breaking reactions. In order to correct for the inherent deficiency of configuration interaction methods, one would have to use methods which are based on multiple determinants necessary to describe appropriately the dissociation of a molecule into atoms or fragments. Such techniques are classified as multiconfigurational self-consistent-field (MCSCF) methods.

Developing useful solutions to the electron correlation is one of the most important achievements of modern electronic structure theory.[16,17] Despite this, there is still ample space for improvement in order to arrive at a point of feasible implementation for most of these methods. For instance methods such as HF, MP2, CCSD, and CCSD(T) are unable to give an accurate description of global potential surfaces despite their success at equilibrium geometries. Even at equilibrium, achieving chemical accuracy requires very large AO basis sets, which coupled with the prohibitive cost of the calculations means they can only be applied to relatively small molecules at present.

2.4.6. Density Functional theory

Density functional theory methods are, similar to other ab initio methods, aimed at solving the Schrodinger equation. However, unlike these methods, DFT is not formulated within the framework of a conventional wave function. Rather the electron density which can be represented as $\rho(\mathbf{r})$ is taken as the fundamental quantity.[1,7,8] Essentially, the goal of all DFT methods is to obtain information about properties of atoms and molecules from the electron density. DFT methods have become the most popular electronic structure methods for computational chemists due to three unique properties of the electron density. Firstly, in contrast to the wave function, the electron density is measurable experimentally, e.g. by X-ray diffraction

or electron diffraction. Secondly, it is readily understood intuitively by chemists, since it depends only on atomic positions i.e. three variables (x,y,z). Lastly, the electron density is easier to manipulate mathematically, hence its use as an attractive alternative to the wave function. In contrast, the wave function is a function of 4n variables; three spatial coordinates and one spin coordinate for each electron. Consequently, the main advantage of DFT is that one can obtain results of comparable quality to MP2 calculation in about the same time needed for an HF calculation of the same property for a given molecule.

DFT is based on the proof that the ground-state electronic energy is determined completely by the electron density ρ. There are two fundamental theorems put forward by Hohenberg and Kohn about 50 years ago which underpin all DFT methods[16,17]:

(i) The first Hohenberg-Kohn theorem states that the electron density uniquely determines the ground state energy and all properties of a molecular system.

$$E_\circ = F[\rho_\circ] = E[\rho_\circ] \tag{2.18}$$

(i) The second theorem states that any trial electron density function will give an energy higher than (or equal to, if it were exactly the true electron density function) the true ground state energy.

$$E_v[\rho_t] \geq E_\circ[\rho_\circ] \tag{2.19}$$

where ρ_t is a trial electronic density and $E_o[\rho_o]$ is the true ground state energy, corresponding to the true electronic density ρ_o. The trial density must satisfy the conditions $\int \rho_t(r) = n$ where n is the number of electrons in the molecule (this is analogous to the wavefunction normalization condition; here the number of electrons in all the infinitesimal volumes must sum up to the total number in the molecule. and $\rho_r \geq 0$ for all r (i.e. the number of electrons per unit volume cannot be negative).

Research within the DFT field stems from the fact that the first theorem does not tell us what is the functional dependence of the energy on the electron density function. It only assures us of the existence of such functional for a molecule in its ground state. Hence the aim of DFT methods is to design accurate functionals connecting electron density with the energy. An important milestone in this regard was achieved in 1965 by Kohn and Sham when they observed that it was possible to largely bypass the difficulty associated with constructing a kinetic energy density functional by an ingenious reformulation of DFT. This reformulation is the well-known Kohn-Sham density functional theory and is the framework for virtually all current density functional theory methods.

The two main fundamental ideas behind Kohn-Sham approach are (i) the molecular energy can be expressed as a sum of energy terms, only one of which, a relatively small term, involves the "unknown" functional. This implies that even moderately large errors in this term will not introduce large errors into the total energy and (ii) we can use an initial guess of the electron density ρ in the KS equation (analogous to the HF equations) to calculate an initial guess of the KS orbitals and energy levels; this initial guess is then used to iteratively refine this orbitals and energy levels in a manner similar to that used in the HF SCF method. The final KS orbitals are used to calculate an electron density that in turn is used to calculate the energy. Taking a hint from the fact that the HF single determinant wave function is exact for the simple albeit unphysical case of electrons that do not interact due to the absence of electron correlations, and the fact that the HF expression in this case in terms of orbitals is also exact, Kohn and Sham proved that it is possible to construct an artificial reference system of *noninteracting* electrons which has exactly the same electron density as the real molecular system of interacting electrons. Hence, the kinetic energy of a molecular system can be approximated as that of a noninteracting reference system and can be precisely evaluated in terms of the Kohn-Sham orbitals. The difference between the kinetic energy of the real and noninteracting system which is rather small is then incorporated within the exchange-correlation (XC) functional. In implementing the Kohn-Sham approach, the first step is to decompose the electronic energy of our molecule into a portion which can be calculated accurately without using DFT, and a relatively small term which requires the elusive functional. To achieve this, a fictitious non-interacting reference system of electrons whose ground state electron density distribution given by ρ_r is exactly the same as that in our real ground state system (ρ_o) i.e. $\rho_r = \rho_o$ is defined. The non-interacting electrons are readily treated exactly, and the deviations from the real behaviour of electrons are incorporated into a small term involving a functional which we have to unravel.

For a given molecule, the ground state electronic energy of the real molecule is the sum of the electron kinetic energies, the nucleus-nucleus attraction potential energies, and the electron-electron repulsion potential energies respectively:

$$E_\circ = \langle T[\rho_\circ] \rangle + \langle V_{Ne}[\rho_\circ] \rangle + \langle V_{ee}[\rho_\circ] \rangle \qquad (2.20)$$

After a series of mathematical derivations and manipulations[1], equation (2.20) can be expressed as:

$$E_\circ = \int \rho_\circ(r) v(r) dr + \langle T[\rho_\circ] \rangle_{ref} + \frac{1}{2} \iint \frac{\rho_\circ(r_1)\rho_\circ(r_2)}{r_{12}} dr_1 dr_2 + \Delta\langle T[\rho_\circ] \rangle + \Delta\langle V_{ee}[\rho_\circ] \rangle \qquad (2.21)$$

The last two "delta terms" in equation (2.21) summarize the main problem with DFT and when combined together give the so-called exchange-correlation energy. The exchange-correlation energy is the sum of the kinetic energy deviation from the reference system and the electron-electron repulsion energy deviation from the classical system. In essence, the exchange-correlation term accounts for all quantum-mechanical effects resulting from a dynamic wave-particle description of the electron. For each of the terms an unknown functional transforms electron density into an energy, kinetic and potential respectively. This exchange-correlation energy is a functional of the electron density function and is expressed as:

$$E_{XC}[\rho_\circ] \equiv \Delta\langle \mathrm{T}[\rho_\circ]\rangle + \Delta\langle \mathrm{V_{ee}}[\rho_\circ]\rangle \tag{2.22}$$

The $\Delta\langle \mathrm{T}\rangle$ term is the kinetic correlation energy of the electrons and $\Delta\langle \mathrm{V_{ee}}\rangle$ is the potential correlation and exchange energy term. It should be noted however that the exchange and correlation energy in DFT do not have exactly the same significance as in HF theory.

Hence, equation (2.21) can be expressed as

$$E_\circ = \int \rho_\circ(\mathrm{r})\nu(\mathrm{r})\mathrm{dr} + \langle \mathrm{T}[\rho_\circ]\rangle_{\mathrm{ref}} + \frac{1}{2}\iint \frac{\rho_\circ(r_1)\rho_\circ(r_2)}{r_{12}} dr_1 dr_2 + E_{XC}[\rho_\circ] \tag{2.23}$$

Subsequently a Kohn-Sham-DFT calculation proceeds in a similar fashion to a HF calculation in that one iteratively solves for the KS orbitals which yield a self-consistent field and an associated electron density that minimizes the DFT energy. The KS orbitals are one-electron functions which are constructed in a similar fashion as the one-electron orbitals in HF theory – through a Slater determinant and with a Hamiltonian operator for the non-interacting system. This Hamiltonian operator is defined as:

$$\mathrm{H_s} = -\frac{1}{2}\sum_{\mathrm{i}}^{\mathrm{N}} \nabla^2 \mathrm{i} + \sum_{\mathrm{i}}^{\mathrm{N}} \mathrm{V_s(r)} \tag{2.24}$$

Where $V_s(r)$ is the effective, local potential of the non-interacting system and does not include any electron-electron interactions. Connecting the non-interacting system to the real system is achieved by selecting $V_s(r)$ in such manner that the electron density of the non-interacting system is equal to the true ground state density. Note that the electron density is constructed in the same way as implemented in the HF scheme – by summation of the square modulus of each one-electron function. Under the usual constraint of orthonormality for the one-electron coefficients, the final energy is then minimized. With the exception of $E_{xc}[\rho]$, each term in equation (2.23) can be calculated explicitly, with equations that are similar to the ones used in HF theory. It is also instructive to note that compared to the HF situation where the Slater determinant composed

of one-electron functions is used to approximate the wave function, the one-electron function Slater determinant is in principle exact for DFT methods.

From this point onwards, DFT methods are distinguished from another based on the level of accuracy of their exchange-correlation functionals in describing the exchange correlation contributions to the energy of a molecular system. There exists a myriad of DFT methods ranging from those based on a uniform electron gas, with local approximations for the exchange energy to the very sophisticated and complicated functionals that involve multiple empirical parameters. Surprisingly, the reason for the success of some of the most commonly used functionals cannot be explained. A class of functionals which have achieved tremendous success, due to their relatively cheap computational cost and high accuracy for generalized purposes, are the so called hybrid functionals. These class of functional combine HF exchange with one or more exchange and correlation functionals in a weighted fashion. A good example is the Becke's three parameter hybrid functional combined with Lee-Yang-Parr correlation (B3LYP) which we have used extensively in this study with exceptional results.

Generally, all density functional methods can be classified broadly in order of increasing sophistication under the following families (i) Local density approximation (LDA) (ii) Local spin density approximation (LSDA) (iii) generalized gradient approximation (GGA), (iv) meta-GGA (v) hybrid GGA or adiabatic connection (ACM) methods (vi) hybrid meta-GGA (hybrid MGGA) methods, and (vii) fully nonlocal theory. Interestingly, an analogy of the biblical ladder reaching up to heaven has been used to aptly describe the hierarchy of DFT methods. It is the ultimate goal of DFT developers that this DFT Jacob's ladder will one day climax in what has been enthusiastically tagged "the divine functional". Some of the properties that this divine functional must possess have been enumerated on theoretical ground.[1] The interested reader can consult the paper by Perdew and co-workers[18] in order to gain more insight into the various families of functionals, their strengths and weaknesses.

The B3LYP functional which was used extensively in this work is a classic example of the Hybrid GGA functional. These are functionals to which the Hatree-Fock exchange has been added. Hybrid functionals are classified according to the percentage of HF exchange energy used in them. They are developed based on the adiabatic connection method which suggests that the exchange-correlation energy $E_{xc}[\rho]$ can be taken as a weighted sum of the DFT exchange-correlation energy and HF exchange energy. B3LYP is the most popular and successful hybrid method. In fact it has been pointed out in an extensive comparison of DFT methods[1] that B3LYP still remains a valid and efficient functional for tackling typical quantum chemistry problems.

There is no doubt that DFT has become a method of choice and a very efficient tool for solving many problems in electronic structure theory with an accuracy roughly comparable to the MP2 method. Also, due to the avoidance of an explicit construction of molecular wave function, the basis set requirements for DFT are far more modest than those needed to obtain reliable results via the MP2 method. In addition, for a given basis set, the computational demands of KS-DFT methods are significantly lower than that of conventional correlation methods. Instead, they are roughly equivalent to those of HF calculations. Despite all it successes, DFT theory has a few shortcomings. Firstly it is not regarded as an ab initio electronic structure theory because the correct mathematical form of the DFT functional is not known. Secondly unlike wave function methods which can be improved in a conceptually straightforward way by going to bigger basis sets and higher electron correlation levels, there is so far no rational or systematic way to improve the performance of a given functional. Hence most improvements result from intuition and comparison of results with experiment or high-level conventional ab initio calculations.

2.5. Thermodynamic Properties

All the various electronic structure theory methods that we have described thus far are only able to describe the structure and energy which are microscopic properties of molecules i.e. the electronic energy obtained from such calculations is for a single molecule. However, most of the properties of matter we are interested in as chemists e.g. rate constants, thermochemical and kinetic properties are macroscopic in nature (i.e. they depend on an immensely large number of molecules). Fortunately, this macroscopic information can be accessed from the results of electronic structure calculations using statistical mechanics techniques.[8] Vibrational frequencies are an important ingredient of such derivations. The theoretical evaluation of harmonic vibrational frequencies is efficiently done in modern electronic structure programs by evaluation of analytic second derivatives of the total energy with respect to the Cartesian coordinates.[6] Alternatively, if the second derivatives are not available analytically, they are obtained by numerical differentiation of analytic first derivatives (i.e. by evaluating gradient differences obtained after finite displacements of atomic coordinates.

In applying statistical mechanics to a collection of molecules, it is assumed that certain macroscopic conditions will be held constant by external influence. The specification of these conditions defines an ‘ensemble’. For most practical statistical applications, the canonical ensemble is usually utilized.[8] Here, the constants are the total number of particles *N* (identical molecules in this case), the volume *V*, and the temperature *T*. This ensemble is typically depicted

as the (*N*,*V*,*T*) ensemble. The relationship between a microscopic property and a given macroscopic quantity is obtained using the partition function. This is a fundamental function that describes a macroscopic system just as the Schrodinger equation characterizes the microscopic system in quantum mechanics. It is written as:

$$Q(N,V,T) = \sum_i e^{-E_i(N,V)/K_B T} \tag{2.25}$$

where *i* is summed over all possible energy states of the system having energy E_i and K_B is Boltzmann's constant. It is clear from the above that Q is a complex many-body function hence a number of simplifying assumptions are made in order to access necessary information from the partition function. The most prominent of this is that our ensemble is an ideal gas. The interested reader is directed to the excellent white paper on thermochemistry documented by Joseph Ochterski of Gaussian Inc.[19] for a detailed explanation of how thermochemical quantities such as enthalpy, entropy and free energy are obtained from the electronic structure calculations by employing the partition function.

2.6. Solvation Models

Typically, electronic structure calculations are performed in vacuum (gas phase) despite the fact that in reality, most chemical reactions of interest especially those involving biological processes actually take place in the presence of solvent. Hence there has been serious questions about calculations that ignore solvent effects when biomolecules are involved.[1] An in-depth appraisal of literature would lead to the conclusion that for some chemical processes, gas phase computations are most suitable, while for others, taking into account solvation effects is very crucial to obtain any meaningful results. Consequently, in any theoretical treatment of molecules such as the one in this research work where solvent effect play a major role, an appropriate choice of solvation model must be made out of all the various solvation models that have been implemented in commercial softwares.

A pragmatic way of making this choice is to examine the purpose of the theoretical investigation in question. In situations where the purpose of a calculation is to probe the intrinsic properties of a molecule, or of a phenomenon that depends on isolated molecules, then it is better to exclude solvation effects. On the other hand, if our study involves the determination of a particular phenomenon or property (e.g. reaction rates, equilibria, and molecular conformations) that depends on the interaction of the molecule with water molecules, then it would be important to include solvent effects in every aspect of the calculation.

The three major ways of accounting for solvation effects in QM calculations are implicit, micro- and hybrid solvation. In implicit solvation (or continuum solvation) the solute molecule is placed in bulk solvent described by a dielectric continuum medium which simulates a myriad of solvent molecules. Microsolvation on the other hand involves placing between one and six solvent molecules around the solute molecule and using this aggregate as input structure for the electronic structure and Gibbs free energy calculations. Hybrid solvation combines the strength of both implicit and microsolvation techniques by surrounding the solute molecule with a number of solvent molecules and then placing this "supermolecule" in a dielectric continuum to simulate bulks solvent properties. The explicit solvation approach is completely different in that the solute is immersed in a periodic box of solvent molecules and free energy simulations are usually performed using classical force fields and newton equations of motion to determine how the properties (i.e. energy, structure and interactions) of the system change with time. This is usually referred to as molecular dynamics simulations. More recently, there has been an upsurge in the development of QM/MM methods to account for bulk solvation effects in a more realistic way.[20] This involves partitioning the molecular system into two regions; the MM region which is referred to as the environment and the QM region which is referred to as the model system. Due to efficiency and low computational cost of MM, a large number of explicit water molecules could be used in such cases to model bulk solvent effects while the solute with the first solvent shell is treated using the more expensive but accurate QM methods. The major challenge with this methods is how to define an effective QM/MM region boundary such that the system will be able to detect the movement of a solute molecule from the MM to the QM region or vice-versa and then treat it accordingly. A lot of research activity is on-going toward achieving this objective as reported in recent publications.[20]

The dielectric continuum solvation models consider the solvent as a uniform polarizable medium with a dielectric constant of ε, with the solute placed in a suitably shaped cavity in the medium. Continuum solvation models are referred to as implicit because a continuum is used to imply the presence of individual solvent molecules.[21–23] To implement this in commercially available codes, the solute is placed in a cavity immersed a dielectric medium and the interaction between the cavity and the solute is investigated. The simplest cavity for a solute molecule is usually designed to have a spherical shape. Due to the fact that a vast majority of molecules are not spherical, recent continuum models design the cavity in such a way that it matches the shape of the molecule. Basically, the molecular shape is determined by considering the exposed surface of an overlapping spheres molecular model. However due to the fact that V-shaped crannies formed between the overlapping spheres are inaccessible to solvent, more advanced

implementations involving the smoothing of the overlapping spheres surface by projecting it onto a large number of polygons or tessellations has been implemented in some solvation models e.g. the COSMO solvation model of Klamt and coworkers.[24–25] The use of a continuum instead of individual solvent molecules helps to average out the effect of a large number of solvent molecules. This eliminates the need for statistical analysis based on delicate sampling averages as required with microsolvation.[26] Hence dielectric continuum calculations are the easiest and most popular way of treating solvent effects in computational chemistry.

Examples of the common solvation models implemented in commercial ab initio quantum chemistry programs include:

(i) Polarizable Continuum Model: This model was originally introduced by Tomasi and co-workers to describe the apparent surface charge (ASC) approach developed in Pisa.[21–23] As at today, all family of continuum solvation models that use the apparent surface charge boundary condition formulation (e.g. IPCM, SCIPCM, IEFPCM, CPCM) are all classified as PCM models. Each of these variants of the original PCM model differ in their approaches one with respect to another, in terms of the electrostatic expression used to define the ASC density. Each of these alternative formulations has its own strengths and weaknesses documented in literature[21] which can be used as guide in choosing the appropriate solvation model for particular applications. PCM methods have the advantage over microsolvation of being able to account simultaneously for long-range electrostatic interactions and polarization effects.[1,22]

(ii) COSMO Models: Similar to other dielectric continuum models COSMO solvation model also represents the solvent as a dielectric continuum. The solute is placed in a molecular cavity which is then placed in the dielectric continuum. Hence, COSMO is in principle an efficient alternative of the apparent surface charge dielectric continuum solvation models[24] (e.g. PCM models). However, the COSMO model is distinct from these apparent surface charge models (ASMs) in that it uses a scaled conductor boundary condition to calculate dielectric screening charges and solvation free energy.[24] This is in contrast to the dielectric boundary condition utilized in all other dielectric continuum models. The scaled boundary condition is much more efficient compared to the solution of the dielectric boundary conditions. It has been argued that COSMO is superior to the ASM continuum solvation models due to a number of reasons.[24,25,27] For instance, while COSMO is able to distinguish between two solvents with essentially identical dielectric constants but quite different solvent properties e.g. cyclohexane and benzene, dielectric continuum solvation

approaches cannot.[25] There are also various modifications of this model such as COSMO, COSMO–RS and DCOSMO–RS and they differ from one another in the details of their cavity design and intended applications.[27]

(iii) SM*x* models: These models are based on a generalized Born approximation for the bulk electrostatic contribution to the free energy of solvation.[28,29] This is an approximation of Poisson's equation suitable for arbitrary cavity shapes. The SM*x* model solves the Poisson-Boltzmann equation in a similar fashion as PCM models, but does so using a set of specifically parameterized radii to model the cavity. All SM*x* (*x* contains information about the version and peculiarities of the model) are based on discrete partial atomic charges. Recently, Cramer and Truhlar have developed the SM8 solvation model as a universal model that can be used to account for both aqueous and non-aqueous solvation as well as treat neutral and charged solutes on the same footing.[28] It has been pointed out that this is the most accurate continuum model till date.

2.7. Bonding and Wavefunction Analysis

In the previous sections, we have discussed the basic theory behind electronic structure methods used to obtain the wave function, energy, and electron density distribution of molecules. However, the need to derive chemically relevant information from the results of such calculations has led to the development of various interpretative techniques which are either wave function based or electron density based. Most of these techniques are based on the topology of the electron density and have been grouped together under the umbrella of Quantum Chemical Topology methods.[30,31] These QCT techniques such as QTAIM, ELF and NCI are popular within the computational chemistry community. Their popularity is as a result of the chemically intuitive nature of the electron density and the fact that it can be observed experimentally. In addition, electron density methods are based on three dimension of atomic position in contrast to wave function based methods which depend on four dimensions and are therefore difficult to understand conceptually. Moreover, QCT techniques are generally independent of the level of theory used in the electronic structure calculations. A major disadvantage of these methods is their inability to obtain information about excited states hence limiting their use for studying molecular reactivity.[32] In the present study, we have used QTAIM and NCI methods to characterize bonding interactions; hence the following brief discussion of these methods.

2.7.1. Quantum Theory of Atoms in Molecules

QTAIM uses a topological analysis of the gradient vector field of the electron density to partition a molecule into its constituent quantum atoms, each of which can be described by the physics of an open system, and to define the bonding environments of the atoms in a molecule.[31,33–36] The origin of the electron density distribution function does not matter: it can be obtained both from the quantum mechanical calculations and on the basis of the high-resolution X-ray diffraction data.[34] The latter fact is of particular interest to experimental chemists (especially X-ray crystallographers) as it provides an avenue for a direct comparison of the QTAIM calculations with experimental results. The properties of each atom can then be computed by integrating over each atomic basin. This theory is based on the premise that atoms in molecules in a sense retain their identities rather than dissolving into a molecular pool of nuclei and electrons.[1] Hence it directs focus of chemists to atoms rather than bonds as advocated by the molecular orbital theory. It defines an atom as an open system (Ω) which is bounded by a surface S(Ω;r) of zero-flux in the gradient vector field of the electron density $\rho(\mathbf{r})$,

$$\nabla\boldsymbol{\rho}(\mathbf{r})\bullet\mathbf{n}(\mathbf{r})=\mathbf{0} \qquad (2.26)$$

for all r on the surface. This definition makes it possible to evaluate the properties of an atom in a similar fashion as that of a molecule. For any given molecular property, summation of the individual atomic contributions to this property should give the molecular value for this property hence making it possible to find out the contributions of each atom to a given experimental observable quantity such as molecular charge.

By evaluating $\nabla\rho(r)$ at any given point in a scalar field ρ(r), then following this vector for an extremely small distance and evaluating $\nabla\rho(r)$ again, we can define a piecewise continuous gradient path in the scalar field. In the limit of an infinitely small step, the path is continuous and corresponds to the true gradient path. Proceeding from each of the local maxima in electron density (usually nuclear positions), one can follow all possible gradient descent paths until we arrive at a spatial location where $\nabla\rho(r)=0$. With that we can define the enclosing atomic basin as the region enclosed by all such paths emanating from the corresponding maximum. Inter-atomic surfaces which satisfy the zero-flux condition $\nabla\boldsymbol{\rho}(\mathbf{r})\bullet\mathbf{n}(\mathbf{r})=\mathbf{0}$ can then be defined, separating adjacent basins. The gradient vector field consists of gradient paths which originate at infinity, moves along the path of steepest ascent and terminate at the nuclear position which is a local maximum. In essence, a gradient path is a trace of the path of steepest ascent from the lowest to the highest electron density point in a molecule.[31] A collection of atomic basins and

all such lines connecting them define a given molecular structure. There are certain points within the gradient vector field where more than one gradient path meet, such points are referred to as critical points and they have the important property that $\nabla\rho(r) = 0$, i.e. at these points, the first derivatives of $\rho(\mathbf{r})$ vanish.

The structure and topological properties of a multielectron system are uniquely and conveniently summarized in terms of the number and kind of its critical points (CPs). Analysing the properties of the Hessian matrix of $\rho(\mathbf{r})$ at each CP allow us to differentiate between various types of critical points. A diagonalization of the Hessian matrix of $\rho(\mathbf{r})$ gives the ordered eigenvalue set $\lambda_1 < \lambda_2 < \lambda_3$, with the Laplacian of the electron density being the algebraic sum of these eigenvalues (i.e. the trace of the Hessian). Each critical point is described by its rank (ω) and signature (σ) and is conventionally labelled using the notation (ω, σ), where ω is the rank (the number of non-zero λ_i and (σ) is the signature (the algebraic sum of the signs of each λ_i. The rank of a critical point is equal to the number of non-zero eigenvalues or non-zero curvatures of ρ at the critical point. The signature, denoted by σ, is simply the algebraic sum of the signs of the eigenvalues, i.e. of the signs of the curvatures of ρ at the critical point. Using this notation, there are four topologically stable types of critical points for $\nabla\rho(r)$;

- (3, –3), All curvatures are negative and $\rho(\mathbf{r})$ is a local maximum at $\mathbf{r_c}$. These are generally found only at the positions of the nuclei. The nuclei acts as the attractors of the gradient vector field of $\rho(\mathbf{r})$. This is generally associated with a nuclear critical point.
- (3, –1), Two curvatures are negative and $\rho(\mathbf{r})$ is a maximum at r_c in the plane defined by their corresponding axes. $\rho(\mathbf{r})$ is a minimum at r_c along the third axis which is perpendicular to this plane. A (3, –1) critical point is usually found between every pair of nuclei which are considered to be linked by a chemical bond. It is a bond critical point (BCP), found between interacting atoms, is the saddle-point in the "ridge-line" in the electron density between interacting atomic basins.
- (3, +1), Two curvatures are positive and $\rho(\mathbf{r})$ is a minimum at r_c in the plane defined by their corresponding axes. $\rho(\mathbf{r})$ is a maximum at r_c along the third axis which is perpendicular to this plane. A (3, +1) critical point is usually found at the centre of bond paths which are linked so as to form a ring of bonded atoms. It is generally found in near the centroid of a ring structure and is named a ring critical point (RCP).
- (3, +3), All curvatures are positive and $\rho(\mathbf{r})$ is a local minimum at r_c. This occurs in the interior of a cage resulting from the arrangement of bond paths in such a way that they

enclose the interior of a molecule with ring surfaces. The charge density is a local minimum at a cage critical point. A cage critical point (CCP) is the intersection point of all the atomic basins for atoms which together form a cage structure.

Within the gradient vector field, certain lines of maximum electron exist, which connects atom pairs. These pair of special gradient paths linking a BCP with two nuclei and along which $\rho(\mathbf{r})$ is a maximum with respect to any neighbouring path is known as an atomic interaction line (AIL). A bond path is then said to exist between two nuclei linked by an AIL when the forces on the nuclei vanish (i.e. the structure is in equilibrium or at an energetically optimum geometry). QTAIM therefore provides a universal definition of what constitutes bonding; if a bond-path exists between two nuclei, they are considered to be bonded. Bond-paths are usually linear, corresponding to the traditional definition of a chemical bond, but in electron-deficient structures and /or easily ruptured bonds, they may be significantly curved.[36] All gradient lines of this field beginning at infinity end in CP (3, –3), therefore such points (i.e. atomic nuclei) are regarded as attractors of the vector field. All gradient trajectories ending on a nucleus in CP (3, –3) define the spatial region of a chemical system called the basin of this nucleus. The unity of the nucleus and electron density within the boundaries of its basin defines both free and bound atom. Among all

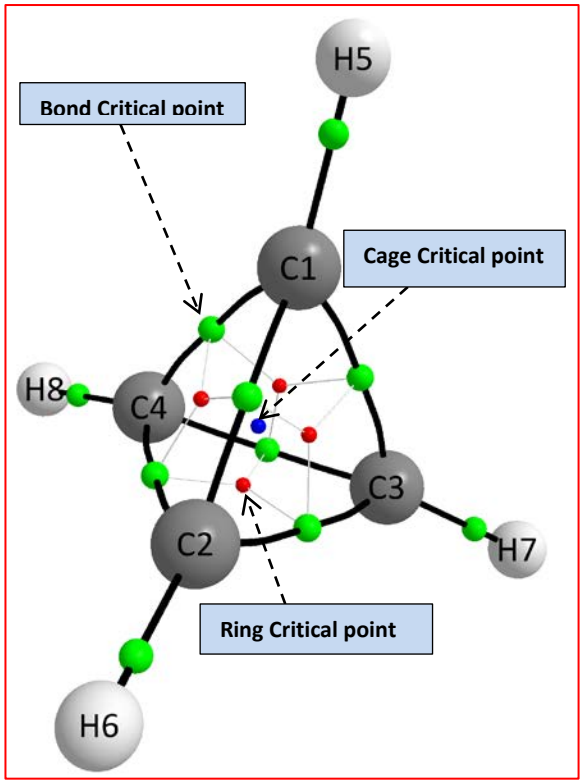


Figure 2.1. QTAIM molecular graph of tetrahedrane indicating bond paths (thick dark lines), bond critical point (green spheres), ring critical point (red spheres) and cage critical point (blue sphere).

gradient trajectories, those that start in CP (3, –1) and end in CP (3, –3) are of special importance. These gradient trajectories define a line along which $\rho(\mathbf{r})$ attains maximum values in respect to any nearest horizontal point. The presence of such gradient trajectories suggests the accumulation of $\rho(\mathbf{r})$ in the interatomic space and is a necessary and sufficient condition of the existence of a chemical bond. That is why the saddle CP (3, –1) and gradient lines passing through this point and CP (3, –3) were christened bond critical point and bond path respectively.

Additionally, the total local energy density at the BCPs ($H(r_b)$), is defined as:

$$H(\mathbf{r}_b) = G(\mathbf{r}_b) + V(\mathbf{r}_b) \tag{2.27}$$

where $G(\mathbf{r}_b)$ and $V(\mathbf{r}_b)$ are the local kinetic and potential energy densities at the BCP, respectively. $H(\mathbf{r}_b)<0$ indicates a bond-path with a degree of covalent character, and conversely $H(\mathbf{r}_b)>0$ reveals a lack of covalent character for the closed-shell interaction, the latter being the case for the very weakest types of interactions, such as van der Waals interactions and hydrogen-bonds. All the above described QTAIM bonding descriptors are only valid at equilibrium geometries and are succinctly summarized in the molecular graph.

2.7.2. Non-Covalent Interaction (NCI) method

The non-covalent interaction index method[37] is based on a graphical interpretation of the electron density (ρ) and its reduced gradient $s(\rho)$ (which is also a derivative of the electron density) to identify regions of space where the electronic density distribution deviates from homogeneity due to the formation of an inter or intramolecular interaction.[38] The reduced density gradient is defined as:

$$s = \frac{1}{2(3\pi^2)^{1/3}} \frac{|\nabla\rho|}{\rho^{4/3}} \tag{2.28}$$

Typically these non-covalent interaction regions occur far from the nuclei and are revealed by singularities in the electron density. In addition, by using the sign of the second eigenvalue (λ_2) of the electron density Hessian, NCI analysis is able to classify interactions as stabilizing or destabilizing according to the topology of electron density distribution. However it has been noted recently in literature[38] that this technique should be used cautiously in regions of high electron density e.g. in assessing the (de)stabilizing character of intramolecular interactions, as by definition it is designed to only explore regions where both electron density and its reduced gradient are weak. Specifically, it was pointed out that if λ_2 is too close to zero (typically $|\lambda_2| <$

0.0001a.u), its sign is spatially unstable and one should not rely on it as a means of classifying an interaction as (de)stabilizing.

NCI information is usually presented with the aid of both 2D and 3D plots. The 2D plot shows variation in reduced density gradient as a function of the electron density oriented by the sign of its λ_2. Within this plot, the presence of an intramolecular interaction is revealed by the appearance of a spike/trough in regions of low electron density. To visualize the isosurfaces that results from these interactions, a 3D plot is used. The 3D plot enables the display of resulting isosurfaces as closed regions in the molecular space thereby revealing the spatial location of inter or intramolecular interactions and their nature within a molecular system in three dimensions.[37,38] In order to rank interactions according to their strength an RGB colour scheme is utilized. According to this scheme stabilizing interactions also referred to as type I NCI interactions are coloured in blue, type II/ destabilizing interactions in red and Van der Waals/type III delocalized weak interactions in green. The colour intensity corresponds to the strength of an interaction i.e. a greater intensity refers to stronger interaction and vice-versa.

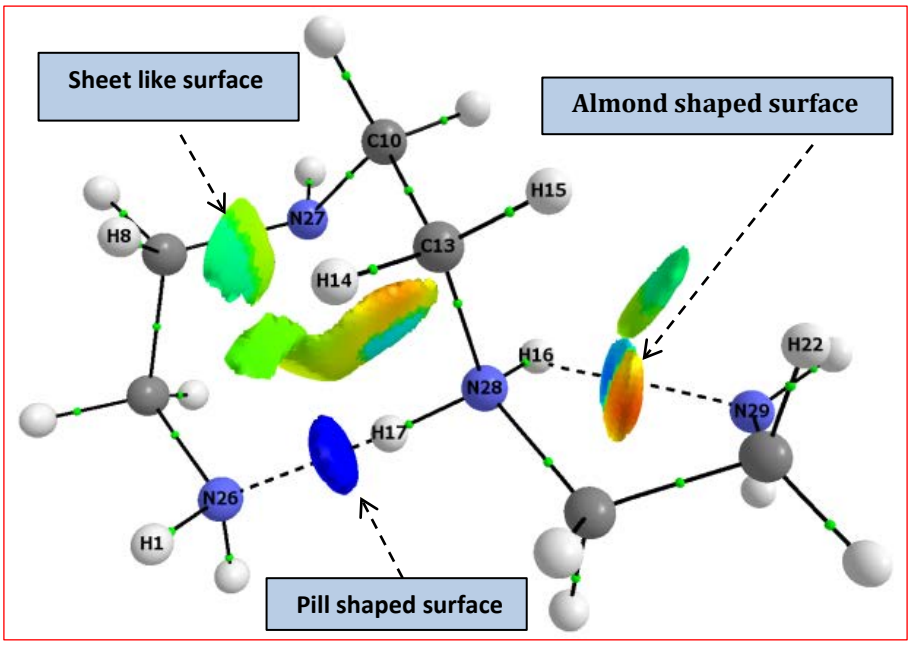


Figure 2.2. 3D NCI plot of a low energy conformer of triethylenetetramine showing the various types of NCI isosurface and their colour code (i) blue for stabilizing, (ii) red for destabilizing and (iii) green for Van der Waals interactions.

The shapes of the isosurfaces of the three main types of NCI interaction also varies according to the type of interaction involved:

- Small, flat, pill-shaped for stabilizing interactions; this has been attributed to their bicentric nature.

- Red cigar-shaped surfaces elongated along directions of increasing density for destabilizing interactions and
- Sheet-like extended surfaces for weak Van der Waals interactions.

In the case of flexible molecules such as amino acids and aliphatic polyamines, the occurrence of a non-covalent interaction often results in steric crowding and intramolecular ring formation. This feature typically appears as a bicoloured almond shaped NCI isosurface in the 3D NCI plot with one colour such as blue or green representing the stabilizing contribution and another e.g. red indicating multicentric repulsion amongst atoms involved in ring formation. NCI analysis has been demonstrated to provide a correct picture of delocalized interactions where other QCT techniques such as QTAIM and ELF fail to describe these kinds of interactions appropriately.

To identify the various intramolecular interactions present in a given conformer, we have used the NCI Isosurfaces generation feature of AIMALL software to generate 3D NCI plots starting from their MP2 electron density.

2.7.3 Interacting Quantum Atoms Method

The interacting quantum atoms energy partitioning scheme[39,40] is based on the QTAIM theory but free from its constraints of using the bond path as an indicator of interactions between atoms as well as its validity only at the equilibrium geometry of the molecule. Starting from its many electron wave function, the IQA analysis uses the first-order (non-diagonal) and second order (diagonal) density matrices to compute the expectation value of the total electronic energy (*E*) of a given molecule which is further decomposed into its separate components within the Coulomb Hamiltonian scheme. These components include the intra-atomic, also referred to as the *self-energy* (E^{A}_{self}) and the pair interaction contributions due to the interactions between all its constituent atoms ($\sum_{B\neq A} E^{AB}_{int}$). The self-energy and half of the interaction contribution terms can then be grouped and summed up together for each atom in a molecule to obtain the additive atomic energy(E^{A}_{add}) which necessarily adds up to the molecular energy (*E*).

$$E = \sum_A E^{A}_{add} \tag{2.29}$$

$$\sum E^{A}_{add} = E^{A}_{self} + \frac{1}{2}\sum_{B\neq A} E^{AB}_{int} \tag{2.30}$$

For any two interacting atoms *A* and *B*, the interaction energy, which is negative for stabilizing

and positive for destabilizing interactions according to the IQA convention, is expressed as the sum of their electron-nuclear (V_{en}, V_{ne}), nuclear-nuclear (V_{nn}) and electron-electron (V_{ee}) interactions:

$$E_{int}^{AB} = V_{nn}^{AB} + V_{en}^{AB} + V_{ne}^{AB} + V_{ee}^{AB} \tag{2.31}$$

Furthermore, the electron-electron interaction contribution (V_{ee}^{AB}) can be partitioned into classical electrostatic interaction and quantum exchange and correlation contributions:

$$V_{ee}^{AB} = V_{C}^{AB} + V_{X}^{A|B} + V_{corr}^{A|B} \tag{2.32}$$

The electrostatic terms ($V_{C}^{AB}, V_{nn}^{AB}, V_{en}^{AB}$ and V_{ne}^{AB}) can also be grouped together to obtain an expression for the total classical electrostatic interaction energy (E_{int}). Likewise, summation of the quantum-mechanical exchange ($V_{X}^{A|B}$) and correlation terms ($V_{corr}^{A|B}$) accounts gives the exchange-correlation energy contributions to a given interaction.

The intra-atomic energy of an atom consists of contributions from T^{A}, V_{en}^{AA} and $V_{\text{ee}}^{\text{AA}}$ and their sum is referred to as the self-energy of the atom i.e.

$$E_{self}^{A} = T^{A} + V_{en}^{AA} + V_{ee}^{AA} \tag{2.33}$$

Because it required a well-defined second order density matrix, the IQA method was originally designed to work with only HF wavefunctions. However upon its implementation in AIMALL software[41], recent modifications of the theory have extended its use to a number of DFT functionals such as LSDA, B3LYP and M062X.[41,42]

2.8. Conclusions

Spectroscopic methods e.g. NMR spectroscopy have become invaluable tools in the hands of experimental chemists for understanding chemical structure and reactivity. Likewise, theoretical and computational chemistry methods are also fast evolving to become indispensable means of understanding and explaining experimental results that would be otherwise difficult to interpret such as the protonation sequence of aliphatic polyamines. In this chapter we have given a brief overview of the various computational/theoretical methods that was used to investigate this problem, their strengths and weaknesses as well as inherent limitations. This should serve as a means of understanding and appraising the results of our investigations of the protonation sequence of polyamines which are presented in subsequent chapters.

2.9. References

[1] E. G. Lewars, in *Computational Chemistry:Introduction to the theory and Applications of Molecular and Quantum Mechanics*. Second Edition, **2011,** Springer Dordrecht.

[2] F. Jensen, in *Introduction to Computational Chemistry*, Second Edition, **2007**, John Wiley & Sons, Ltd.,West–Sussex.

[3] A. R. Leach, D. Schomburg, in *Molecular Modelling: principles and applications*, **1996**, Longman, London.

[4] D. C. Young, in *Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems*, **2001**, John Wiley & Sons Inc.

[5] P. R. Westmoreland, P. A. Kollman, A. M. Chaka, P. T. Cummings, K. Morokuma, M. Neurock, E. B. Stechel, P. Vashishta, in *WTEC Panel Report on applications of molecular and materials modelling* , **2002**.

[6] J. B. Foresman, E. Frisch, in *Exploring Chemistry with Electronic Structure Methods.* Second Edition, **1996**, Gaussian, Inc Pittsburgh, PA.

[7] W. Koch, M. C. Holthausen, in A Chemist's Guide to Density Functional Theory, Second Edition, **2001**,Wiley–VCH Weinheim.

[8] C. J. Cramer, in *Essentials of Computational Chemistry: Theories and Models*, Second Edition, **2004,** John Wiley & Sons, Ltd. West–Sussex.

[9] T. A. Halgren, *J. Comput. Chem.*, **1996**, *17*, 490–519.

[10] W. J. Hehre, in *A Guide to Molecular Mechanics and Quantum Chemical Calculations*, ***2003***, Wavefunction, Irvine.

[11] I. Kolossvàry, W. C. Guida, *J. Am. Chem. Soc.*, **1996**, *118*, 5011–5019.

[12] S. R. Wilson, W. Cui, *Tetrahedron Lett.* , **1988**, *29*, 4373–4376.

[13] N. Foloppe, I. Chen, *Curr. Med. Chem.*, **2009**, *16*, 3381–3413.

[14] K. Gaalswyk, C. N. Rowley, *PeerJ. 4:e2088*, **2016,** https://doi.org/10.7717/peerj.2088.

[15] P. A. M. Dirac, *Proc. R. Soc.*, **1929**, *A123*, 714–733.

[16] M. Head-Gordon, *J. Phys. Chem.*, **1996**, *100*, 13213–13225.

[17] S. Bell, T. J. Dines, B. Z. Chowdhry, R. Withnall, *J. Chem. Edu.*, **2007**, *84*, 1364–1370.

[18] J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria, G. I. Csonka, *J. Chem. Phys*., **2005**, *123*, 062201:1–9.

[19] J. W. Ochterski, *Gaussian white paper "Thermochemistry in Gaussian"*, http://www.gaussian.com/g_whitepap/thermo.htm.

[20] L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, H. Li, L. Ding, K. Morokuma, *Chem. Rev*., **2015**, *115*, 5678−5796.

[21] J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev*., **2005**, *105*(8), 2999−3094.

[22] B. Mennucci, *J. Phys. Chem Lett*., **2010**, *1*, 1666−1674.

[23] J. Tomasi, M. Persico, *Chem. Rev*., **1994**, *7*, 2027−2094.

[24] A. Klamt, *WIREs Comp. Mol. Sci*., **2011**, *9* (1), 609−620.

[25] A. Klamt, *J. Phys. Chem*. **1995**, *99*, 2224−2235.

[26] J. Ho, *Aust. J. Chem*., **2014**, *67*, 1441–1460.

[27] A. Klamt, *WIREs Comput. Mol. Sci*., **2011**, 1, 699–709.

[28] C. J. Cramer, D. G. Truhlar, *Acc. Chem. Res*., **2008**, *41*,760−768.

[29] Cramer, C. J.; Truhlar, D. G. in *Trends and Perspectives in Modern Computational Science; Lecture Series on Computer and Computational Sciences* Vol. 6; Maroulis, G., Simos, T. E., Eds.; Brill/VSP, Leiden, **2006**; pp. 112-140.

[30] P. L. A. Popelier, É. A. G. Brémond, *Int. J. Quant. Chem*., **2009**, *109*, 2542–2553.

[31] P. L. A. Popelier In *The Nature of the Chemical Bond Revisited*. **2014**, pp. 271–308, Wiley-VCH Germany.

[32] P. Cassam-Chenai, D. Jayantilka, *Theor. Chem. Acc*., **2001**, *105*, 213–218.

[33] R. F. W. Bader, Atoms in Molecules: A Quantum Theory. **1990**, Oxford University Press, Oxford, UK.

[34] I. S. Bushmarinov, K. A. Lyssenko, M. Y. Antipin, *Russian Chemical Reviews*, **2009**, *78*, 283–302.

[35] J. R. Maza, S. Jenkins, S. R. Kirk, J. S. M. Anderson, P. W. Ayers, *Phys. Chem. Chem. Phys*., 2013, 15, 17823–17836.

[36] D. Stalke, *Chem. Eur. J.*, **2011**, *17*, 9264–9278.

[37] E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen, W. J. Yang, *J. Am. Chem. Soc.*, **2010**, *132*, 6498–6506.

[38] R. Chaudret, B. de Courcy, J. Contreras-García E. Gloaguen, A. Zehnacker-Rentien, M. Mons, J. P. Piquemal, *Phys Chem Chem Phys.*, **2014**, *16*, 9876–9891.

[39] M. A. Blanco, A. M. Pendás, E. Francisco, *J. Chem. Theory Comput.*, **2005**, 1, 1096–1109.

[40] E. Francisco, A. M. Pendás, M. A. Blanco, *J. Chem. Theory Comput.*, **2006**, 2, 90–102.

[41] T. A. Keith, AIMAll (Version 13.11.04), TK Gristmill Software, Overland Parks KS, USA, 2013 (aim.tkgristmill.com).

[42] P. Maxwell, A. M. Pendás, P L. A. Popelier, *Phys Chem Chem Phys.*, **2016**, DOI: 10.1039/c5cp07021j.

Chapter 3

Structural-topological preferences and protonation sequence of aliphatic polyamines: a theoretical case study of tetramine *trien*

This chapter is essentially the published paper in *J. Mol. Model.* , **2015**, *21:162*, 1–18.

Summary

As a means of understanding the preferred sequence of protonation of aliphatic polyamines, a dedicated conformational search protocol (CSP) was developed in this section to identify relevant low energy conformations of triethylenetetramine. Using lowest energy conformers isolated with this CSP, theoretically predicted mixture of primary ($\mathbf{HL_p}$) and secondary ($\mathbf{HL_s}$) forms was found to be in accord with recent literature reports. On the contrary, using linear conformers resulted in predicting $\mathbf{HL_s}$ as the only tautomer formed.

Furthermore, a large set of lowest and medium energy conformers was used to uncover structural-topological preferences in mono- and diprotonated triethylenetetramine, which in principle, could be extended to other aliphatic linear polyamines. Numerous common structural features among **HL** and $\mathbf{H_2L}$ tautomers were identified, *e.g.*, H-atoms of protonated functional groups are always involved in stabilizing intramolecular NH•••N interactions and they result in as large and as many as possible rings in lowest energy conformers. Largest, 11-membered, rings stabilize a molecule most and they appeared to be strain free whereas 5-membered-rings were most strained (all formed due to NH•••N interactions). In addition, CH•••HC interactions with QTAIM-defined atomic interaction lines were also found for the first time in aliphatic polyamines but, surprisingly, mainly in the lowest energy conformers of **HL** tautomers.

According to the Non-covalent Interaction-based (NCI) analysis, 5-memebered rings formed by CH•••HC interactions are not strained and, in general, 3D NCI isosurfaces mimic those obtained for weaker NH•••N interactions. Also, 3D NCI isosurfaces found for NH•••N and CH•••HC interactions, regardless whether linked or not by an atomic interaction line, appeared to be indistinguishable.. In contrast to HF, the overall performance of B3LYP was found satisfactory for the purpose of the study; it reproduced MP2 results well.

3.1. Introduction

Some 'families' of organic compounds, even though they might play significant roles in many fields of chemistry or biochemistry, have not been extensively explored computationally at higher than MM-levels of theory. In many cases, this can be attributed to an almost 'infinite' number of possible conformations that higher homologues can adopt due to their enormous flexibility making this kind of investigation an extremely challenging task. A typical example might be aliphatic linear polyamines (ALPs) which are found in most living organisms and are important in the regulation of cell proliferation and cell differentiation.[1-3] Their function in living organisms is based on their ability to act as natural polycations when protonated under physiological conditions and therefore interact with natural polyanions such as DNA and RNA thereby influencing important cell functions.[3] Due to their crucial roles, among others in physiological processes, the protonation sequence of ALPs has been investigated with various experimental techniques for decades but a consensus has not been reached as to which of the nitrogen atoms present in a given polyamine is protonated in the first step.[4-12] For instance, Paoletti et al. suggested, from thermochemical studies of diethylenetriamine (*dien* or 2,2-tri) and triethylenetetramine (*trien* or 2,2,2-tet), that the first step of protonation results in the formation of an equilibrium mixture of tautomers with either the primary ($\mathbf{HL_p}$) or secondary ($\mathbf{HL_s}$) nitrogen atoms protonated.[4] Hague and Moreton, on the other hand, concluded from ^{13}C NMR studies of a wide spectrum of polyamines, ranging from 2,2-tri to tetraethylenepentamine (2,2,2,2-pent), that all the nitrogen atoms in the monoprotonated form share the proton equally.[5] Contrary to these, Delfini et al[7] proposed from ^{13}C NMR study of a series of triamines that (i) in the case of symmetrical triamines, either of the identical primary N-atoms is involved in the first protonation step for symmetrical molecules, but (ii) for asymmetrical triamines, the primary nitrogen atom attached to the longer aliphatic chain is protonated first. Recently, Borkovec et al.[8] used cluster expansion analysis on experimental NMR data to investigate a number of di-, tri-, tet- and pentamines. They concluded that both forms, $\mathbf{HL_p}$ and $\mathbf{HL_s}$, are formed in the first protonation step but $\mathbf{HL_p}$ is predominant with a %-fraction of 86%. Clearly, even in some cases, where the same experimental technique was used, such as ^{13}C NMR, different research groups proposed contradicting conclusions.[8,9]

Conformational preference of ALPs with more than two nitrogen atoms has not been investigated computationally. However, better understanding of the properties of polyamines and their activity, as observed from experimental investigations, requires knowledge of their conformational preferences.[13] Furthermore, if one could theoretically predict a reliable and representative set of the lowest energy conformers (LECs) of all protonated forms of polyamines

then this, *e.g.*, (i) should be of great aid in interpretation of experimentally recorded NMR spectra (often used to predict most likely protonation sequence) or (ii) might result in significantly improved quality of theoretically computed protonation constants, as it has been demonstrated in the case of carboxylic acids.[14,15]

Therefore, we have embarked on the development and testing of a dedicated protocol for conformational analysis of mono- and di-protonated (**HL** and $\mathbf{H_2L}$) forms of ALPs and, as a case study, results obtained for triethylenetetramine (*trien*) are reported in this work. The singly protonated form of *trien* is of particular interest to us mainly because (i) there is no consensus in the literature as to the preferred site of protonation and (ii) in the case of symmetrical ALPs, such as *trien*, the basicity of the primary and secondary N-atoms is of the same order of magnitude.[7,16] Furthermore, the apparent difficulty in characterizing the singly protonated forms of ALPs by, *e.g.*, XRD crystallography, has been a major hindrance; we were able to find only one example of the singly protonated ALP, namely that of 2,2-tri reported by Ilioudis et al.[17] We were also motivated by the fact that, except for few alicyclic diamines which were mainly studied using molecular mechanics, MM, methods[14,19-20] the conformational analysis of these important compounds, as far as we could establish, has not been reported to date. Both Raman spectroscopy and MM-based theoretical results[18-20], however, led to the conclusion that the most stable conformers are those in which either intra- or intermolecular interactions can occur depending on whether the isolated diamine was studied in the solid or aqueous phase. Also the results showed that even though the conformational barriers obtained with MM method compare very well to those from *ab initio* results for cyclic polyamines, this is not always the case for aliphatic polyamines.[19,20]

Hence, in this work apart from characterizing the lowest energy conformers of mono and di-protonated *trien*, we have also embarked on a study of their conformational preferences in terms of structural and topological properties. Unavoidably, this kind of study is computationally expensive because it involves modelling of hundreds (if not thousands) of conformers. With an attempt to suggest most economical methodology, we have also evaluated, relative to MP2, performance of HF and selected, B3LYP and B97D, DFT techniques.

3.2. Methodology

3.2.1. Computational details

Energy minimized structures and wavefunctions were generated from Gaussian 09, Revision B[21] at the HF, B3LYP, B97D and MP2 theory levels in solvent using the polarizable continuum

model[22-24] in conjunction with the UFF cavity model with water as a solvent (ε = 78.39). A scaling factor of 1.1 was used in the PCM calculations. Unless otherwise stated, the 6-311++G(d,p) basis set was used; it has been shown that this basis set contains adequate polarization and diffuse functions sufficient to describe molecules like polyamines.[25,26] General shortcomings as well as advantages of each levels of theory tested here are well-known. However, due to a formidable computational task associated with the aim of this work, it was important to find out which level of theory is best suited for the purpose. In other words, which level of theory performs sufficiently well to identify the lowest energy conformers and characterize their structural preferences without getting involved in unnecessarily time-consuming computational operations.

The Quantum Theory of Atoms in Molecules (QTAIM)[27], as implemented in AIMAll[28]. together with the recently developed non-covalent interaction (NCI) index technique[29], were used to carry out a topological analysis of the calculated electron densities in order to identify and characterize intramolecular interactions which may be responsible for conformational preference. All topological data (QTAIM and NCI) as well as molecular graphs and NCI 3D plots were generated with AIMAll. Frequency calculations were performed on fully optimized geometries to ensure that the true stationary points were found; in case of MP2, the frozen core approximation was used. No imaginary frequencies were found among (i) all MP2-optimized conformers and (ii) lowest energy conformers at all levels of theory. Few conformers of diprotonated form of *trien* with a single imaginary frequency are listed in Table A1 in Appendix A. The conformational search was done using MM force field (MMFFaq) and the Monte Carlo method as implemented in the Spartan[30] quantum chemistry package.

3.2.2. Conformational search protocol

The potential energy surface of extremely flexible biomolecules, such as aliphatic polyamines, is characterized by the presence of an enormous number of local minima, thus making an exhaustive conformational analysis impracticable, particularly if DFT was to be employed exclusively. Hence, a very efficient (for the study of polyamines) conformational search protocol involving both, MM and *ab initio* techniques was developed. One of the major goals was to identify a 'final' subset of LECs containing representative structures within a narrow (few kcal/mol, typically within 5 kcal/mol) relative energy range, $\Delta E = E_{conf} - E_{LEC}$ (E_{conf} stands for the electronic energy of a conformer). Instead of ΔG, ΔE values was used because (i) the former requires frequency calculations, an additional and extremely time

consuming operation to be avoided (if possible), (ii) this work is not concerned with protonation constants calculations (this would require ΔG values) and (iii) representative sets of polyamines should be generated at minimum computational time and, fortunately, under conditions employed the relative trends in the ΔE and ΔG values did not differ significantly enough to justify the use of ΔG values as a selection criterion.

The capped-stick representations (with numbering of atoms) of linear structures of only two possible $\mathbf{HL_p}$ (primary N-atom is protonated) and $\mathbf{HL_s}$ (secondary N-atom is protonated) tautomers used as inputs for conformational search by MM are shown in Figure. 1(a and b). Hypothetically, there are four possible tautomers of the di-protonated form, but only two of them, $\mathbf{H_2L_{ps}}$ (with primary and secondary N-atom protonated) and $\mathbf{H_2L_{pp}}$ (with both primary N-atoms protonated) were considered in this work, see Figure. 3.1(c) and (d), based on minimum charge separation constraints required by electrostatics.

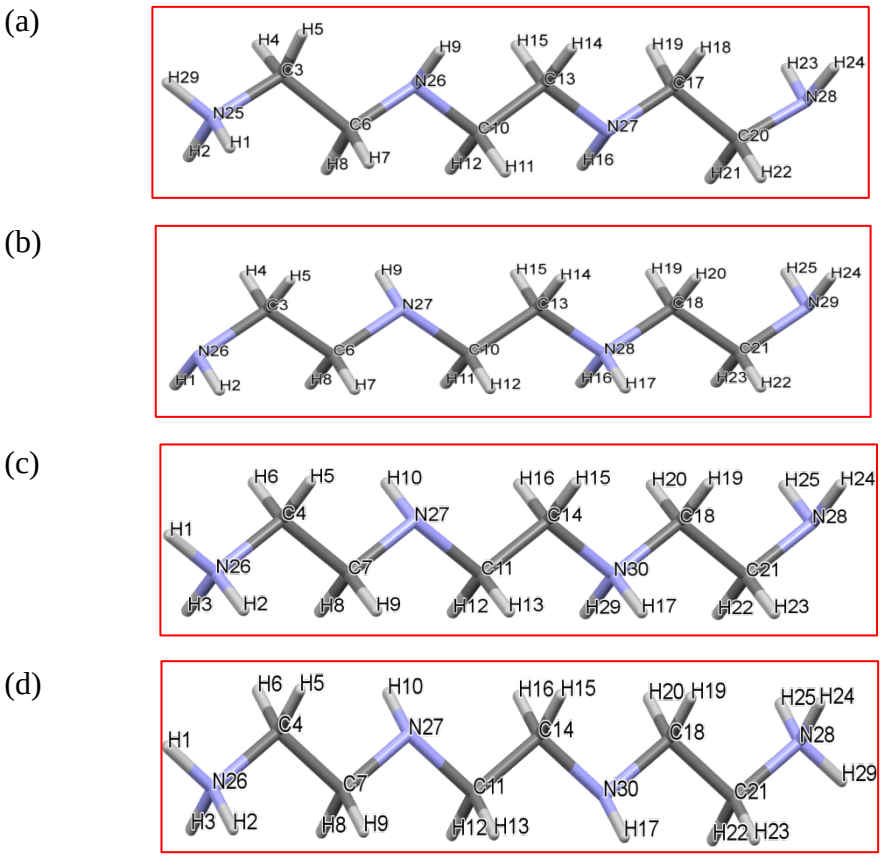


Figure 3.1. Capped-stick representation of linear structures of the mono-protonated: (a) - HL_p and (b) - HL_s, and di-protonated: (c) – H_2L_{ps} and (d) – H_2L_{pp}, tautomers of trien used as inputs for conformational search by MM, also showing atoms' numbering.

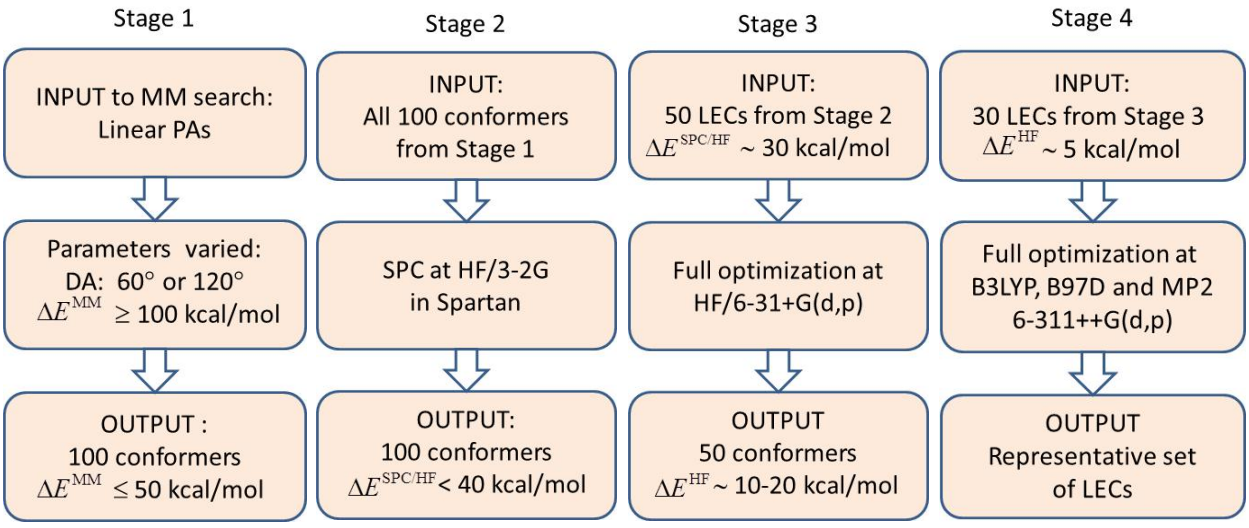
Each tautomer has nine single bond (N–C and C–C) torsional degrees of freedom. For each rotatable bond, a systematic variation in six steps of 60° was allowed in conformational searches.

Due to the enormous number of possible conformers generated from MM-based search (*e.g.*, there are over 10 million theoretically predicted by Spartan possible conformers for each of the **HL** tautomers) one is immediately faced with a challenge, namely (i) how many to retain and (ii) how large a MM-energy window must be used within which retained conformers should fall.

Furthermore, it is important to stress that this study requires not only LECs but also representative structures of medium and higher energy conformers necessary to investigate their structural preferences and topological characteristics. To achieve these goals, a multi-stage approach which makes it possible to eliminate, after each stage, the 'redundant' structures when aims of this work are concerned was implemented- see Scheme 3.1.

Stage 1 involved MM-based search and we tested the influence of the pre-set MM-energy window, ΔE^{MM}, on the energy difference between the highest (HEC) and the lowest energy conformer generated among 100 structures retained, $\Delta E^{MM} = E^{MM}_{HEC} - E^{MM}_{LEC}$. To obtain the representative wide energy spectrum of conformers we were setting the MM-energy window to rather large values, $\Delta E^{MM} \geq 100$ kcal/mol. Surprisingly, the resultant set of MM-generated 100 structures always produced $\Delta E^{MM} < 50$ kcal/mol but, fortunately and conveniently, each time with the same (or very much similar) set of LECs which were most sought after.

In **Stage 2** we performed (in Spartan) a single point calculation (SPC) at the HF/3-2G level of theory on all 100 structures retained. Our aim was to generate more realistic values of molecular energies, hence energy differences between conformers. We have further assumed that the a set of LEC and HEC conformers required for the study of conformational preferences should be located within the energy window of ~30 kcal/mol obtained at the HF/3-2G level of theory and



Scheme 3.1. Four-stage protocol implemented in the search of representative lowest energy conformers.

we decided to take the 50 lowest in energy conformers for further optimizations. In other words, our reasoning was that none of the rejected conformers would change its energy (at higher level of theory) so much that it would fit into the final set of the lowest energy conformers

Next, in **Stage 3**, we fully optimized these 50 structures at the HF/6-31+G(d,p) level of theory; this resulted in the energy windows ΔE^{HF} of about 12, 22, 10 and 11 kcal/mol for HL_p, HL_s, H_2L_{ps} and H_2L_{pp} tautomers, respectively. We selected 30 LECs (their relative energies were within about 5 kcal/mol window) and, during the final **Stage 4** of our CSP, they all were fully energy-optimized at B3LYP, B97D and MP2 (in each case a 6-311++G(d,p) basis set was used).

3.3. Results and discussion

The structural and topological data generated at lower than MP2 level of theory are presented as differences rather than absolute values. As an example, when interatomic distances d(H,N) for the intramolecular NH•••N interactions are analysed, we will report data obtained at, *e.g.*, HF, as $\Delta d(H,N)_{HF} = d(H,N)_{HF} - d(H,N)_{MP2}$. Using this approach made it easier to evaluate the relative performance of a lower level of theory. Moreover, one must remember that performing all the calculations at the MP2 level on a large number of conformers is extremely time-consuming and, importantly, not always necessary to achieve goals of interest. To facilitate data analysis and to monitor the performance of the conformational search implemented here, we have consistently labelled conformers of the primary (**HL_p**) and secondary (**HL_s**) forms of monoprotonated *trien* as C_pn and C_sn, respectively. Those of the di-protonated forms, **H_2L_{ps}** and **H_2L_{pp}**, have been labelled as $C_{ps}n$ and $C_{pp}n$, respectively, where $C_{ps}n$ is used for conformers of the tautomeric form where one primary and one secondary nitrogen atoms are protonated whereas $C_{pp}n$ is used to denote a structure in which the two terminal nitrogen atoms are protonated. For all tautomers *n* stands for the identification number of a conformer as obtained from the MM-based conformational search where $n = 1$ applies to the lowest MM-energy conformer.

3.3.1. Analysis of conformers' relative energies

Relative energies and Boltzmann distribution, as %-fraction of the total population of conformers considered, of ten LECs obtained from **Stage 4** of the CSP at different levels of theory are listed in Table 3.1; for graphical illustration, data set for the **HL_p** conformers is shown in Figure. 3.2. The full data sets of **HL_p**, **HL_s**, **H_2L_{ps}** and **H_2L_{pp}** conformers optimized at **Stage 4** of the conformational protocol are included in Table A2 of Appendix A and relevant bar-graphs are shown in Figure A1 of Appendix A.

Let us first focus on the MP2-generated data from which it is apparent that:

- There are only few (between 3 and 5) conformers of each tautomer that contribute significantly (>5%) to their populations according to the Boltzmann distribution .
- There are always one or two conformers which are significantly lower in energy when compared with the remaining LECs. Importantly, we found that these two conformer are always among the top 12 conformers generated from MM search; among the lowest energy conformers, the largest n = 12 was found for the di-protonated $\mathbf{H_2L_{pp}}$ form.
- It appears that top 25 MM-generated conformers (of the lowest energy) 'guaranty' fully representative set of MP2-optimized structures, each one contributing to the total population 3% and above. If $C_{ps}41$ is excluded (note that the energy of this conformer is essentially the same as that of $C_{ps}10$; this exemplifies the need of optimizing small sets of DFT-generated LECs at the MP2 level) then the $C_{pp}23$ has the largest n = 23 value.

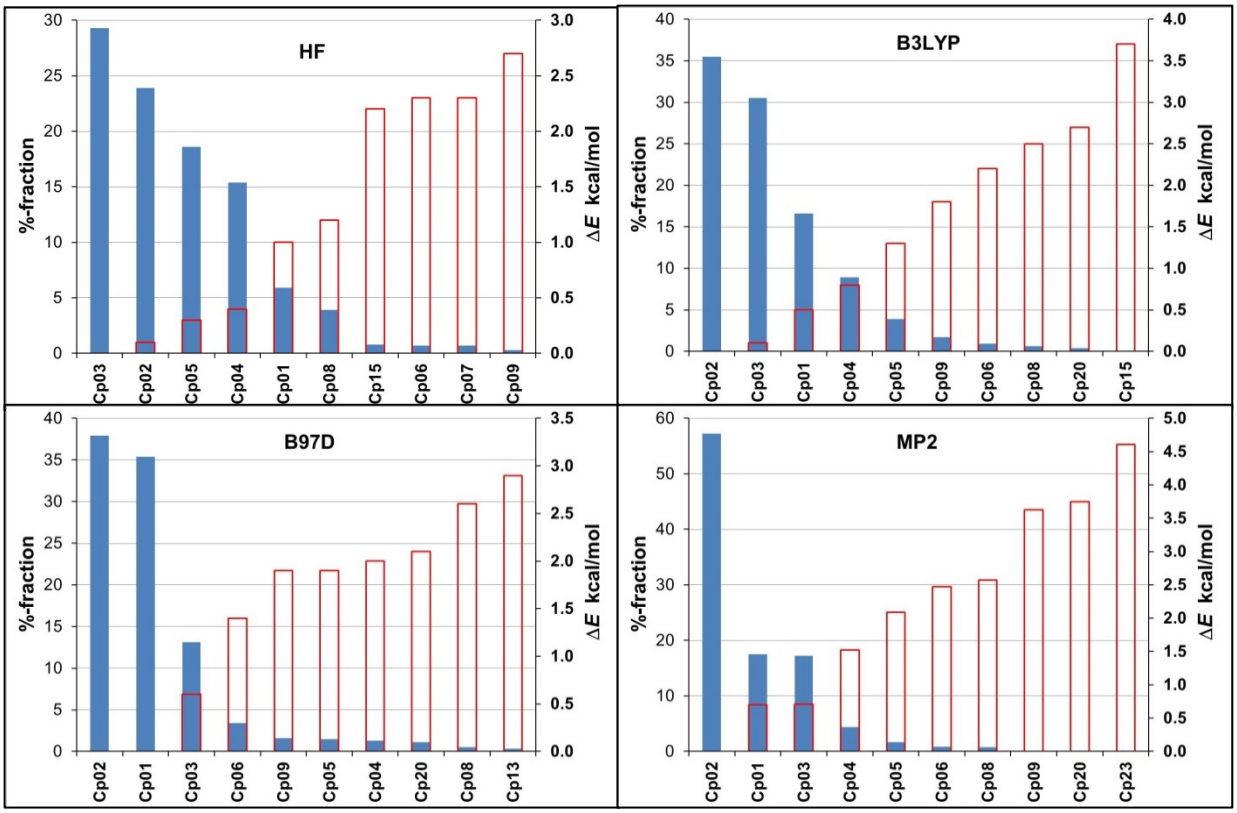


Figure 3.2. Graphical presentation of ten lowest energy HL_p conformers of trien, in terms of %-fraction computed from Boltzmann distribution (solid bars) and relative energies (in kcal/mol) obtained at the indicated level of theory.

- The energy window of 30 conformers (output from **Stage 4** in Scheme 1) decreases from $\mathbf{HL_p}$ to $\mathbf{H_2L_{pp}}$ (from 15 to 5.6 kcal/mol, respectively); the larger number of conformers of

similar energy is observed for the **H_2L_{pp}** form of *trien* and this is due to specific structural features of this tautomer, as discussed in the following section.

Table 3.1. Relative electronic energies ($\Delta E = E_{Conf} - E_{LEC}$ in kcal/mol) of ten lowest energy conformers found at indicated levels of theory and Boltzmann distribution, as a %-fraction, of the total population for: part (a) - HL_s, part (b) - H_2L_{ps} and part (c) - H_2L_{pp}

Part (a)

HF			B3LYP			B97D			MP2		
HL_s	**ΔE**	**%**	**HL_s**	**ΔE**	**%**	**HL_s**	**ΔE**	**%**	**HL_s**	**ΔE**	**%**
C_s08	0.00	23.2	C_s04	0.00	41.2	C_s01	0.00	33.2	C_s04	0.00	40.7
C_s04	0.14	18.5	C_s01	0.28	25.5	C_s04	0.04	31.2	C_s01	0.14	32.0
C_s01	0.26	14.9	C_s03	0.43	19.9	C_s02	0.27	21.1	C_s03	0.64	13.9
C_s03	0.33	13.4	C_s02	0.83	10.1	C_s03	0.56	12.8	C_s02	0.71	12.3
C_s09	0.47	10.4	C_s06	2.36	0.8	C_s05	2.36	0.6	C_s05	2.72	0.4
C_s06	0.69	7.3	C_s22	2.44	0.7	C_s22	2.52	0.5	C_s22	2.90	0.3
C_s02	1.16	3.3	C_s07	2.75	0.4	C_s13	2.76	0.3	C_s06	3.58	0.1
C_s07	1.49	1.9	C_s08	2.81	0.4	C_s14	2.77	0.3	C_s13	3.59	0.1
C_s22	1.71	1.3	C_s05	2.91	0.3	C_s06	5.03	0.0	C_s14	3.60	0.1
C_s32	1.79	1.1	C_s09	3.05	0.2	C_s07	5.18	0.0	C_s07	3.98	0.0

Part (b)

HF			B3LYP			B97D			MP2		
H_2L_{ps}	**ΔE**	**%**	**H_2L_{ps}**	**ΔE**	**%**	**H_2L_{ps}**	**ΔE**	**%**	**H_2L_{ps}**	**ΔE**	**%**
$C_{ps}01$	0.00	58.0	$C_{ps}01$	0.00	82.0	$C_{ps}01$	0.00	55.6	$C_{ps}01$	0.00	71.8
$C_{ps}03$	0.79	15.3	$C_{ps}41$	1.30	9.2	$C_{ps}41$	0.29	34.0	$C_{ps}41$	0.99	13.6
$C_{ps}41$	0.90	12.7	$C_{ps}10$	1.87	3.5	$C_{ps}10$	1.00	10.2	$C_{ps}10$	0.99	13.4
$C_{ps}02$	1.89	2.4	$C_{ps}03$	1.89	3.4	$C_{ps}03$	3.84	0.1	$C_{ps}03$	2.60	0.9
$C_{ps}05$	1.91	2.3	$C_{ps}02$	3.24	0.3	$C_{ps}14$	4.38	0.0	$C_{ps}14$	3.69	0.1
$C_{ps}10$	2.10	1.7	$C_{ps}14$	3.31	0.3	$C_{ps}15$	6.12	0.0	$C_{ps}05$	4.98	0.0
$C_{ps}06$	2.19	1.4	$C_{ps}05$	3.51	0.2	$C_{ps}21$	6.24	0.0	$C_{ps}17$	5.07	0.0
$C_{ps}17$	2.48	0.9	$C_{ps}12$	3.62	0.2	$C_{ps}12$	6.29	0.0	$C_{ps}02$	5.07	0.0
$C_{ps}13$	2.64	0.7	$C_{ps}04$	3.81	0.1	$C_{ps}18$	6.46	0.0	$C_{ps}34$	5.23	0.0
$C_{ps}20$	2.64	0.7	$C_{ps}17$	3.88	0.1	$C_{ps}04$	6.50	0.0	$C_{ps}43$	5.24	0.0

Part (c)

HF			B3LYP			B97D			MP2		
H_2L_{pp}	ΔE	%	H_2L_{ps}	ΔE	%	H_2L_{ps}	ΔE	%	H_2L_{ps}	ΔE	%
$C_{pp}02$	0.00	19.9	$C_{pp}12$	0.00	16.4	$C_{pp}12$	0.00	26.7	$C_{pp}12$	0.00	32.7
$C_{pp}01$	0.09	17.0	$C_{pp}02$	0.20	11.7	$C_{pp}10$	0.37	14.4	$C_{pp}10$	0.73	9.5
$C_{pp}05$	0.46	9.2	$C_{pp}01$	0.31	9.6	$C_{pp}14$	0.78	7.1	$C_{pp}07$	0.86	7.6
$C_{pp}17$	0.51	8.4	$C_{pp}05$	0.36	8.9	$C_{pp}21$	0.87	6.1	$C_{pp}16$	1.09	5.2
$C_{pp}03$	0.52	8.2	$C_{pp}10$	0.50	7.1	$C_{pp}19$	0.87	6.1	$C_{pp}11$	1.09	5.2
$C_{pp}19$	0.54	7.9	$C_{pp}09$	0.59	6.1	$C_{pp}11$	0.89	5.9	$C_{pp}09$	1.13	4.9
$C_{pp}08$	0.76	5.5	$C_{pp}04$	0.60	5.9	$C_{pp}16$	0.89	5.9	$C_{pp}05$	1.26	3.9
$C_{pp}23$	0.91	4.3	$C_{pp}07$	0.66	5.4	$C_{pp}27$	1.05	4.5	$C_{pp}02$	1.38	3.2
$C_{pp}04$	0.97	3.9	$C_{pp}03$	0.75	4.6	$C_{pp}13$	1.19	3.6	$C_{pp}23$	1.40	3.1
$C_{pp}06$	1.08	3.2	$C_{pp}23$	0.81	4.2	$C_{pp}18$	1.19	3.6	$C_{pp}04$	1.41	3.0

The above analysis strongly suggests that the protocol implemented here has indeed provided us with (i) representative sets of the lowest energy conformers (recall that we have fully optimized 30 top MM-generated conformers at all levels of theory) as well as (ii) sufficient spread in conformers' energies (see Table A2 of Appendix A) needed to explore conformational preferences in terms of structural and topological properties. Furthermore, data presented in Table A2 of Appendix A also shows that:

- Without an exemption, the top MP2 conformer of all tautomers is also reproduced as the lowest energy at the B3LYP level.
- Top four MP2 conformers of **HL_p**, **HL_s** and **H_2L_{ps}** are also among top four at B3LYP and B97D whereas top five MP2 conformers of **H_2L_{pp}** (they contribute above ~5% to the total population) are among top 10 conformers at the B3LYP level (they contribute above ~3% to the total population) when conformers with one imaginary frequency, $C_{pp}01$ and $C_{pp}04$, are excluded.
- Grouping of conformers in terms of their relative energies is similar at the DFT and MP2 levels. It means that if just few lowest energy conformers are observed at DFT then these conformers will be even more dominant, as LECs, at MP2. It also appears that if a large number of similar in energy conformers is generated at the DFT level, the same is also obtained at MP2 although each time MP2 differentiates the relative energies better.
- If only the lowest energy conformers are required then it is sufficient to select those from the B3LYP optimization which contribute above ~3% to the total population and subject them to optimization at MP2.

In general, analysis of the relative energies indicates that, as observed for other structurally flexible compounds, such as amino acids[25], the Hartree-Fock method, irrespective of basis set used, is only, if at all, useful for pre-optimization of conformers generated from the MM-search. In addition, energy differences between conformers are the smallest at HF making it difficult to decide on which ones should be retained for further analysis.

3.3.2 Validation of conformational search protocol used

We are of an opinion that the data presented in Table 3.1 and Table A2 in Appendix A in combination with tracing the changes in relative energy of the conformers throughout all the stages of CSP (from MM to MP2) can be used to assess the performance of the CSP implemented in this study. Although, in principle, one cannot guarantee that the global minimum energy conformer has been discovered for any of the protonated forms studied in this work, there are several strong indications that we might have achieved that:

1) We have performed repeatedly conformational search with different set of parameters (such as the energy window, number of conformers to be retained, number of steps used to change the torsional angles, *etc*.) and in each case similar sets of LECs were generated from Stage 1. This allowed us to collate the top conformers (with the lowest MM-energy) for further optimizations.

2) The LEC found from MP2 as well as B3LYP are always among the top LEC generated from MM. It was gratifying to note that the discovered HL_p and HL_s LECs (they constitute over 99 and 96 % of the total population at MP2 and B3LYP, respectively) were within the five and four, respectively, top conformers found from the MM-based conformational search. In the case of both H_2L tautomers, the LECs at MP2 level of theory were among the twelve lowest in energy conformers obtained from the MM conformational search.

3) Medium and higher energy conformers (with $n > 30$) retained after MM search were never found to be among (i) the lowest energy conformers at MP2 and (ii) 10 top lower energy conformers discovered at all levels of theory for all tautomers studied (recall that $C_{ps}41$ and $C_{ps}10$ have virtually the same energy) .

However, to make 'absolutely' sure that none of the remaining 50 conformers (all with $n > 50$ were rejected after **Stage 2** of the CSP where SPC at the lowest level of theory, HF/3-2G, was performed on 100 MM-generated structures) belongs to the set LECs already established, we optimized them all at the HF/6-31+G(d,p) level, as implemented at **Stage 3** of the CSP. For the HL_s and H_2L_{pp} tautomers, all the fifty conformers were higher in energy when compared with the

ones we had selected for optimization in **Stage 4**. However for HL_p and H_2L_{ps}, nine and sixteen conformers fitted within the ΔE_{30}^{HF} ~5 kcal/mol energy window, respectively, used to select conformers for optimization at **Stage 4**. Although they were placed well below the energy range from which our LEC originated, we decided to optimize them at B3LYP/6-311++G(d,p); all of them had relative energy ΔE > 6 kcal/mol (for HL_s) and δE > 4 kcal/mol (for H_2L_{pp}), hence were not suitable for any further analysis.

All the above observations gave us confidence that our search protocol has been able to recover representative (i) samples of conformers needed for analyses of factors influencing conformational preferences and (ii) sets of top lowest energy conformers of each tautomer with a high likelihood that the true LEC was also found. Hence, we would like to conclude this section by proposing a simple and affordable (when time required is concerned) conformational search protocol which should be suitable to discover lowest energy conformers of ALPs. Note, that because of poor performance of HF it is not included in the proposed protocol and our decision will become even more obvious from sections that follow:

- **Stage 1**: Perform a MM conformational search on input structures with all N-C-C-C or N-C-C-N dihedral angles of 180° (all-trans conformers). It is sufficient to vary the dihedral angles in 120° steps. Retain all (but not more than 50) unique conformers within 25 kcal/mol relative MM-energy window, $\Delta E^{MM} = E_{HEC}^{MM} - E_{LEC}^{MM}$.

- **Stage 2**: Select top conformers within ΔE^{MM} = 20 kcal/mol energy window and 'pre-optimize' them at the B3LYP/6-31+G(d,p) level (all top conformers found at MP2 in this work were within ΔE^{MM} ~18 kcal/mol).

- **Stage 3**: Select top conformers from **Stage 2** (within $\Delta E_{pre\text{-}opt}^{B3LYP}$ = 3.5–4 kcal/mole energy window) and optimize them at the B3LYP/6-311++G(d,p) level of theory (note that ΔE^{HF} < 3.5 kcal/mol was sufficient in this study and, importantly, B3LYP provides better differentiation in conformers' energy).

- **Stage 4**: Finally, select conformers within ΔE^{B3LYP} = 3–3.5 kcal/mol relative energy window obtained in **Stage 3** and optimized them at MP2. This should assure 10 top (the lowest energy) conformers at the MP2 level (in our studies, top 5 MP2 conformers of HL and H_2L tautomers came from ΔE^{B3LYP} < 3 kcal/mol).

3.3.3. Structural preferences

Full sets of 30 structures for **HL_p**, **HL_s**, **H_2L_{ps}** and **H_2L_{pp}** are shown in Figures A2–A5 in Appendix A. For convenience, the structure of the LEC and a representative sample of conformers from medium and higher energy spectrum of all tautomers obtained at MP2 are displayed in Figures. 3.3. and 3.4. An examination of all the conformers of both tautomers shows some interesting general structural features and trends (we must stress that similar geometrical features, in relation to *E*, are observed at all levels of theory examined in this work) and we will describe them separately for each tautomer of both protonated forms of *trien*.

*The **HL_p** conformers*:

a) Two H-atoms of the $–NH_3^+$ group in the LECs are involved in the formation of the NH--N short contacts. One H-atom is forming the leading 11-membered molecular ring (11m-MR) and the other forms two additional intramolecular NH--N short contacts; as a result, all N-atoms are involved in the intramolecular interactions and the largest possible number of rings is formed (11m-MR, one 7m-intramolecular ring, 7m-IR, and two 5m-IRs).

b) Among LECs which form the 11m-MR, those with the lowest energy (C_p02, C_p01, C_p03 and C_p04; they constitute over 90 % of the population) form the shortest contacts between terminal $–NH_3^+$ and $–NH_2$ groups with d(H,N) ~1.745±0.05 Å.

c) The somewhat higher energy conformers (they are characterized by either 11m-MR or largest possible intramolecular ring) have the shortest contact which involves either the terminal $–NH_3^+$ group or the secondary N-atom, *e.g.*, C_p05, C_p06, C_p07 and C_p08 where we observe d(H,N)~1.695±0.02 Å.

d) Medium energy **HL_p** conformers are characterized by the presence of an 8m-IR, due to the NH--N short contact between the $–NH_3^+$ terminal group and the farther secondary N-atom, leaving the $–CH_2CH_2NH_2$ fragment free to rotate and bent.

e) The highest energy conformers form the smallest possible 5m-IR, typically between the terminal $–NH_3^+$ and an adjacent –NH– group.

*The **HL_s** conformers*:

a) Both H-atoms of the $–NH_2^+$– group of the LECs are involved in the formation of the NH--N short contacts. The shortest contact with d(H,N)~1.735±0.01 Å (this generates the 8m-IR) is formed with the terminal N-atom. The other NH--N contacts involve two of the remaining unprotonated N-atoms. As a result, additional two 5m-IRs are formed, one with the terminal

–NH_2 group where d(H,N)~2.197±0.01 Å and the other with the secondary N-atom where much longer intramolecular distance, d(H,N)~2.457±0.05 Å, is observed. In general, as one observes for **HL_p**, three rings are observed in the LECs except C_s04 (the lowest energy conformer) where only two intramolecular rings are present.

b) The medium and highest energy **HL_s** conformers have mainly smallest possible 5m-IRs and it appears that their consecutive placement is preferred.

*The **H_2L_{ps}** conformers*:

a) As found for **HL_p**, all lowest energy **H_2L_{ps}** (C_{ps}01, C_{ps}41 and C_{ps}10) conformers (i) form the 11m-MRs due to the NH--N contact between terminal functional groups but the average interatomic distance, d(H,N)~1.975±0.13 Å, is about 0.2 Å longer and (ii) with 11m-MR constitute almost 100% of the population.

b) Unlike found for the **HL_p** conformers, the NH--N contact forming the 11m-MR in the lowest energy C_{ps}01 conformer is not the shortest; the 5m-IR has shorter, by about 0.1 Å, the NH--N contact.

c) All three LECs form an additional intramolecular NH--N contact and this involves the H-atom from the –NH_2^+– group of the protonated secondary N-atom and the adjacent –NH– group; as a result two, 10m- and 5m-IRs are formed.

*The **H_2L_{pp}** conformers*:

a) In general, a total of three (or minimum two) 5m-IRs are found in the LECs; we have not found a conformer with a larger intramolecular ring.

b) Among LECs, a H-atom from each of the terminal –NH_3^+ groups is always involved in an interaction with the N-atom of the adjacent –NH– group (regardless whether two or three 5m-IRs are found); this results in the formation of two terminal 5m-IRs. In case of three 5m-IRs being present, they are formed in consecutive fashion; the additional NH--N interaction involves two adjacent –NH– groups.

c) The medium and high energy conformers of **H_2L_{pp}** (as well as **H_2L_{ps}**) generally consist of structures with two 5m-IRs resulting from various NH--N interactions of different atoms depending on the energy range into which a particular conformer falls. Not surprisingly, their highest in energy conformers have only one NH--N interaction resulting in the formation of the smallest possible 5m-IR, as we observed in the HECs of **HL_p** and **HL_s**.

To conclude this section, we would like to summarize and generalize all the above observations:

1) At least a single intramolecular NH--N short contact, with d(H,N) significantly shorter than the sum of the van der Waals radii, is observed in all conformers and it always involves either $–NH_3^+$ or $–NH_2^+$ group, hence protonated N-atom; this applies to both tautomers of mono and di-protonated forms.

2) When more than one intramolecular NH--N contact is formed then H-atoms of both protonated groups are always involved.

3) In general, the largest number of the intramolecular NH--N short contacts, typically three, is found in the LECs for both tautomers of mono and di-protonated forms.

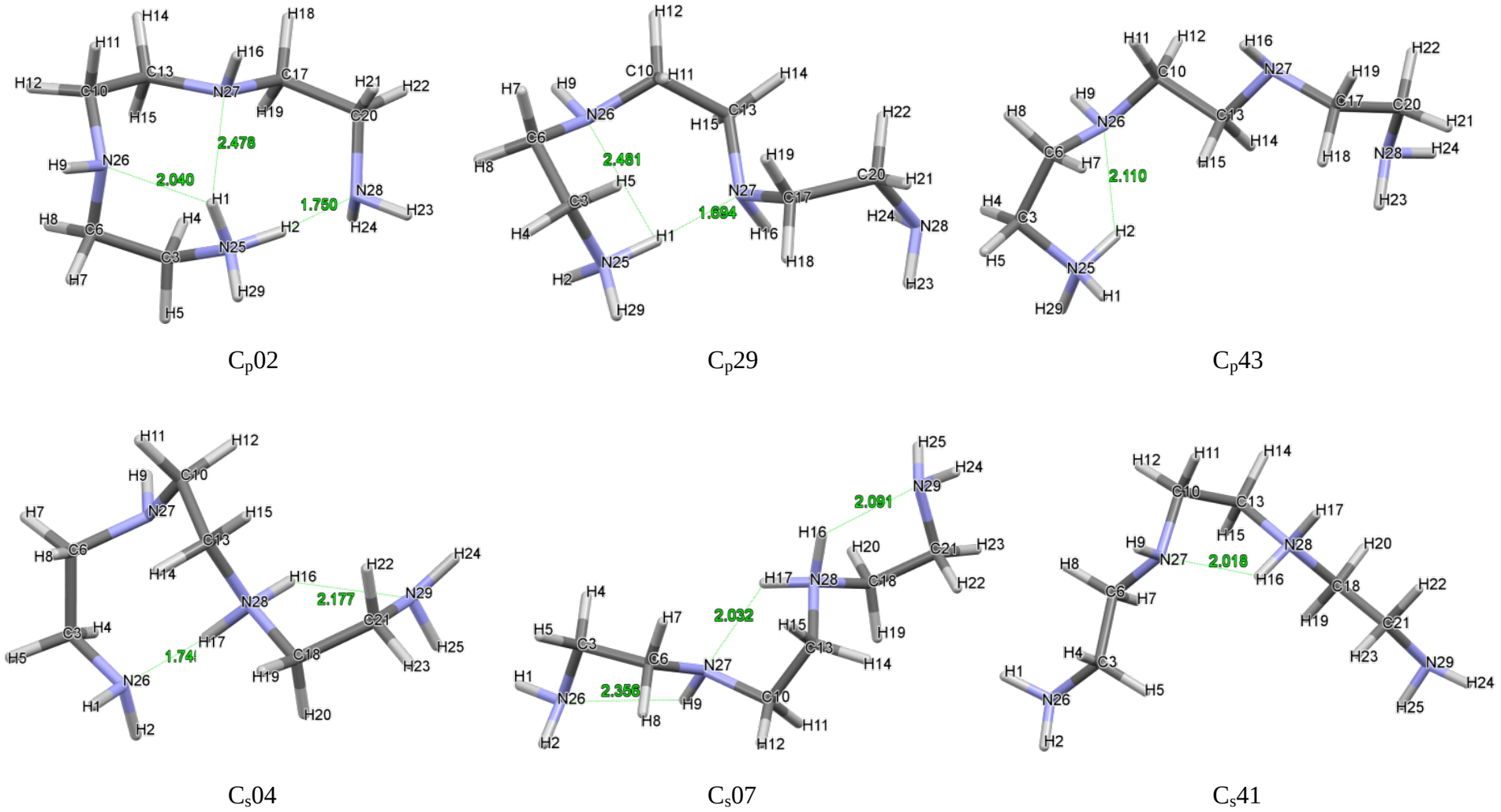
4) The size of an intramolecular ring formed by the NH--N contacts appears to be of significance and most stable conformers form largest possible rings.

5) The latter trend does not apply to conformers with two terminal groups being protonated; only 5m-IRs are formed and, typically, three consecutive such rings are present among the LECs.

6) When a single terminal group is protonated, then the largest possible, 11m-MR is preferentially formed. This feature is observed among ten $\mathbf{HL_p}$ and three $\mathbf{H_2L_{ps}}$ LECs.

7) When a single –NH–group is protonated, then largest possible 8m-IR is preferentially formed; this is observed among top six lowest energy $\mathbf{HL_s}$ conformers.

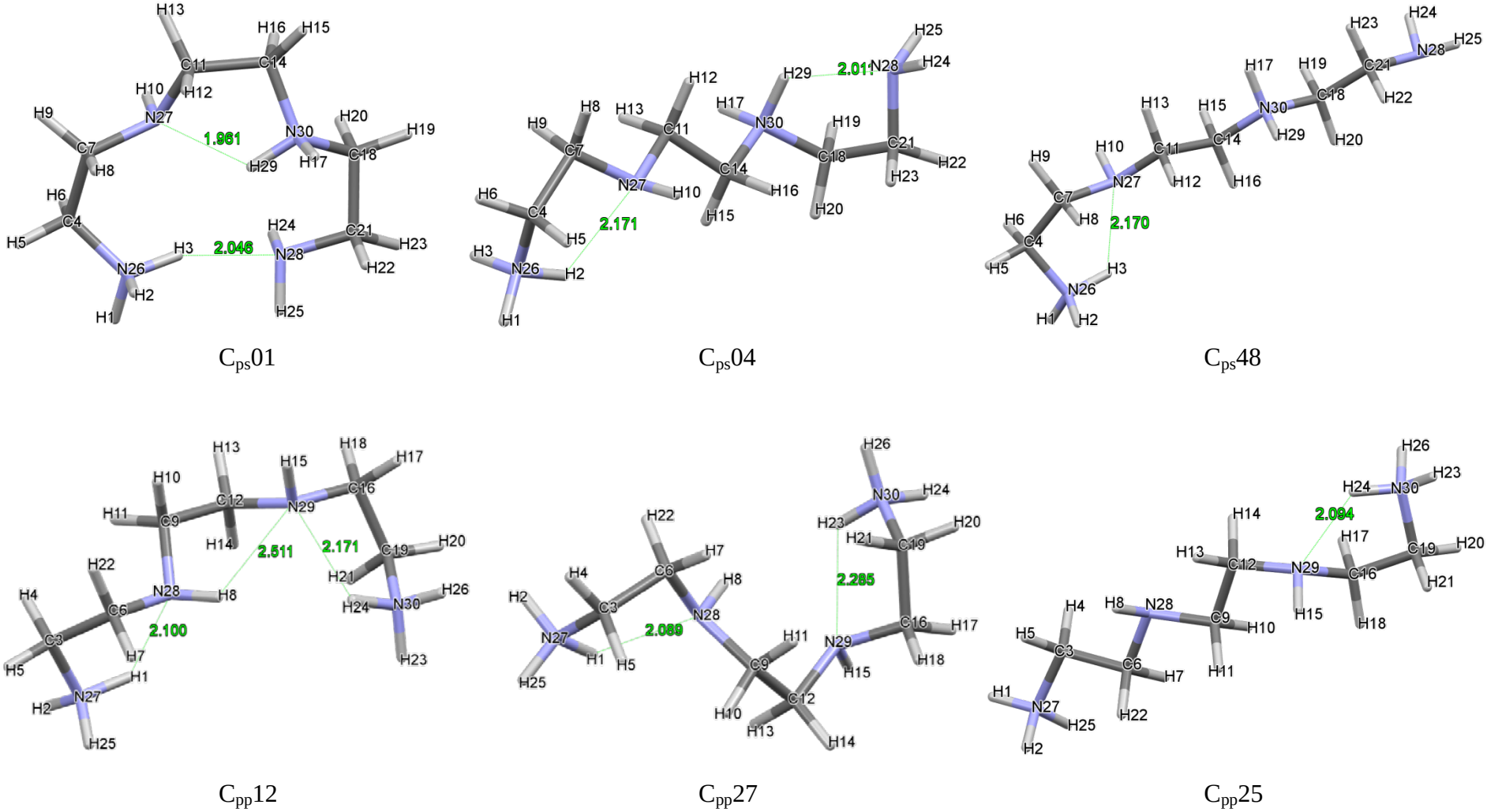
8) It appears that large rings contribute in stabilizing manner to molecular energy more than smaller rings. We found significant differentiation in relative energies of $\mathbf{HL_p}$, $\mathbf{HL_s}$ and $\mathbf{H_2L_{ps}}$ conformers for which small sets of LECs, all with either 11m- or 8m-rings, are observed. Note that all the $\mathbf{H_2L_{pp}}$ conformers form only 5m-IRs and we observe (i) many of them within a relatively small energy window and (ii) typically, most stable conformers form more rings.



C_p02 C_p29 C_p43

C_s04 C_s07 C_s41

Figure 3.3. Representative structures of the lowest (C_p02 and C_s04), medium (C_p29 and C_s07) and higher (C_p43 and C_s41) energy conformers of HL_p and HL_s, respectively, generated at MP2 during the fourth and final stage of the conformational protocol developed in this work, also showing atoms' numbering as well as interatomic distances in Å of the short NH--N contacts



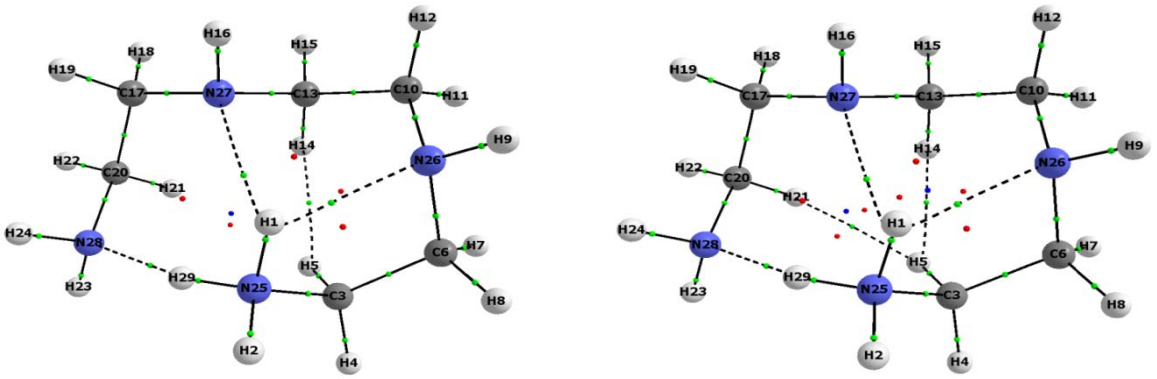
$C_{ps}01$ $C_{ps}04$ $C_{ps}48$

$C_{pp}12$ $C_{pp}27$ $C_{pp}25$

Figure 3.4. Representative structures of the lowest ($C_{ps}01$ and $C_{pp}12$), medium ($C_{ps}04$ and $C_{pp}27$) and higher ($C_{ps}48$ and $C_{ps}25$) energy conformers of H_2L_{ps} and H_2L_{pp}, respectively, generated at MP2 during the fourth and final stage of the conformational protocol developed in this work, also showing atoms' numbering as well as interatomic distances in Å of the short NH--N contacts

3.3.4. Topological preferences of ALPs

Molecular graphs were generated for all 480 final structures obtained from Stage 4 of the CSP, *i.e.*, 30 **HL$_s$**, **HL$_p$**, **H$_2$L$_{sp}$** and **H$_2$L$_{pp}$** conformers optimized at all four levels of theories were analysed. Unfortunately, direct comparison of topological properties was not a straightforward process because often different number of atomic interaction lines (AILs), with associated critical points, CPs, was observed for the same conformer optimized at different LoT. An example is shown in Figure 3.5 where four and five dashed lines representing intramolecular AILs are observed for the C$_p$01 conformer optimized at B3LYP and MP2, respectively; one short CH--HC contact at B3LYP is not linked by an AIL.



(a) (b)

Figure 3.5. Molecular graphs of the C$_p$01 conformer showing different sets of AILs due to different levels of theory these structures were optimized at: (a) – B3LYP and (b) – MP2

Furthermore, it is important to stress that AILs are observed for most (but not all) of the short intramolecular NH--N contacts identified from the geometrical analysis of conformer. Because of that, we based our comparative approach on contacts linked by AILs at MP2. For each molecular graph generated at MP2, a dedicated table was prepared showing atoms involved in a particular interaction, interatomic distance and electron density at a CP, ρ_{CP}, as well as short contacts without an AIL – an example for C$_p$02, the LEC found at MP2, is shown in Table 3.2. which is supplemented (for convenience) by a molecular graph shown in Figure 3.6. A similar approach was adopted for the top 15 conformers of **HL$_p$**, **HL$_s$**, **H$_2$L$_{ps}$** and **H$_2$L$_{pp}$** and relevant data is provided in Tables A3–A6 of Appendix A.

From the analysis of data presented in Tables A3–A6 of Appendix A, it became clear that the appearance of AIL is not governed by the interatomic distance. For instance, (i) one observes an AIL in C$_p$04 with d(N,H) = 2.512 Å, but there is no AIL in C$_p$02, C$_p$05, C$_p$06 or C$_s$06 where shorter d(N,H) = 2.478, 2.482, 2.473 and 2.365 Å are observed, respectively, (ii) there is only one AIL observed even though two NH--N contacts have the same d(N,H) = 2.559 Å - see C$_{pp}$10

Table 3.2. Interatomic distance and electron density at a CP of interactions found on molecular graph, shown in Figure. 4, of Cp02 (the LEC at MP2) together with a short contact which is not linked by AIL

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N28	H2	1.750	0.0507
	N26	H1	2.040	0.0287
CH•••N	N27	H4	2.601	0.0115
CH•••HC	H4	H15	2.124	0.0105
N•••N	N25	N27	2.897	0.0147
Contact without AIL				
NH--N	N27	H1	2.478	–

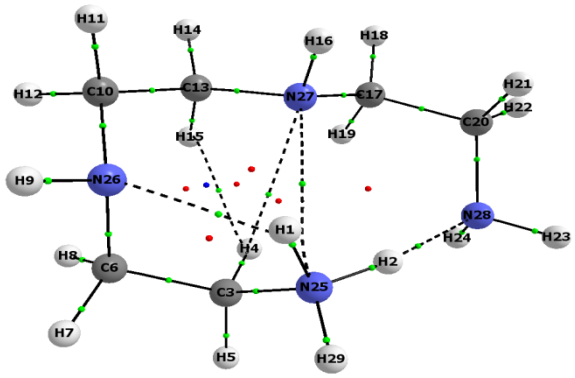


Figure 3.6. A molecular graph of C_p02, the lowest energy conformer obtained at MP2

in Table A6 of Appendix A where N29 and H24 are linked by AIL but N29 and H8 are not, and finally (iii) atoms separated by larger distance might be linked with AIL, *e.g.*, for $C_{pp}12$ shown in Table A6 of Appendix A, an AIL is observed for N29 and H24 with d(N,H) = 2.566 Å but no AIL is observed for N29 and H8 with d(N,H) = 2.511 Å. Moreover, there are numerous additional AILs (with small ellipticity) linking atoms one would not expect; we refer here to CH•••HC intramolecular interactions which, surprisingly, are mainly present in the LECs of the monoprotonated forms but, even more unexpectedly, are not observed at all in the ten highest in energy conformers of **HL** tautomers. Considering $\mathbf{H_2L}$ tautomers, these CH•••HC interactions were completely absent in the conformers of the $\mathbf{H_2L_{ps}}$ tautomer and only were found in four highest in energy conformers of the $\mathbf{H_2L_{pp}}$ tautomer. Furthermore, some conformers of the **HL** tautomers might be considered as topologically 'unstable' (even though no negative frequencies are present, hence energetically they are stable) because unusual atoms are linked with highly bent AILs, *e.g.*, we observe the CH•••N (in C_s03 shown in Table A4 of Appendix A), N•••N (in

C_p02) or CH•••C (in C_p04) interactions which are shown in Table A3 of Appendix A.

It is now well established that the ρ_{CP} *vs.* d(A,B) should follow an exponential relationship when the same kind of interaction in very much similar molecular environment takes place. A good example one can find in the work by Arabieh and co-workers[31] where a large number of intramolecular H-bonds was examined in a numerous conformers of pamidronate (bisphosphonate). Clearly, one might expect that the NH•••N interactions in both tautomers of **HL** should fully meet this requirement and we have decided to test whether topological 'anomalies' (this also includes appearance of AILs representing CH•••HC) have some influence on ρ_{CP} *vs.* d(N,H) relationships involving 'legitimate' interactions, classically regarded as the intra-molecular H-bonds. We noted that the NH•••N interactions, *e.g.*, in MP2-generated $\mathbf{HL_p}$ conformers, could be grouped in terms of interatomic distances. We found fifteen short contacts with d(N,H) = 1.70±0.03 Å for which ρ_{CP}(H•••N) = 0.058±0.05 a.u., eight medium length contacts, d(N,H) = 2.04±0.08 Å with ρ_{CP}(H•••N) = 0.029±0.04 a.u. and five longer contacts, d(N,H) = 2.47±0.09 Å with ρ_{CP}(H•••N) = 0.014±0.02 a.u. and they all follow an excellent ρ_{CP} *vs.* d(N,H) relationship - see Figure 3.7a for $\mathbf{HL_p}$ conformers where, as it should be, an exponential decay in ρ_{CP}(NH•••N) is observed with an increase in the interatomic distance regardless on the presence of additional AILs or whether a structure could be regarded as topologically stable or not. Considering the $\mathbf{HL_s}$ conformers, the NH•••N interactions are characterised by overall longer interatomic distances; they are dominated by eighteen medium d(N,H) of 2.11±0.05 Å with ρ_{CP}(H•••N) = 0.025±0.03 a.u. rather than short range contacts; we found only eight shorter contacts with d(N,H) = 1.74±0.05 Å and ρ_{CP}(H•••N) = 0.054±0.05 a.u.

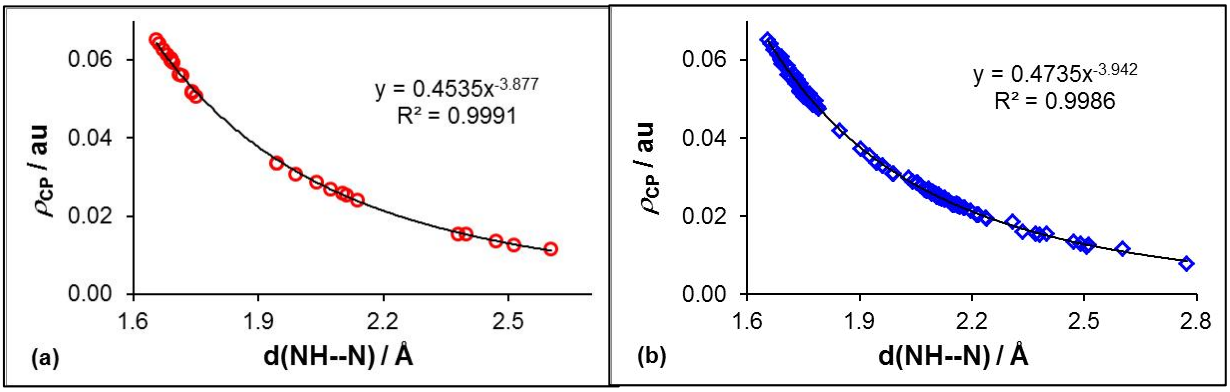


Figure 3.7. Exponential decrease in ρ_{CP} with interatomic distance d(N,H) for all NH•••N interactions in 15 LECs for: part a –HLp at MP2, part b – HLp and HLs at MP2 and B3LYP

We have decided to combine the data related to all the NH•••N interactions in 15 lowest energy

conformers of both, **HL$_p$** and **HL$_s$**, tautomers and generated the ρ_{CP} *vs.* d(N,H) relationship; it was pleasing to note that the trend and quality of the resultant relationship was as seen in Figure. 3.7a; as an example, a relationship obtained at B3LYP is shown in Figure. A6 in Appendix A. Finally, we combined all the relevant data obtained at MP2 and best performing B3LYP and, as seen in Figure 3.7b, this had no influence on the quality of the relationship. This observation can also be used in support of our recommendation related to the use of B3LYP. The trend seen in Figure 3.7b shows that the overall density distribution between atoms does not depend to a large extent on the level of theory used even though (i) the electronic energies of the same conformers and (ii) interatomic distances between the same atoms involved in the NH--N contacts differ significantly at the MP2 and B3LYP level of theory. One might rationalize this observation in terms of the same underlying physical fundamentals, namely the Ehrenfest and Feynman forces. The interplay between potential and kinetic energy densities appears to be such that density distribution at critical points throughout the molecules varies exponentially with the interatomic distance; note that the combined ρ_{CP}(H•••N) data comes from **HL$_s$** and **HL$_p$** tautomers which are characterized by the same kind and a number of atoms, hence the same molecular environment is present in both tautomers. Similar quality relationships were found for the NH•••N interactions also in **H$_2$L** tautomers which are characterized by dominance of medium d(N,H) values; forty d(N,H) = 2.11±0.04 Å and thirty d(N,H) = 2.12±0.08 Å in the **H$_2$L$_{pp}$** and **H$_2$L$_{ps}$** tautomers, respectively.

According to QTAIM, the 'glue' that binds atoms together is the shared electron density at critical point, ρ_{CP}, and this can be used as a measure of relative strength of an interaction. Focusing on the leading and by far strongest NH•••N intramolecular interactions, it appears that, on average, they are weaker in **HL$_s$** than in **HL$_p$** when measured by the individual values of ρ_{CP}. Also, the stronger the interaction (the larger ρ_{CP}) the more significant stabilizing energy contribution to a molecular system is expected. Because there are several NH•••N interactions in the LECs and assuming that these stabilizations are additive, it is reasonable to evaluate their combined contribution[32], denoted here as $\rho_{CP}(H\bullet\bullet\bullet N)_{total}$, which represents summed densities at CPs of these interactions within a conformer. From preliminary inspection of topological data we found, just as an example, that the lowest energy conformers of **HL$_p$** (C_p02) and **HL$_s$** (C_s04) have only two NH•••N interactions and the $\rho_{CP}(H\bullet\bullet\bullet N)_{total}$ value of 0.079 a.u. in C_p02 is larger than that in C_s04 (0.074 a.u.); however, the electronic (E) and Gibbs free (G) energies of C_s04 are more negative (C_s04 is more stable than C_p02). In contrast, the $\rho_{CP}(H\bullet\bullet\bullet N)_{total}$ value obtained for $C_{ps}01$ (0.073 a.u.) is larger than that found in $C_{pp}12$ (0.049 a.u.) and this corresponds to their

relative stability; the values of *E* and *G* of $C_{ps}01$ are ~4 and 2 kcal/mol, respectively, lower (more negative) than that of $C_{pp}12$.

To gain better assessment and understanding, we initially plotted the sum of ρ_{CP} values for all NH•••N interactions linked with AIL in each conformer against its relative energy for the fifteen lowest in energy conformers of each tautomer optimised at MP2. The best such relationship was obtained for **HL$_s$** which is shown in Figure 3.8a – points marked with asterisks follow a reasonable trend with R^2 = 0.936 (dashed line); an example of rather poor relevant trend (R^2 = 0.496) is shown for **HL$_p$** as asterisks in Figure 3.8b.

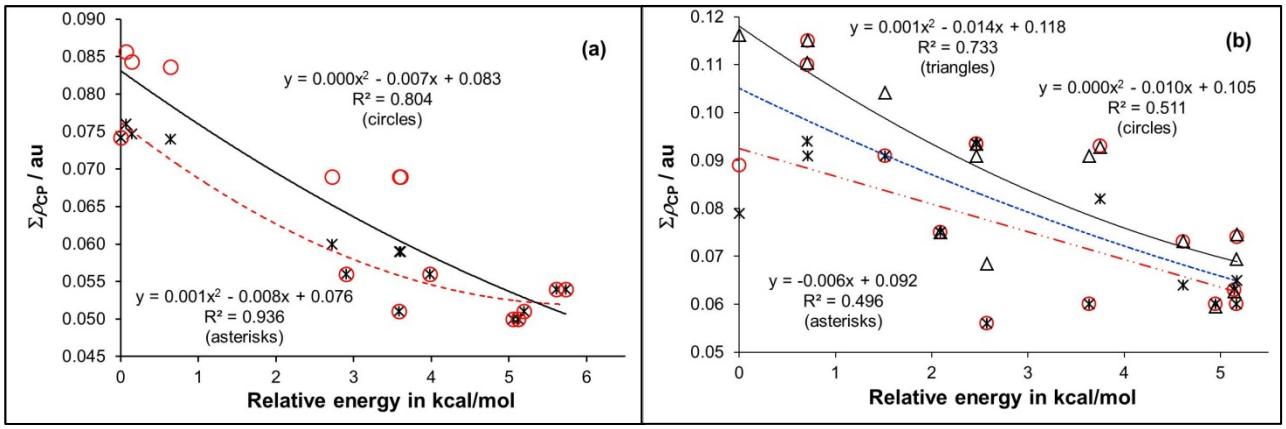


Figure 3.8. Relationship between $\Sigma\rho_{CP}$ and relative energy of top 15 conformers of (a) HLs and (b) HLp. Asterisks represent the NH•••N interactions; circles stand for sum of the NH•••N and the CH•••HC interaction, and triangles (in b) are used for the sum of all interactions linked by AILs in a conformer

Thereafter we plotted the sum of ρ_{CP} values for all AIL-linked NH•••N and CH•••HC interactions found for each conformer against their relative energy – see circles in Figure 3.8. This resulted in significantly worse trend (R^2 = 0.804, solid line) for **HL$_s$** but somewhat better trend (R^2 = 0.511, dotted line) in case of **HL$_p$** in Figure 3.8b.

Finally, we have decided to sum up densities at all CPs within a conformer, regardless whether AIL could be regarded as representing acceptable for a chemist an intramolecular interaction – an example of much improved relationship, with R^2 = 0.733 , is shown as triangles in Figure 3.8b.

At first glance the overall picture obtained from these relationships appears to be inconsistent and highly confusing. However, one must recall that (i) the appearance of an AIL (or a bond path) was interpreted as a privileged exchange-correlation channel, hence it does not have to be present between all short geometrical contacts and (ii) it is well-known that an increase in electron density in the bonding interatomic region always results in a stabilizing energy

contribution made by an intramolecular interaction.[33] An analysis of Tables A3-A6 in Appendix A provides further explanation. Namely, there are numerous conformers with either NH--N and/or CH--HC contacts (some shorter than those linked by AIL) which are not linked with AILs, hence it was impossible to include their contributions in the relationships drawn in Figure 3.8. Moreover, the best trend obtained for $\mathbf{HL_p}$ includes, besides NH•••N and CH•••HC, also CH•••N and N•••N interactions when linked by AILs and this significantly improved the regression coefficient, from $R^2 = 0.496$ to 0.733. All the above leads to the conclusion that it is most likely impossible to generate high quality $\Sigma\rho_{CP}$ *vs*. ΔE relationships in the case of large(r) molecules with densely packed atoms. Perhaps, one would need input from all interatomic areas characterized by non-uniform density distribution, a characteristic feature of QTAIM- as well as NCI-defined critical points, to improve such relationship.

3.3.5. Insight from an NCI analysis

From the above analyses it follows that the stability of *trien* conformers depends mainly on the strength and number of the NH•••N interactions present. It is also clear that the (de)stabilizing role of the CH•••HC interactions in the LECs of *trien* (or more generally, in ALPs) cannot be neglected. Furthermore, since the ρ_{CP} value of an interaction does not necessarily give sufficient information as to the stabilizing or otherwise nature of an interaction[34-37], we therefore decided to gain further insight from the recently developed non covalent interaction index (NCI) technique.[38] This is also because the presence of other intramolecular interactions, which are invisible to the QTAIM technique but must contribute to (de)stability of a conformer, can be uncovered by the NCI method.

NCI is not widely used yet, hence to facilitate interpretation of NCI-generated data, it is in order to give a brief outline of NCI descriptors used to describe a nature and kind of interaction. This technique is making use of the real space visualization and interpretation of the electron density (ρ) and its reduced gradient $s(\rho)$ to locate regions of space where the electronic density distribution deviates from homogeneity due to the formation of a non-covalent inter- or intramolecular interaction.[39] Using the sign of the second eigenvalue (λ_2) of the Hessian matrix, NCI analysis is used to classify interactions as (de)stabilizing according to the topology (decrease/increase) of electron density distribution. Conveniently, results obtained from the NCI analysis are displayed using either 2D or 3D plots. The 2D plot shows variation in the reduced density gradient as a function of the electron density oriented by the sign of its λ_2. Within this plot, the presence of an intramolecular interaction is shown by the appearance of a spike/trough

in regions of low electron density. To visualize the interactions, 3D isosurfaces are generated which reveal the spatial location of inter or intramolecular interactions and their nature within a molecular system.[38,39] In order to rank interactions according to their strength and nature, a colouring scheme is utilized. According to this scheme: (i) stabilizing interactions – also referred to as type I NCI interactions characterized by density increase in the interatomic region – are coloured in blue, (ii) destabilizing – NCI type II – interactions are coloured in red, indicative of density depletion, and (iii) van der Waals – NCI type III – delocalized weak interactions are shown in green. The colour intensity corresponds to the strength of an interaction, *i.e.*, greater intensity refers to stronger interaction and *vice-versa*. The shape of the isosurfaces of the three main NCI-types interactions can also vary, *e.g.*, (i) small, flat, pill-shaped isosurface for stabilizing interaction has been attributed to its bi-centric nature, (ii) red cigar-shaped isosurface (elongated along directions of decreasing density within an intramolecular ring) was linked with destabilizing interaction, and (iii) sheet-like extended surfaces were interpreted as weak van der Waals interactions. In the case of flexible molecules, such as amino acids and aliphatic polyamines, the formation of non-covalent interactions often leads to steric crowding and intramolecular rings. A closure of a ring, an interatomic region between interacting atoms, typically appears as a bicoloured almond-shaped NCI isosurface in a 3D NCI plot with one end being coloured in green-to-blue (this represents a stabilizing contribution) and the other end is usually red which indicates multi-centric repulsion amongst atoms involved in ring formation. All the above are generally used and accepted NCI descriptor and we will attempt to utilize them in the interpretation of intramolecular interactions.

Selected examples of the NCI 3D plots for each tautomer are shown in Figure. 3.9 and relevant plots for top five lowest energy conformers of each tautomer are shown as Figures. A7-A10 in Appendix A. Let us first focus on the strongest, NH•••N, interactions:

a) All large rings, 11m-MR and 9m-IR, are characterized by a dark blue pill-shaped NCI isosurface which corresponds very well to highly stabilizing contribution recovered by QTAIM in the form of AILs with relatively large ρ_{CP} values. Furthermore, these dark blue discs are not surrounded by red rings; this indicates that there is no strain in this interatomic region and this agrees well with general notion related to this type of interaction.

b) Without an exemption, the formation of all 5m-IRs with AILs is recovered by the presence of an almond-shaped bicoloured NCI isosurface. The blue region, as it should, coincides with the presence of an interaction critical point and the red region is due to the steric

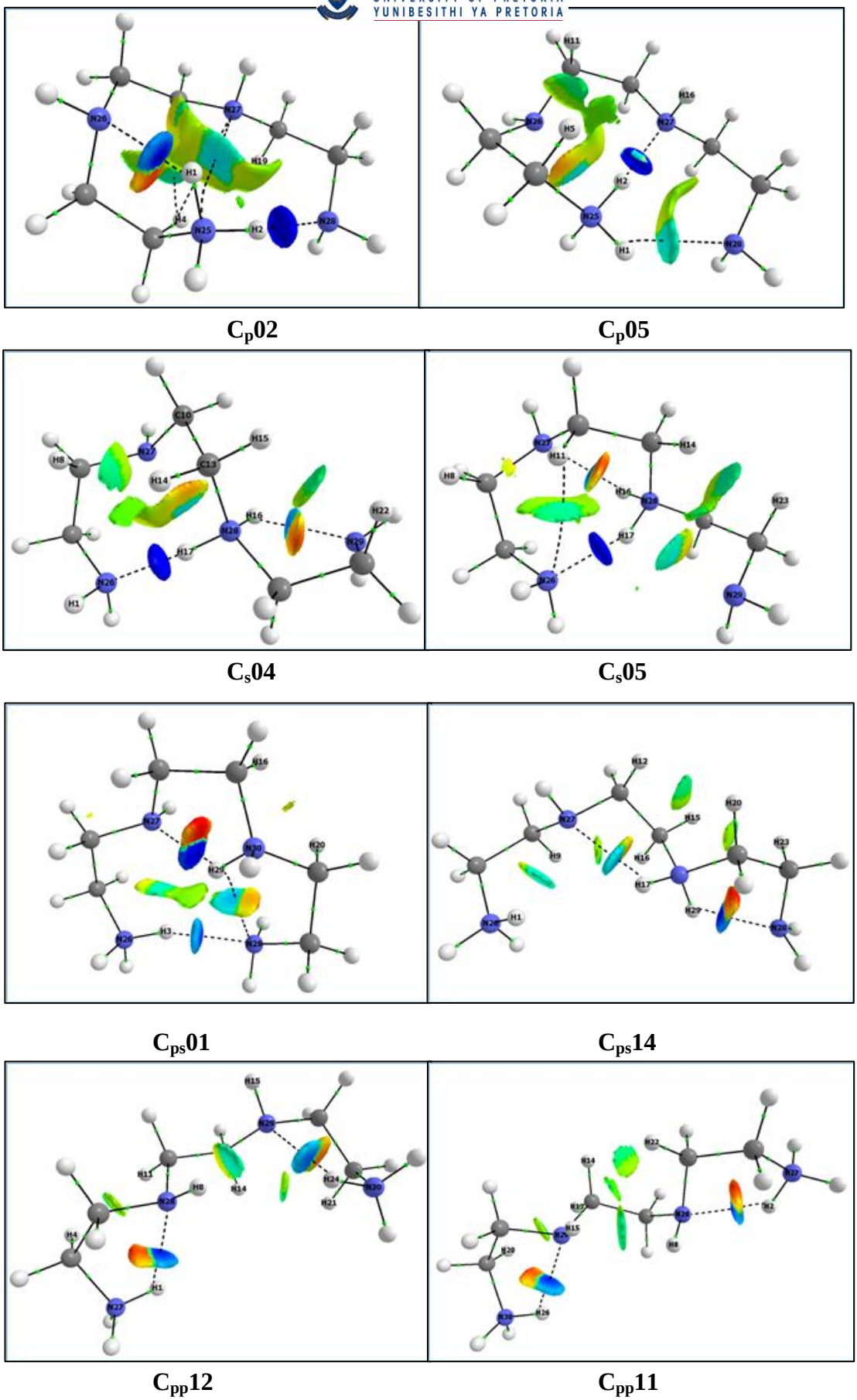


Figure 3.9. NCI isosurfaces (RDG isovalue = 0.5 a.u.) for selected conformers of (a) HL_p, (b) HL_s, (c) H_2L_{ps} and (d) H_2L_{pp}. Isosurfaces are coloured from blue to red using a $-0.03 \leq \rho(r) \times \text{sign}(\lambda_2) \leq +0.03$ range

crowding (caused by an intramolecular ring formation) and it points at the area where a ring critical point (point of lowest density within a ring) is identified from a QTAIM analysis.

c) The NCI analysis also revealed the presence of additional NH•••N interactions which were not recovered from the QTAIM-based analysis – a typical example is seen for C_s05 in Figure 3.9 where a bluish colour of the NCI plot is observed between N29 and H17 (similar is found *e.g.*, for the contact between N27 and H29 in $C_{ps}03$ in Figure. A9 in Appendix A). These AIL-absent NH•••N interactions which form intramolecular 5-membered rings also resulted in almond-shaped bicoloured NCI isosurfaces (as found for AIL-linked 5-membered rings) but their isosurfaces have less intense colours which is indicative of being weaker interaction.

We change our focus on few and rather unexpected interactions now:

a) There is a N27--N25 contact in the MP2 lowest energy C_p02 conformer which is linked by an AIL, hence it might be represented as N27•••N25. Classically, such an interaction would be seen as repulsive and a cause of intramolecular strain, hence destabilizing a molecule. It is then most interesting to note that indeed, as one would expect due to the presence of AIL, a blue isosurface is observed between these two N-atoms but, most surprisingly, it is not surrounded by red-coloured area. The blue in the centre and red on the outskirts discs are synonymous with strain caused by crowded atomic environment and have been reported for, *e.g.*, water dimer where d(O,H) of intermolecular interaction was smaller than that at the equilibrium[40] or for water dimer with forced-to-be O-atoms in close proximity.[34]

b) Another example of unusual interaction is seen in Cs05 (Figure 3.9.) where the AIL-linked CH11•••N26 interaction is observed; also in this case no trace of intramolecular strain is recovered by the NCI-defined isosurface and, in general, we observe very much the same features as described in details for the N27•••N25 interaction in C_p02.

Considering the CH•••HC interactions identified by NCI technique, we note that:

(a) None of them appears to be strained; notice that all relevant isosurfaces, within a bonding region of the interactions, are not coloured in red.

(b) Some of interatomic H--H regions display isosurfaces coloured in blue (which might be interpreted as being of significantly stabilizing nature, *e.g.*, CH14•••H20C in $C_{ps}14$ in Figure 3.9. or CH14•••H20C in $C_{ps}14$ in Figure. A9 in Appendix A.

(c) Interestingly, these blue (or light-blue) interatomic regions are parts of almond-shaped bicoloured NCI isosurfaces and, importantly, the other colour is not red (or reddish) which strongly suggests that the resultant 5-memebred rings are not strained either. What one observes, instead, is the change from blue to a mixture of yellow and green colours, a transition from stabilizing (increased density) to van der Waals – NCI type III - delocalized weak interaction.

(d) In many instances, the number of the CH•••HC interactions identified by NCI technique is larger when compared with QTAIM, *e.g.*, two QTAIM- and three NCI-identified interactions in C_p01 or none QTAIM- and three NCI-identified interactions in $C_{ps}14$. However, the NCI isosurfaces are virtually indistinguishable for both, AIL-linked and AIL-absent, interactions and also compare well with weaker NH•••N interactions without AILs.

Finally, we would like to focus on intramolecular regions in conformers where large rings are formed. There is a striking difference between intramolecular regions of the 11m-MRs and 8m-IRs. One observes large multi-coloured NCI isosurfaces which represent multi-centric interactions and are (i) predominantly blue-to-yellow-to-green in the case of 11m-MRs (typically in C_p conformers) whereas (ii) a significant degree of red colouring is present in the case of the 8m-IRs as seen, *e.g.*, in C_s conformers. When this observation is combined with isosurfaces obtained for 5m-IRs, it appears that the intramolecular strain increases with a decrease in the size of a ring and this correlates well with earlier observation where we found that the most stable conformers form 11m-MR and largest possible intramolecular rings. The fact that NCI managed to recover some degree of intramolecular strain within 8m- and 5m-IRs, which are formed by some of the NH--N contacts, provides additional credibility to our interpretation of the CH--HC contacts as strain free.

We have not performed the NCI analysis on all conformers (hundreds of them) because (i) this is simply not feasible and (ii) this methodology is not perfectly suited for quantitative analysis. Instead, we focused on the lower energy conformers where numerous CH--HC and some other unusual contacts were found and, understandably, uncovering their nature was of paramount importance. The NCI results showed that indeed (i) there are many additional and not negligible intramolecular interactions which, if quantified, might better explain the relative stabilities (energies) of LECs found here, and (ii) the presence of the CH--HC contacts in the LECs has not resulted in the increase of these conformers' energies and most likely, at least in some cases, it might have resulted in additional stabilizing contributions. One must stress, however, that this does not mean that some of CH--HC are indeed destabilizing a molecule and

clearly one would have to perform an in depth examination of each case to fully uncover their nature, but this is not the theme of this work. Moreover, from NCI isosurfaces one can gain an insight on localized to two-atom fragment nature of an interaction as well as the presence of intramolecular strain can be deduced, but this tells us nothing about an energy contribution made by such interaction to the entire molecular system.

There is another important observation to be made. As shown in Figure 3.8(b), an improved trend was obtained when all AIL-linked interactions were accounted for and this correlates well with the above interpretations of the 3D NCI plots. It is possible, as mentioned above, that inclusion of all remaining interactions (which are not linked by AIL) might improve such trend even further. We are of an opinion, however, that this will not result in an excellent relationship (ΔE *vs.* ρ_{CP}) which could be used as a reliable predictive tool because (i) the density at critical points is not a direct measure of the strength of an interaction; similar ρ_{CP} for different kind of interactions might represent different energy contributions, (ii) very much the same density can be observed for attractive and repulsive interactions, and (iii) one still has to account for intramolecular strain but this is not possible at present.

On the positive note, although the quantitative interpretation of relative stability of conformers, not only ALPs, appears to be just a dream, one can gain an invaluable insight on qualitative interpretation using geometric and topological descriptors and features which, as demonstrated in this work, can identify definite lower and definite higher in energy conformers. It is reasonable to assume that findings of this work should be useful also in the analysis of conformers in the gas phase.

3.3.6. Theoretical prediction of protonation sequence

As we have stressed in the introduction, our main focus is to predict theoretically the first protonation site of polyamines because of the 'contradicting' interpretations of experimental results reported to date.

It appears, however, that most recent papers, particularly those involving cluster expansion analysis (CEA) of NMR data, (i) are consistent in that a mixture of primary and secondary tautomers is always proposed but (ii) report significantly different %-fractions for each tautomer, $\mathbf{HL_p}$ and $\mathbf{HL_s}$. For instance, considering well-studied tetramine spermine, 30% ($\mathbf{HL_p}$) and 70% ($\mathbf{HL_s}$) was reported by Borkovec at al.[8] whereas essentially equal molar ratio for both tautomers was found by Albelda et al.[41], both results came from CEA. This is not entirely surprising when one recalls that the cluster expansion analysis (i) involves a lot of unknowns to be fitted

into experimental data (*e.g.*, there are 16 parameters in the case of tetramine spermine[42]) and (ii) the protonation microconstant of each protonation site must be determined from the NMR studies (often experimental NMR protonation constants differ significantly from those determined by most accurate technique, glass electrode potentiometry[43]) and, hence this unavoidably carries some degree of mathematical/statistical and experimental uncertainty.

Furthermore, theoretical studies[12], which were based on linear triamines, predicted a tautomer with secondary N-atom to be the only **HL** form present and, interestingly, we found the same when linear conformers of *trien* were used. It was then of utmost interest and importance to find out whether this prediction will still hold when lowest energy conformers of $\mathbf{HL_p}$ and $\mathbf{HL_s}$, those with %-fraction above 5, were used. To do this, we have selected lowest energy conformers at MP2 with %-fraction > 5% and used their ZPVE-corrected electronic energies, *E*, to predict their relative contributions from Boltzmann distribution based on their relative energies. As can be seen in Table 3.3, it was pleasing to note that a mixture of about 40 to 60% of $\mathbf{HL_p}$ and $\mathbf{HL_s}$, respectively, was obtained in an implicit water environment. This correlates well with recent experimental/theoretical models and clearly indicates an importance of using LEC of each tautomer. Also, one must realize that the obtained 40:60 ratio must not be taken literally as it must depend on a solvation model as well as level of theory used (we found *e.g.*, 1% and 54% of $\mathbf{HL_p}$ from modelling at HF and B97D level of theory, respectively) and, unfortunately, there is no easy way to *rigorously* verify any theoretical model experimentally. In support of this, Weisell et al.[44] found, from QM-modelling of pentamine 1,12-diamino-3,6,9-triazadodecane (SpmTrien) involving explicit water molecules, that 91% ($\mathbf{HL_p}$) and 9% ($\mathbf{HL_s}$) is formed with short chain being protonated preferentially (66% short and 25% long chain for the $\mathbf{HL_p}$ form) which clearly contradicts all earlier as well recent studies whereas from the CEA of

Table 3.3. Boltzmann distribution, as %-fractions, of lowest energy conformers of HL and H_2L at MP2.[a] (conformers with %-fraction > 5 were used).

HL	δ*E*	%	H_2L	δ*E*	%
C_s1	0.00	26.13	$C_{ps}1$	0.00	72.6
C_p2	0.08	22.71	$C_{ps}9$	0.99	13.8
C_s3	0.14	20.59	$C_{ps}10$	0.99	13.6
C_s2	0.64	8.93	$C_{pp}02$	4.27	0.1
C_s4	0.71	7.89	$C_{pp}08$	5.01	0.0
C_p1	0.79	6.91	$C_{pp}01$	5.14	0.0
C_p3	0.79	6.85	$C_{pp}13$	5.36	0.0
	Total HL_p	36.5	Total H_2L_{ps}		99.9
	Total HL_s	63.5	Total H_2L_{pp}		0.1

NMR data they obtained 94% (**HL_p**) and 6% (**HL_s**) with long chain, as expected, being protonated preferentially, ~84%. Moreover, note that their findings, in general, are not in agreement with other studies as they predicted a very small fraction of **HL_s** whereas this tautomer was found to be significant in many polyamines and even dominating, *e.g.*, in spermine.[5,8,41]

On the other hand, one can argue that experimentally or theoretically predicted variation (within reasonable limits) in the %-fraction of these two tautomers, **HL_p** and **HL_s**, should not

have any significance in real, *e.g.*, biological environment; from that perspective most important finding is that indeed these two tautomeric forms can easily change from one form to another implying that each is available to suite a specific biological process best.

Finally, few comments on the **H_2L** form. It has been stressed[42] that two protons must be separated by the minimum of four $–CH_2-$ groups with one proton being located on the terminal nitrogen. From our conformational search it follows that the lowest energy conformers of **H_2L** are characterized by a large number of structures with two terminal 5m-IRs involving protonated either primary or secondary N-atom (the difference in their energy being negligible). Because the lowest energy conformers of **H_2L_{ps}** are characterized by the presence of the largest possible 11m-MR (a feature found also for most stable **HL** conformers), they are also predicted to dominate a mixture of the two, **H_2L_{ps}** and **H_2L_{pp}**, tautomers. Although this does not correlate very well with the cluster expansion analysis of NMR data[8] (it predicts **H_2L_{pp}** to constitute 76% of diprotonated forms of *trien*) one must note that (i) the presence of **H_2L_{pp}** was predicted here whereas this is not the case at all when linear structures are used, (ii) formation of **H_2L_{ps}** does meet the requirement of minimum separation by the $–CH_2-$ groups, and (iii) it is well known that a proton exchange is among fastest reactions known and it is quite possible that conformers with terminal 5m-IRs can easily exchange a proton between the primary and secondary N-atom. Furthermore, in the case of spermine, **H_2L_{ps}** was found to be either a dominant tautomer with 67 %-fraction[8] or making only 40% contribution to the population[41]; interestingly, in both cases data came from the same, CEA protocol and this exemplifies again significant degree of uncertainty in NMR-generated experimental data and their interpretation by CEA.

It has been suggested from the analysis of ^{13}C-NMR studies[42] that terminal N-atoms are protonated first because their better disposition for solvation with the aqueous solvent. To test this, explicit water molecules in computational modelling must be used. Even though this must be seen as extremely time-demanding modelling, results obtained in this work thus far indicate that it is a feasible project because (i) we can make use of an effective conformational search

protocol developed and tested here as well as (ii) data generated at the B3LYP level were found here to be sufficient for the purpose of the study (the same level of theory was also used recently in modelling of SpmTrien.[44] Hence result obtained with explicit solvation is discussed briefly in chapter 5 of this work where four explicit solvent molecules were included in the computational modelling of trien.

3.4. Relative performance of levels of theory

3.4.1. Comparative structural analysis

Two main types of intramolecular short contacts, NH--N and CH--HC, were identified in the various conformers. Classically, the CH--HC contact is interpreted as a steric clash destabilizing a molecule but, surprisingly, it was present in many low energy conformers. Although the van der Waals atomic radii values given by Bondi (1.20Å and 1.55 Å for H- and N-atoms, respectively) are the most widely accepted and cited in chemical literature, subsequent investigations[45-47], based on crystallography and electron density topology, have argued that they were overestimated for certain atoms, among them hydrogen. Consequently it is possible that some of the CH--HC contacts identified here might be due to this overestimation.

We have shown here that the general structural features of each group of conformers is determined by their intramolecular contacts, hence we compared the distances of all NH--N and CH--HC short contacts in the MP2 structures to those of relevant structures obtained at HF, B3LYP and B97D. The absolute value of each deviation from the distance at MP2, $|\Delta d(A,B)|$, was then used to calculate average distance deviation (and its standard deviation) for each conformer. Full data sets for the two types of short contacts identified in LECs of **HL** and **H_2L** tautomers are presented in Tables A7-A14 of Appendix A and, as an example, see Table 3.4.

The comparative geometrical analysis of all NH--N contacts in conformers of the **HL** and **H_2L** tautomers leads to the following observations and trends:

- Relative to MP2, the best estimates of d(N,H) were obtained at B3LYP; we found the following trend, HF > B97D > B3LYP, for the $\left|\Delta d(N,H)\right|$ values.
- HF consistently overestimates d(N,H) and, on average, the $\left|\Delta d(N,H)\right|$ value of ~0.22 Å is an order of magnitude larger when compared with data obtained at B3LYP for all conformers of **HL** and **H_2L**.

Table 3.4. Relative to MP2 values of d(N,H), performance of HF, B3LYP and B97D in terms of Δd(N,H) obtained for top LECs of HLp (all values in Å)

Conformer	Contact	d(N,H)	Δd_{HF}	Δd_{B3LYP}	Δd_{B97D}
C_p03	N28--H29	1.741	0.214	0.029	–0.060
	N26--H2	2.136	0.136	0.032	0.107
	N27--H2	2.379	0.138	–0.044	–0.131
C_p02	N28--H2	1.750	0.219	0.025	–0.072
	N26--H1	2.040	0.127	0.038	0.081
	N27--H1	2.478	0.066	0.025	–0.001
C_p01	N28--H29	1.740	0.227	0.016	–0.079
	N26--H1	2.073	0.125	0.036	0.086
	N27--H1	2.398	0.052	–0.029	–0.043
C_p05	N27--H2	1.680	0.276	0.031	–0.061
	N28--H1	2.470	0.303	0.020	–0.050
	N26--H2	2.482	0.053	0.050	0.078
	N28--H2	2.646	0.008	0.038	0.033
C_p04	N28--H2	1.740	0.242	0.032	–0.072
	N26--H29	2.101	0.162	0.012	0.077
	N27--H29	2.512	0.157	–0.005	–0.031
C_p09	N27--H2	1.689	0.245	0.001	–0.061
	N28--H1	1.989	0.563	0.003	–0.140
C_p08	N27--H1	1.717	0.315	0.036	–0.083
C_p06	N27--H29	1.691	0.286	0.043	–0.018
	N28--H2	1.944	0.914	0.016	–0.124
	Overall Average for |Δd|:		**0.220**	**0.023**	**0.065**
	Overall StDev for |Δd|:		**0.183**	**0.014**	**0.032**
	Overall Average for Δd:		***0.220***	***0.019***	***0.006***
	Overall StDev for Δd:		***0.183***	***0.019***	***0.073***
	Combined data for HL_p and HL_s				
	Overall Average for |Δd|:		**0.220**	**0.036**	**0.082**
	Overall StDev for |Δd|:		**0.159**	**0.020**	**0.040**
	Overall Average for Δd:		***0.216***	***0.032***	***0.009***
	Overall StDev for Δd:		***0.164***	***0.026***	***0.092***

- Although, in most cases, B3LYP also overestimates d(N,H), it consistently generates the averaged $\left|\Delta d(N,H)\right|$ of ~0.02 Å which, in turn, is about halve of what we obtained at B97D.
- B97D consistently shows somewhat larger number of underestimated diatomic distances, particularly in the case of **HL_p** conformers.

Focusing now on analysis of Δd(H,H), we found the following:

- HF was unable to reproduce the MP2-generated CH--HC contacts in many cases, particularly in the **H_2L_{pp}** conformers.
- Generally, the magnitude of deviation from the MP2 values for d(H,H) follows the HF > B3LYP > B97D trend.
- Both, HF and B3LYP, overestimate d(H,H) values, but B97D shows an almost equal number of positive and negative deviations.

In summary, (i) B3LYP performs significantly better in reproducing an overall geometry of the lowest in energy conformers than B97D; this is because B3LYP is better in predicting the strongest interactions, NH•••N, which play a decisive role in determining the ALPs structures, (ii) HF is not suitable for a rigorous analysis as it performed much inferior when compared with B3LYP or B97D, and (iii) if analysis of the CH--HC contacts was the main focus of an investigation then, not surprisingly, one should use B97D rather than B3LYP.

3.4.2. Comparative topological analysis

As an example, the computed differences, *e.g.*, at HF, $\Delta\rho_{CP}^{HF} = \rho_{CP}^{HF} - \rho_{CP}^{MP2}$, for a number of LECs of **HL_p** are shown in Table 3.5 where, in addition, the ρ_{CP} values obtained at MP2 as well as the combined data for both, **HL_p** and **HL_s**, tautomers are also included.

The full data sets for both tautomers of **HL** and **H_2L** are presented in Table A15-A21 in Appendix A. The analysis of all $\Delta\rho_{CP}^{LoT}$ values, obtained for the NH•••N interactions, reveals that:

- A lot of interactions observed at MP2 were not recovered at HF, *e.g.*, only 1 among 16 was found in the case of C_{pp} conformers.
- All MP2 NH•••N interactions were reproduced at B3LYP but few are missing at B97D.
- The magnitude of deviation from MP2 values, $\Delta\rho_{CP}^{LoT}$, generally follows the $\Delta\rho_{CP}^{HF} > \Delta\rho_{CP}^{B97D} > \Delta\rho_{CP}^{B3LYP}$ trend.
- On average, HF significantly underestimates the values of ρ_{CP} (they differ already at the second decimal place of a.u. when compared with those obtained at MP2) whereas B97D generates somewhat overestimated data, particularly in the case of **HL_p** (typically on third decimal place of au).

- The best performing B3LYP reproduced the MP2 ρ_{BCP} values well, typically, average absolute difference, avr($|\Delta\rho_{CP}^{B3LYP}|$), is characterized by low numbers on the third decimal

Table 3.5. Relative to MP2 ρ_{CP} values, performance of HF, B3LYP and B97D in terms of $\Delta\rho_{CP}$ for NH•••N interactions in top LECs of HL_p (all values in a.u.). In addition, the combined data for HL_p and HL_s is provided.

Conformer	**Interaction**	ρ_{CP}	$\Delta\rho_{CP}^{HF}$	$\Delta\rho_{CP}^{B3LYP}$	$\Delta\rho_{CP}^{B97D}$
C_p03	N28•••H29	0.0517	–0.0215	–0.0025	0.0102
	N26•••H2	0.0241	–0.0057	–0.0016	–0.0047
	N27•••H2	0.0153	–0.0040	0.0007	0.0036
C_p02	N28•••H2	0.0507	–0.0213	–0.0021	0.0115
	N26•••H1	0.0287	–0.0071	–0.0022	–0.0043
C_p01	N28•••H29	0.0521	–0.0224	–0.0010	0.0128
	N26•••H1	0.0269	–0.0064	–0.0019	–0.0042
	N27•••H1	0.0154	–0.0025	0.0000	0.0005
C_p05	N27•••H2	0.0614	–0.0305	–0.0035	0.0118
	N28•••H1	0.0136	–	–0.0007	0.0008
C_p04	N28•••H2	0.0518	–0.0234	–0.0028	0.0120
	N26•••H29	0.0258	–0.0070	–0.0008	–0.0037
	N27•••H29	0.0126	–	–0.0005	0.0000
C_p09	N27•••H2	0.0601	–0.0278	0.0007	0.0115
	N28•••H1	0.0308	–0.0197	0.0001	0.0117
C_p08	N27•••H1	0.0559	0.0238	–0.0037	0.0143
C_p06	N27•••H29	0.0599	–0.0305	–0.0050	0.0046
	N28•••H2	0.0336	–	–0.0007	0.0117
	Overall Average($\|\delta\rho_{CP}\|$):		**0.0169**	**0.0017**	**0.0074**
	Overall StDev($\|\delta\rho_{CP}\|$):		**0.0102**	**0.0014**	**0.0049**
	Overall Average($\delta\rho_{CP}$):		***–0.0137***	***–0.0015***	***0.0056***
	Overall StDev($\delta\rho_{CP}$):		***0.0144***	***0.0016***	***0.0070***
Combined data for HL_p and HL_s					
	Overall Average($\|\delta\rho_{CP}\|$):		**0.0151**	**0.0023**	**0.0063**
	Overall StDev($\|\delta\rho_{CP}\|$):		**0.0097**	**0.0015**	**0.0040**
	Overall Average($\delta\rho_{CP}$):		***–0.0088***	***–0.0008***	***0.0037***
	Overall StDev($\delta\rho_{CP}$):		***0.0158***	***0.0027***	***0.0066***

place of a.u., *e.g.*, 0.0017±0.0014 and 0.0030±0.0019 a.u. was found for **HL_p** and **HL_s**, respectively. For comparison, at B97D we obtained avr($|\Delta\rho_{CP}^{B97D}|$) of 0.0074±0.0049 and 0.0066±0.0038 a.u. for **HL_p** and **HL_s**, respectively.

Similar analysis was performed for the CH•••HC interactions – see data displayed in Tables 3.6 and A19-A20 in the Appendix A. The most significant difference, when compared with trends found for the NH•••N interactions, is the fact (as one would expect) that the MP2 ρ_{CP} values were best reproduced at B97D and we found the following trend, $\Delta\rho_{CP}^{HF} > \Delta\rho_{CP}^{B3LYP} > \Delta\rho_{CP}^{B97D}$.

Table 3.6. Relative to MP2 ρ_{CP} values, performance of HF, B3LYP and B97D in terms of $\Delta\rho_{CP}$ for CH•••HC interactions in top LECs of HL_p. In addition, the combined data for HL_p and HL_s is provided

Conformer	**Interaction**	ρ_{CP}	$\Delta\rho_{CP}^{HF}$	$\Delta\rho_{CP}^{B3LYP}$	$\Delta\rho_{CP}^{B97D}$
C_p03	H8•••H15	0.0134	–0.0026	–0.0024	–0.0016
	H15•••H22	0.0105	–	–0.0009	0.0001
C_p02	H4•••H15	0.0105	–0.0021	–0.0019	–0.0002
C_p01	H5•••H14	0.0113	–0.0023	–0.0024	–0.0004
	H5•••H21	0.0046	–0.0018	–	–0.0002
C_p20	H5•••H14	0.0110	–0.0049	–0.0024	–0.0002
C_p24	H4•••H15	0.0093	–0.0019	–0.0018	0.0000
C_p23	H4•••H15	0.0089	–0.0017	–0.0015	0.0001
	Overall Average($\lvert\delta\rho_{CP}\rvert$):		**0.0022**	**0.0017**	**0.0004**
	Overall StDev($\lvert\delta\rho_{CP}\rvert$):		**0.0014**	**0.0008**	**0.0005**
	Overall Average($\delta\rho_{CP}$):		***–0.0025***	***–0.0019***	***–0.0003***
	Overall StDev($\delta\rho_{CP}$):		***0.0011***	***0.0006***	***0.0006***
Combined data for HL_p and HL_s					
	Overall Average($\lvert\delta\rho_{CP}\rvert$):		**0.0023**	**0.0018**	**0.0003**
	Overall StDev($\lvert\delta\rho_{CP}\rvert$):		**0.0009**	**0.0005**	**0.0004**
	Overall Average($\delta\rho_{CP}$):		***–0.0023***	***–0.0018***	***–0.0003***
	Overall StDev($\delta\rho_{CP}$):		***0.0009***	***0.0005***	***0.0004***

It is important to note that results of our comparative geometrical and topological analysis are in agreement. Therefore, we recommend B3LYP as the optimum level of theory among those we have examined for reproducing MP2 geometries and topological properties of low energy conformers of ALPs. Our recommendation is strengthened by the finding that the NH•••N interaction, which is most dominant, hence determines the structure adopted by conformers, is best reproduced at B3LYP.

3.5. Conclusions

From an in depth analysis (involving structural, QTAIM and NCI properties) of hundreds of conformers of **HL** and **H_2L** *trien* tautomers, a combined (MM/DFT) conformational search protocol (CSP) was developed to generate representative sets of lowest energy conformers (LECs) of aliphatic linear polyamines (ALPs) in reasonable time. For the purpose of this study, sets of lowest, medium and higher energy conformers were investigated at HF, B3LYP, B97D and MP2 levels of theory with an aim of finding out structural and topological preferences in implicit aqueous environment.

Considering structural preferences, we found common numerous trends for both, mono- and di-protonated tautomers, such as (i) when a single NH--N short contact is formed (this is characteristic for higher energy conformers) then it always involves either $–NH_3^+$ or $–NH_2^+$ group, hence the protonated N-atom, (ii) when more than one intramolecular NH--N contact is observed then H-atoms of both protonated groups are always involved, (iii) LECs are characterized by largest possible number of the intramolecular NH--N short contacts and most stable conformers form largest possible rings.

Furthermore, (i) when a single terminal group is protonated, then the largest possible, 11m-MR is preferentially formed. This feature is observed among ten **HL_p** and three **H_2L_{ps}** LECs, (ii) when a single –NH–group is protonated, then largest possible 8m-IR is preferentially formed; this is observed among top six lowest energy **HL_s** conformers and (iii) with two terminal groups being protonated; only 5m-IRs are formed and, typically, three consecutive such rings are present among the LECs (all N-atoms in a molecule are involved). Finally, results obtained showed that large rings contribute in stabilizing manner to molecular energy more than smaller rings. We found significant differentiation in relative energies of **HL_p**, **HL_s** and **H_2L_{ps}** conformers for which small sets of LECs, all with either 11m- or 8m-rings, are observed; the **H_2L_{pp}** conformers form only 5m-IRs and the more rings the more stable conformer is formed.

QTAIM-based topological analysis revealed the presence of many atomic interaction lines (AILs), in LECs in particular, which in most cases could be seen as ‘legitimate’ or an orthodox bond path. In general, most but not all NH--N short contacts were linked by AILs; in some instances contacts with longer d(N,H) were linked by AIL even though a shorter contact was present. Excellent relationships between ρ_{CP} and d(N,H) were found for all tautomers but, more importantly, combined data sets from B3LYP and MP2 also resulted in near perfect exponential relationships for each tautomer. Quite unexpectedly, we found numerous AILs linking CH--HC intramolecular contacts; they were present mainly in the LECs of the mono-protonated forms

but, even more unexpectedly, they were absent in the ten highest in energy conformers of **HL** tautomers. On the other hand, these CH•••HC interactions were completely absent in the conformers of the $\mathbf{H_2L_{ps}}$ tautomer and only were found in four highest in energy conformers of the $\mathbf{H_2L_{pp}}$ tautomer. To gain further insight, we used NCI technique. It revealed the presence of many additional intramolecular NH--N as well as CH--HC interactions. 3D NCI-defined isosurfaces showed that there is no qualitative difference between AIL-linked and AIL-absent contacts of the same kind, NH--N or CH--HC. Also, following classical interpretation, NCI isosurfaces revealed an increase in an intramolecular strain with a decrease in the size of a ring formed by the NH--N; 11m_MR appeared to be strain free whereas 5m-IR showed significant degree of strain. Surprisingly, in all cases of 5m-IR formed by the CH--HC interaction, no strain was observed. It is rather difficult to rationalize significance of QTAIM-defined AILs and NCI-defined isosurfaces in the interatomic region of the CH--HC contacts in terms of their (de)stabilizing energy contribution to a molecular system as these two phenomena simply represent a localized density increase between atoms involved; hence, we embarked on an extensive and dedicated for the purpose separate studies to explain fully the chemical nature and energy contribution made by these and some other totally unexpected contacts discovered mainly in the lowest energy conformers. One might add to this that an attempt to correlate summed density values at critical point against relative energy of conformers failed to generate good quality relationships (regardless whether only the strongest NH•••N or combined, NH•••N plus CH•••HC, interactions were accounted for).

Significance and usefulness of the developed conformational search was confirmed here as we were able to predict (i) a mixture of $\mathbf{HL_p}$ and $\mathbf{HL_s}$ conformers in accord with experimental data reported recently[5]; formation of $\mathbf{HL_s}$ as the only form was predicted when linear conformers were used, and (ii) a possibility of presence of both $\mathbf{H_2L}$ tautomers was also recovered, although a significantly different relative %-fractions were obtained here when compared with cluster expansion analysis of NMR data[8]; using linear conformers, formation of $\mathbf{H_2L_{ps}}$ was predicted as the only species formed. Clearly, this work strongly supports the view that conformational search must be seen as a prerequisite for theoretical studies of molecules characterized by almost an infinite conformational freedom, such as ALPs.

Finally, from the extensive analysis of all the data obtained at different levels of theory it became clear that (i) B3LYP was best reproducing MP2 data; it can be reliably used for this type of investigations and only few (typically 3 to 5) LECs could (should) be optimized and analysed at MP2, and (ii) HF should rather be avoided as it has not performed well neither in the search of LECs nor was able to recover topological properties.

3.6. References

[1] E. Agostinelli, M. P. M. Marques, R. Calheiros, F. P. S. C. Gil, G. Tempera, N. Viceconte, V. Battaglia, S. Grancara, A. Toninello, *Amino Acids*., **2010**, *38*, 393–403.

[2] U. Bachrach, *Plant Phys. Biochem*., **2010**, *48*, 490–495.

[3] L. A. E. Batista de Carvalho, M. P. M. Marques, J. Tomkinson, *J. Phys. Chem. A*., **2006**, *110*, 12947–12954.

[4] P. Paoletti, M. Ciampolini, A. Vacca, *J. Phys. Chem*., **1963**, *67*, 1065–1067.

[5] D. N. Hague, A. D. Moreton, *J. Chem. Soc. Perkin Trans*., **1994**, *2*, 265–270.

[6] S. P. Dagnall, D. N. Hague, M. E. McAdam, *J. Chem. Soc. Perkin Trans*., **1984**, *2*, 1111–1114.

[7] M. Delfini, A. L. Segre, F. Conti. R. Barbucci, V. Barone, P Ferruti, *J. Chem. Soc. Perkin Trans*. , **1980**, *2*, 900–903.

[8] M. Borkovec, D. Cakara, G. J. M. Koper, *J. Phys. Chem. B*., **2012**, *116*, 4300–4309.

[9] G. R. Hedwig, H. K. J. Powell, *J.C.S. Dalton*., **1973**, 793–797.

[10] R. Barbucci, L. Fabbrizzi, P. Paoletti, *J.C.S. Dalton*., **1972**, 745–749.

[11] M. M. Kimberly, J. H. Goldstein, *Anal Chem*., **1981**, *53*, 789–793.

[12] I. Cukrowski, C. F. Matta, *Comput. Theoret. Chem*., **2011**, *966*, 213–219.

[13] M. P. M. Marques, L.A.E. Batista de Carvalho, *Biochem. Soc. Trans*., **2007**, *35*, 374–380.

[14] K. K. Govender, I. Cukrowski, *J. Phys. Chem. A*., **2009**, *113*, 3639–3647.

[15] K. K Govender, I. Cukrowski, *J. Phys. Chem. A*., **2010**, *114*, 1868–1878.

[16] M. L. Reyzer, J. S. Brodbelt, *J. Am. Soc. Mass Spectrom*., **1998**, *9*, 1043–1048.

[17] C. A. Ilioudis, K. S. B. Hancock, D. G. Georganopoulou, J. W Steed, *New J. Chem*., **2000**, *24*, 787–798.

[18] L. A. E. Batista de Carvalho, L. E. Laurenc, M. P. M. Marques, *J. Mol. Struc-Theochem*., **1999**, *482–483*,639–646.

[19] L. Carballeira, RA. Mosquera, M. A. Rios, *J. Mol. Struc-Theochem*., **1988**, *176*, 89–105.

[20] L. Carballeira, R. A. Mosquera, M. A. Rios, C. A. Tovar, *J. Mol. Struc-Theochem*., **1988**. *193*, 263–277.

[21] M. J. Frisch, G. W. Trucks, HB et al, Schlegel, Gaussian 09, Revision D.1, 2009, Gaussian, Inc Wallingford CT.

[22] J. Tomasi, M. Persico, *Chem. Rev*., **1994**, *94*, 2027–2094.

[23] S. Miertusˇ, E. Scrocco, J. Tomasi, *Chem. Phys*., **1981**, *55*, 117–129.

[24] R. Cammi, J. Tomasi, *J. Comput. Chem*., **1995**., *16*, 1449–1458.

[25] J. J. Wilke, M. C Lind, H. F Schaefer, A. G Csaszar, W. D Allen, *J. Chem. Theoret. Comput*., **2009**, *5*, 1511–1523.

[26] M. Sanchez-Lozano, E. M. Cabaleiro-Lago, J. M. Hermida-Ramon, C. M. Estevez, *Phys. Chem. Chem. Phys*., **2013**, *15*, 18204–18216.

[26] R. F. W. Bader, *Atoms in Molecules*: *A Quantum Theory*, **1990**, Oxford University Press Oxford, UK.

[27] T. A. Keith, AIMALL (Version 13.11.04), TK Gristmill Software, **2013**, Overland Parks KS, USA, <aim.tkgristmill.com>.

[28] J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret,J. P. Piquemal, D. Beratan, W. J. Yang, *J. Chem. Theoret. Comput*., **2011**, 7, 625–632.

[30] Spartan ’10, **2010**, Wavefunction, Inc.: Irvine, CA.

[31] M. Arabieh, M. H Karimi-Jafari, M. Ghannadi-Maragheh, *J. Mol. Model*., **2013**, *19*, 427–438.

[32] S. Gronert, R.A. J. O’Hair, *J. Am. Chem. Soc*., **1995**, *117*, 2071–2081.

[33] A. M. Pendás, E. Francisco, M. A. Blanco, C. Gatti, *Chem. Eur. J*., **2007**, *13*, 9362–9371.

[34] I. Cukrowski, J. H. de Lange, A. S. Adeyinka, P Mangondo, *Comput. Theoret. Chem*., **2015**, *1053*, 60–76.

[35] A. Haaland, D. J. Shorokhov, N. V. Tverdova, *Chem. Eur. J*., **2004**, *10*, 441–4421.

[36] M. J. Jablonski, *J. Phys. Chem. A*., **2012**, *116*, 3753–3764.

[37] P. Dem’yanov, P. Polestshuk, *Chem. Eur. J*., **2012**, *18*, 4982–4993.

[38] E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J Cohen, W. J. Yang, *J. Am. Chem. Soc*., **2010**, *132*, 6498–6506.

[39] R Chaudret, B. de Courcy, J. Contreras-García, E. Gloaguen, A. Zehnacker-Rentien, M. Mons, J. P. Piqueml, *Phys. Chem. Chem. Phys*., **2014**, *16*, 9876–9891.

[40] J. Contreras-García, W. J. Yang, *J. Phys. Chem. A.*, **2011**, *115*, 12983–12990.

[41] M. T. Albelda, J. C. Frías, E. García-España, *Encyclopedia of Supramolecular Chemistry.*, **2007**, *1:1*, 1–37.

[42] A. Bencini, A. Bianchi, E. Garcia-España, M. Micheloni, J. A. Ramirez, *Coord. Chem. Rev.*, **1999**, *188*, 97–156.

[43] C. Frassineti, S. Ghelli, P. Gans, A. Sabatini, M. S. Moruzzi, A. Vacca, *Anal. Biochem.*, **1995**, *231*, 374–382.

[44] J. Weisell, J. Vepsäläinen, M. Peräkylä, *J. Phys. Org. Chem.*, **2013**, *26*, 360–366.

[45] A. Bondi, *J. Phys. Chem.*, **1964**, *68*, 441–451.

[46] R. S. Rowland, R. Taylor, *J. Phys. Chem.*, **1996**, *100*, 7384–7391.

[47] R. A Klein, *Chem. Phys. Lett.*, **2006**, *425*, 128–133.

Chapter 4

Evaluating the true nature of intramolecular interactions through IQA interaction energies and 1D cross-sections of the electron and deformation density distributions

4.1 Introduction

QTAIM and NCI techniques fall under the umbrella of QCT methods[1] and are based on a detailed examination of the topology of electron density. They are commonly used to detect and classify inter- and intramolecular interactions in molecular systems due to their relatively cheap computational cost.[2-4]. According to QTAIM, a bond path or in general an atomic interaction line (AIL), which is a bridge of maximal electron density, should be observed in the interatomic region of atoms where chemists expect a classical chemical bond to exist. Therefore, the presence of a QTAIM defined bond path has been extensively used by chemists to infer the presence of a chemical bond[5-8] or, at the very least, a bonding interaction.[9,10] Furthermore, the value of electron density at the BCP (ρCP) of a given AIL has been used as a measure of the strength of that bond. However, many cases have been reported in literature where an AIL is conspicuously absent where classical chemist expect a bonding interaction (or where a bond has been identified by other methods such as energy decomposition schemes[11]). Much more bewildering is the fact that there are several other cases in which an AIL has been observed between atoms believed to be involved in a steric repulsion interaction.[12-22] These problematic cases, have led to heated debates questioning either the validity and interpretation of QTAIM theory, or the meaning of classical chemical concepts such as steric repulsion. In fact certain authors have questioned the definition of the chemical bond itself since Bader emphatically stated that a bond path should be associated with bonding in all cases since it recovers all Lewis structures.[23]

Non covalent interaction index (NCI) solved one aspect of QTAIM-interpretation problems by demonstrating that electron density may be concentrated in the bonding region of an interaction despite the absence of an AIL. NCI reveals inter- and intramolecular interactions by locating regions of low electron density where the reduced density gradient (RDG) approaches (or tends) to zero.[24] These regions are then classified as either stabilizing or destabilizing depending on whether electron density is accumulated or depleted within the identified region.[24-26, 27,28] NCI therefore complements typical QTAIM analysis in three ways; (i) it locates an interaction in 3D space, whereas QTAIM can only identify a bridge of maximal electron density, (ii) it can detect electron density accumulation whether an AIL is present or not, and (iii) it can also identify interactions resulting from electron depletion. In addition, NCI can be used to approximately analyse interactions using the promolecular densities (i.e. electron density based only on the geometries without the need for an electronic structure calculation) thereby allowing it to be used for very large biological systems. Notwithstanding, NCI is marred by the same problem of interpretation of with regards to controversial interactions (such as CH•••HC

interactions in the bay of biphenyl) because it will always show a region of concentration wherever an AIL is found.[29] However, due to their relative simplicity, insight and low computational cost, both NCI and QTAIM methods remain attractive tools for computational chemists. Their only disadvantage is the fact that interpretation of results obtained is still very unclear especially when investigating potentially new and controversial interactions.

Bader's original interpretation (the so-called orthodox interpretation) of an AIL or bond path is that it signals a bonding interaction[30] and only at equilibrium geometry (where the net forces acting on atoms is zero), can it be related specifically to a chemical bond. To arrive at this interpretation, firstly he pointed[30] out that the sign of the eigenvalues λ_1, λ_2 and λ_3 of the Hessian matrix (the ordered matrix of the second derivative of the electron density) as well as the sum of these eigenvalues, the Laplacian ($\nabla^2\rho(\mathrm{r})$) can be related to the concentration or depletion of electron density in a specific axis or at a point, respectively. Specifically, when $\nabla^2\rho(\mathrm{r}) > 0$, the second order change in the electron density is positive at ***r***, the density at ***r*** is less than the average of the its surrounding density, and the electron density is said to be depleted; similarly, electron density is concentrated at ***r*** when $\nabla^2\rho(\mathrm{r}) < 0$. The same can be said of the individual component eigenvalues of the Hessian matrix, but along a specific axis. For an AIL to be present, the topological condition that density must be depleted along the AIL ($\lambda_1>0$), but concentrated across it (λ_2 and $\lambda_3<0$) must be satisfied. However since by convention $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and $\lambda_1>0$ between any two atoms, therefore λ_2 is of particular importance because its sign will determine whether electron density is concentrated or depleted across an interaction. NCI classification of inter- and intramolecular interactions as (de)stabilizing is also based on this premise. The concentration of electron density (as measured by the second-derivative of the electron density) was then linked by Bader through the local statement of the virial theorem to arrive at his concept of bonding[30,31] in which he suggested that a build-up of charge is observed in the bonding region upon formation of a chemical bond.[32-36] Consequently, the presence of an AIL (a bridge of maximal concentration of electron density),which is due to a concentration of electron density, can be regarded as a bonding interaction. Similar interpretation has been applied in the development of the NCI technique, with regions of concentration designated as stabilizing and attractive, and regions of depletion often referred to as steric strain.[24,27-29]

However there is no general consensus at present as to how this "build-up" of charge can be measured. Hence, several contentious questions which have far reaching consequences on the use of QTAIM and NCI methods as interpretative tools to understand structure and reactivity in chemical systems arises such as (i) Is the concentration or depletion of electron density, as used

by the orthodox interpretation of an AIL or NCI regions of interactions truly indicative of increased or decreased electron density in the bonding region, respectively, and thus synonymous with energetic stabilization or destabilization? and (ii) Is it a measure of the electron density of a small region relative only to its environment, or can it be linked with the electron density relative to an unbound state? We seek answers to some of these questions in this chapter by a thorough investigation of electron density and its variation in a number of intramolecular interactions discovered in representative low energy conformers of triethylenetetramine.

Triethylenetetramine (2,2,2-tet) is a member of the homologous series of aliphatic linear polyamines (ALPs), most of which are found in living organisms and play important roles in regulating cell proliferation and differentiation.[37,38] It is also a well-known copper chelator and has been used extensively in the treatment of Wilson's disease.[39,40] A variety of intramolecular interactions (*e.g.*, NH•••N, CH•••HC, CH•••N etc.) have been identified in its protonated forms making it suitable for carrying out this kind of investigation. QTAIM[31] and NCI[24-26] analyses was compared with results from the Interacting quantum atoms (IQA) energy decomposition techniques. In addition, an in-depth analysis of the one dimension cross-section of the electron density was examined. Furthermore, the use of the deformation density as an alternative measure of charge "build-up" in the bonding region was investigated. Finally we discuss the implications of our findings for conformational preference in large molecular systems such as aliphatic linear polyamines.

The work in this chapter has been extracted from a published paper[41] where intermolecular interactions in water dimer as well as intramolecular interactions in both nitrilopropanoic (NTPA) and bipyridine (bpy) was also investigated using the same approach discussed here. Hence some of our discussion here overlaps to some extent with the results obtained for these molecules. Representative structures of these molecules are shown in Figures B1-B3 in Appendix B.

4.2. Methods and computational details

All geometry optimizations and electronic structure calculations were performed in Gaussian 09, revision D[42] at the RMP2/6-311++G(d,p) level of theory in solvent (PCM/UFF). QTAIM and IQA analyses were carried out in AIMAll[43] using the Proaim integration algorithm with very high angular quadrature outside the beta sphere for IQA calculations. NCI calculations were carried out using NCIPlot 2.0[26] and corresponding isosurfaces were visualized in VMD

1.9.1.[44] Finally, one-dimensional cross-sections of the electron and deformation densities along λ_2 eigenvectors were performed using in-house software.

In order to calculate the cross-sections of the electron and deformation densities, the geometric interaction point (GIP) was determined which corresponds to the point of lowest density directly between two nuclei. The eigenvector corresponding to the λ_2 eigenvalue of the Hessian matrix was then calculated, and two new coordinates were generated at a specific distance (usually 0.05 au) in both directions along this vector. The electron densities were then recorded at these points and new coordinates generated based on the eigenvectors corresponding to the λ_2 eigenvalues at these points. This process was repeated until a pre-determined length (usually 2 Å in both directions) was reached. For brevity, the entire path followed through this process is referred to the λ_2-eigenvector. The λ_2-eigenvector therefore will always originate from the GIP (which occasionally may coincide with a QTAIM-defined bond critical point, BCP) and will pass through any corresponding NCI-defined interaction critical point (ICP) or BCP, ring critical point (RCP) and cage critical point (CCP). However, due to congestion of intramolecular interactions in 2,2,2-tet, the λ_2-eigenvector corresponds to the cross-sections of multiple interactions and in such cases the path followed for the cross-section was calculated as a straight line along the initial λ_2-eigenvector at the GIP. In order to calculate the cross-sections of the deformation density, the electron density for each fragment was calculated along the same λ_2-eigenvector as for the cross-section of the molecular electron density. The cross-section of the molecular density was then subtracted from the sum of fragment densities to give the cross-section of the deformation density.

Deformation densities of monoprotonated aliphatic polyamine, triethylenetetramine (2,2,2-tet) were calculated by fragmenting each conformer into three radical fragments (corresponding to two duplet fragments, $\bullet(CH_2)_2(NH_3)$ and $\bullet(CH_2)_2(NH_2)$, and a triplet $\bullet(NH)(CH_2)_2(NH)\bullet$ fragment, as shown in Figure 4.1. Further information about the calculation of deformation

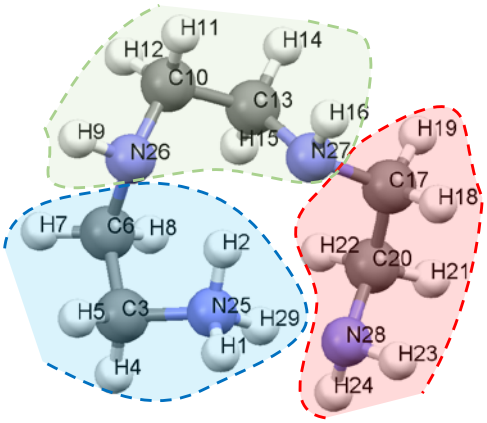


Figure 4.1. Radical fragments used in the calculation of the deformation density in the LEC of 2,2,2-tet.

densities for other molecules discussed here can be found in the published paper.[41]

4.3 Results and Discussion

4.3.1 Interpretation of AIL and NCI-defined Isosurfaces

Molecular graphs of selected lower (L1) and higher (L2) energy conformers of 2,2,2-tet examined in this study is shown in Figure 4.2. Both conformers show a very strong (or leading) classical intramolecular NH•••N hydrogen bonds, with an AIL present, but these conformers differ in the overall congestion of the molecule. As a result, L1 forms an additional NH•••N interaction (with an AIL) and both conformers have unexpected CH•••HC interactions also indicated by the presence of AILs. Similarly, the presence of an AIL has been observed for O•••O interaction at various interatomic distances in water dimer arranged to simulate steric repulsion. Numerous CH•••HC interactions all signified by the presence of AILs were also found in both bpy and NTPA. All these examples suggest that the presence and significance of AILs between two atoms as well as their interpretation is not quite obvious since AILs appeared

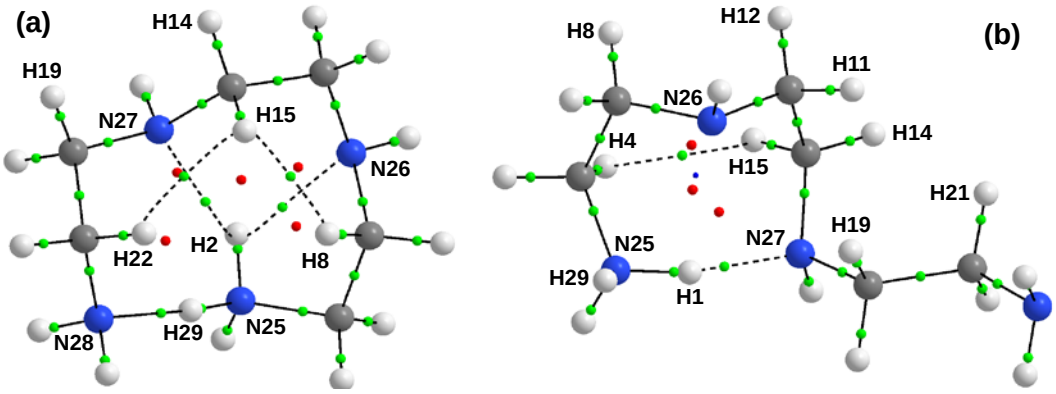


Figure 4.2. Molecular graphs of a) the lower energy L1 and b) higher energy L2 conformer of 2,2,2-tet

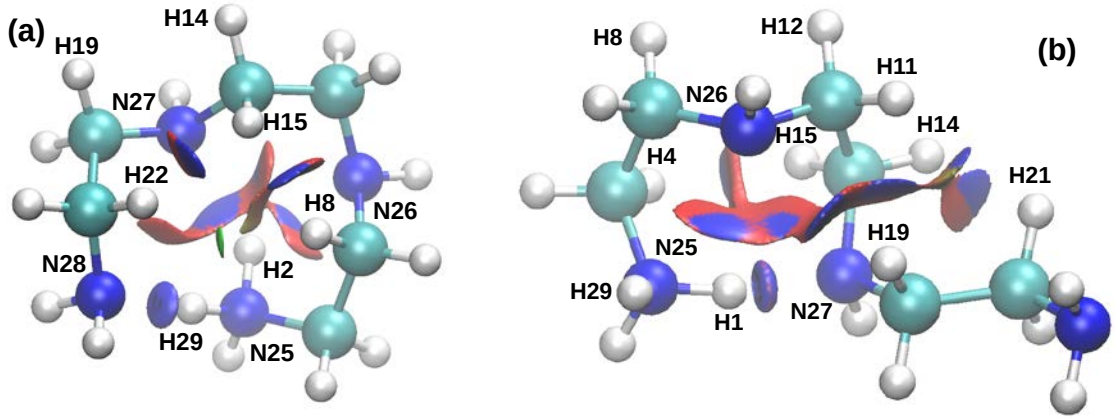


Figure 4.3. NCI isosurfaces of a) the lower L1 and b) higher energy L2 conformer of 2,2,2-tet with a RDG isovalue = 0.5 au and isosurfaces coloured from blue to red using $-0.03 \leq \rho(r) \times \text{sign}(\lambda_2) \leq +0.03$ au.

between atoms involved in both, classically attractive and repulsive interactions in these cases. As one would expect, this ambiguity in meaning of AIL also affects NCI defined isosurfaces hence blue region of electron density concentration was observed for the O•••O and CH•••HC interactions.

The NCI-plots for each conformer of 2,2,2-tet studied in this work, shown in Figure 4.3, disclose an abundance of intramolecular interactions involving N-and H-atoms. Focusing on the CH•••HC interactions, it can be seen that besides those with the presence of AILs, there also exist interactions just showing regions of concentration as well as regions of depletion according to NCI-defined isosurfaces indicators. Cross-sections of the electron density for the NH•••N, CH•••N and CH•••HC interactions are shown in Figures 4.4(a-c).

A clear concentration and local maximum in the density is seen in L1 for the leading NH29•••N28 interaction, whilst only a slight increase in the density is observed for the weaker NH2•••N27 interaction; both maxima correspond to the observed BCP. The difference between

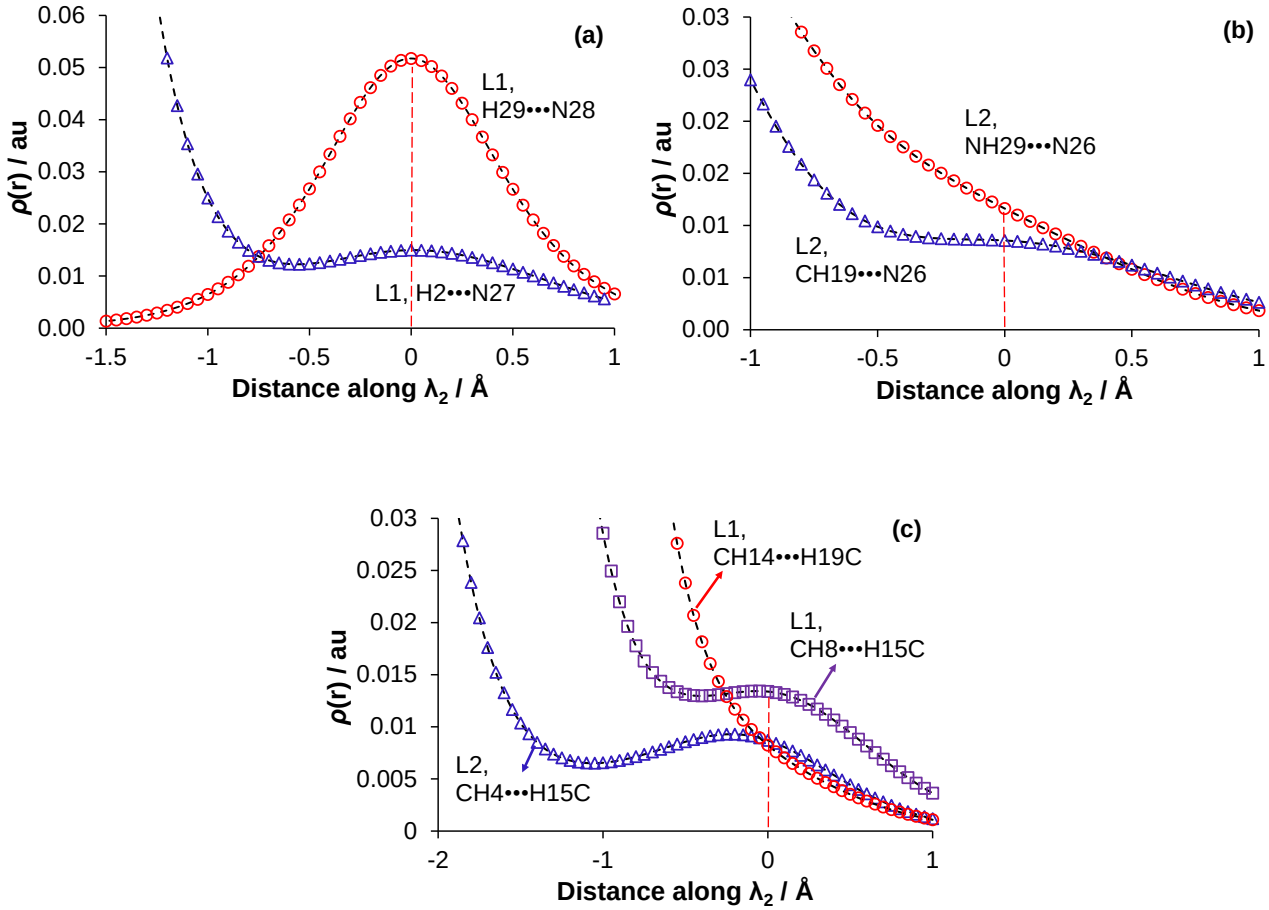


Figure 4.4. Cross-sections of the electron density along the λ_2-eigenvector for indicated a) H•••N interactions with an AIL, b) XH•••N interactions without an AIL and c) CH•••HC interactions in the lower energy conformer, LEC, of 2,2,2-tet.

these two interactions, in terms of density cross-sections, might be rationalized in terms of the local environment; NH29•••N28 is on the 'outskirts' of a molecule, whereas NH2•••N27 occurs within the congested ring in the presence of numerous other interactions. For both interactions $\Delta\rho(\boldsymbol{r}) > 0$, i.e. they result from an inflow of density into the interatomic region of atoms involved. Also, Figure 4.5(a) shows a slight variation in the shape of deformation densities for these two interactions. This suggests that molecular environment of an interaction might also play an important role in determining its nature. Two other H•••N interactions, CH19•••N26 and NH29•••N26 in L2, but without AILs present, are presented in Figure 4.4b. Regions of concentration and depletion are observed in the 1D cross-section of their electron density along the λ_2 eigenvector, respectively, but the deformation density cross-sections in Figure 4.5b show a very slight inflow of density for CH19•••N26 but only an outflow of density for NH29•••N26.

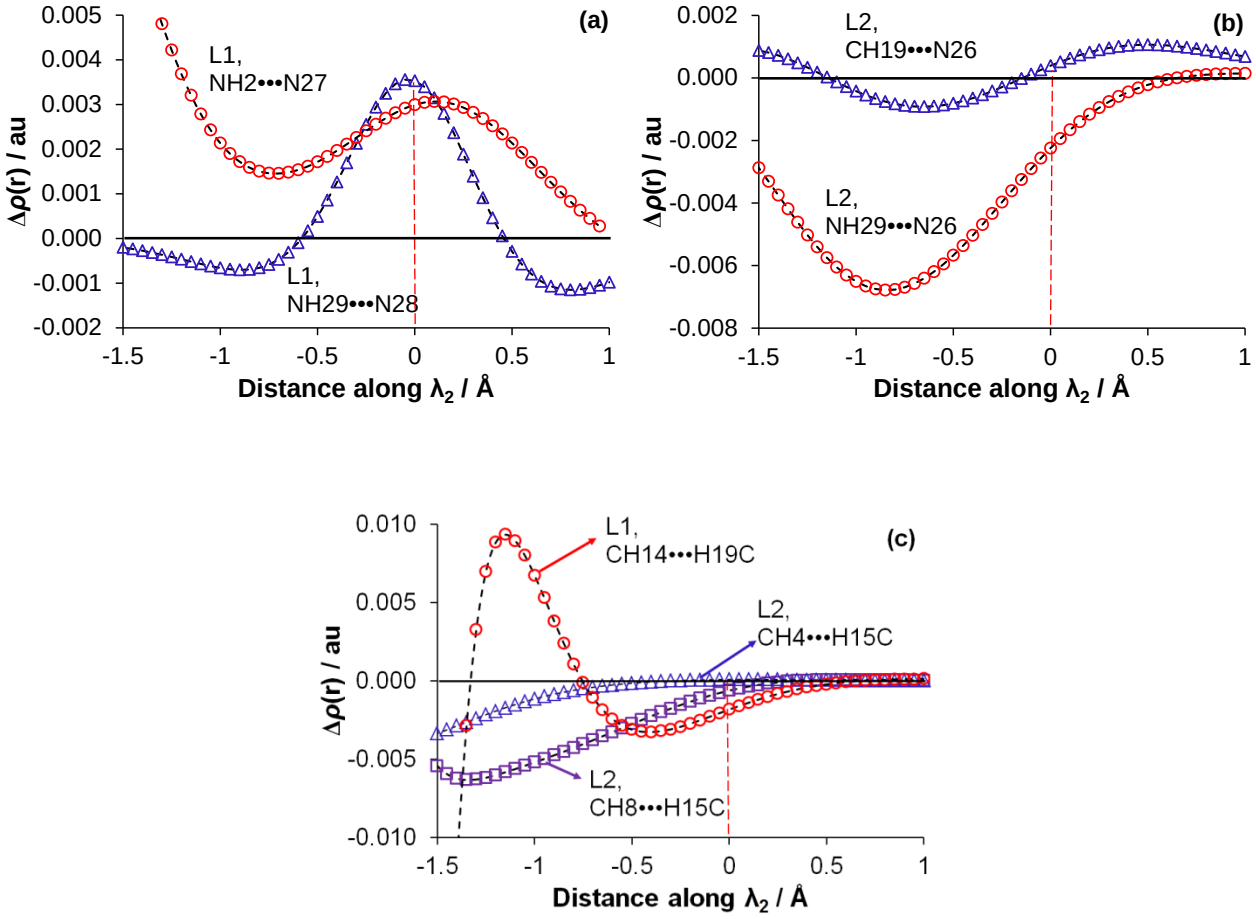


Figure 4.5. Cross-sections of the deformation density along the λ_2-eigenvector for indicated a) XH•••N with AIL present, b) XH•••N without AIL present and c) CH•••HC interactions in the lowest energy conformer of 2,2,2-tet. Red dashed lines indicate the GIP.

Similar phenomenon was observed when water dimer was arranged so as to simulate both OH•••O classical hydrogen bond and O•••O steric clashes (see Figure 4.6). Although the 1D-

cross section of the electron density of both OH•••O and O•••O interactions show regions of concentration corresponding to the appearance of an AIL, the deformation density cross section shows an inflow of density for the classical OH•••O classical hydrogen bond but an outflow of density as one would expect for O•••O steric clash. (see Figure 4.6).

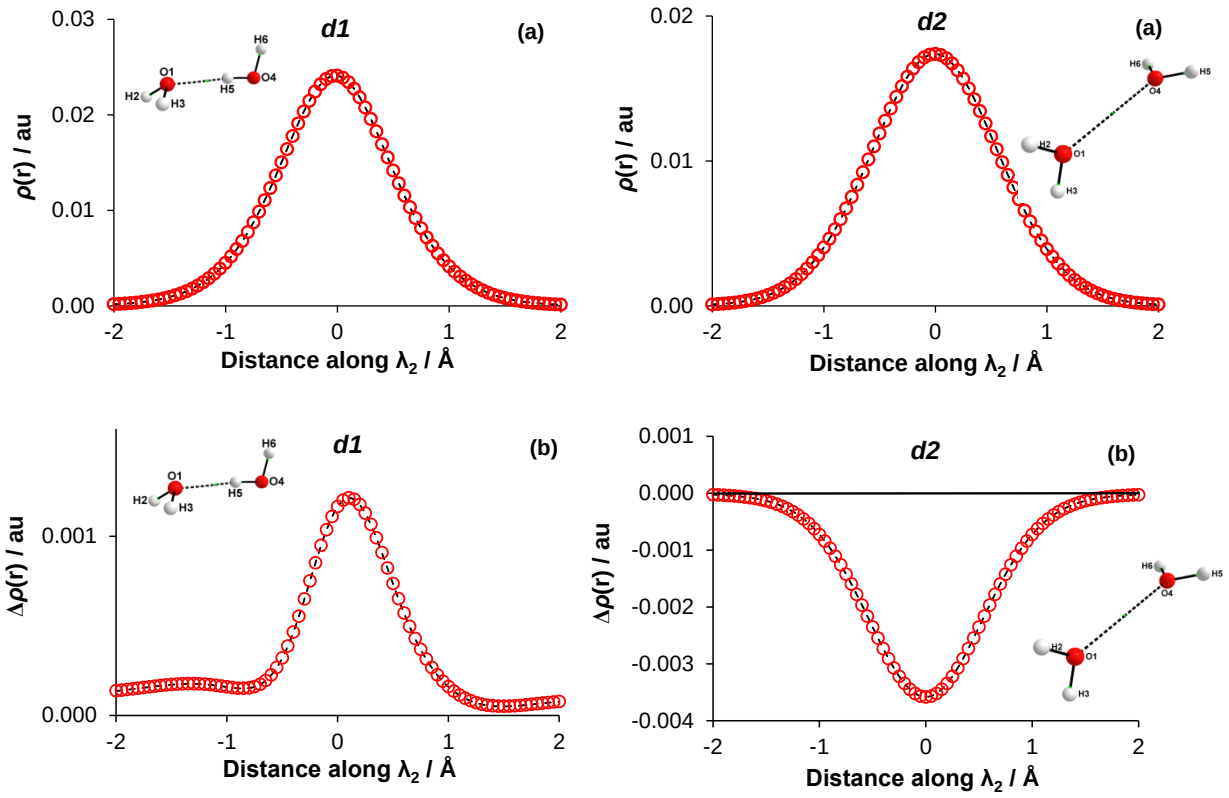


Figure 4.6. Cross section of (a) the electron density and (b) the deformation density along λ_2 eigenvector for ***d1*** at d(O--H) = 1.946 Å and ***d2*** at d(O--O) = 2.6 Å. The origin of the cross-sections for ***d1*** and ***d2*** are the BCP(O1,H5) and BCP(O1,O4), respectively.

These two examples show clearly that the presence of an AIL might be due to either an outflow or inflow of electron density as opposed to the traditional interpretation according to QTAIM that it is due to concentration of electron density in the interatomic region ultimately leading to stabilization.

A large spectrum of various H•••N interactions with different topological indices are observed in these molecules due to their structural arrangement. The density cross-sections of selected CH•••HC interactions (Figure 4.4c) show regions of clear increase in the electron density corresponding to the presence of an AIL and the absence of any shoulders showing regions of depletion. The deformation densities (Figure 4.5c) show a slight outflow of density in the interaction regions but much larger changes in the neighbouring regions between C atoms except

for CH14•••H19C interaction, whose cross-section passes close to lone pair of a nitrogen atom. Similar variation in both electron and deformation densities were also observed in three CH•••HC interactions found in NTPA conformers.

The full list of interacting atoms (as identified by NCI) is shown in Table 4.1, together with their IQA interaction energies. It is important to emphasize that, while the interaction energies of all intramolecular interactions vary greatly, none of the CH•••HC interaction is repulsive and all are showing dominating contribution of the $V_{XC}^{A,B}$ term as observed in other molecules like bpy and NTPA that were studied in our published work; note that $E_{int}^{A,B} < 0$ is observed in all CH•••HC interactions identified in both LECs of 2,2,2-tet, regardless whether density concentration or depletion is observed in the bonding region, contradicting the MM-based notion that this kind of interaction is highly repulsive in nature.

Table 4.1. Analysis of interactions in the protonated lower (L1) and higher (L2) energy conformers of 2,2,2-tet and its protonated forms in terms of interaction energies and electron density in the interatomic region.

Form	Interaction	Atoms A,B	Distance Å	$E_{int}^{A,B}$ [a]	$V_{cl}^{A,B}$ [a]	$V_{XC}^{A,B}$ [a]	AIL	$\rho(r) \times \text{sign}(\lambda_2)$ [b]	$\Delta\rho(r)_{GIP}$ [c]
L1	NH•••N	H29,N28	1.741	–131.8	–107.5	–24.3	Yes	–0.0517	+0.0035
	NH•••N	H2,N27	2.379	–78.2	–73.5	–4.7	Yes	–0.0150	+0.0026
	NH•••N	H2,N26	2.136	–89.8	–81.6	–8.2	Yes	–0.0239	+0.0000
	CH•••HC	H8,H15	2.034	–3.6	+0.0	–3.6	Yes	–0.0134	–0.0006
	CH•••HC	H15,H22	2.133	–3.0	+0.1	–3.1	Yes	–0.0098	+0.0000
	CH•••HC	H14,H19	2.490	–0.9	+0.2	–1.1	No	+0.0082	–0.0018
L2	NH•••N	H1,N27	1.654	–145.8	–116.5	–29.3	Yes	–0.0656	+0.0144
	NH•••N	H29,N26	2.736	–61.4	–60.5	–0.9	No	+0.0116	–0.0022
	CH•••N	H19,N26	2.711	–5.0	–1.9	–3.1	No	–0.0086	+0.0004
	CH•••HC	H4,H15	2.156	–2.7	+0.1	–2.8	Yes	–0.0087	+0.0001
	CH•••HC	H8,H12	2.426	–1.1	+0.2	–1.3	No	+0.0082	–0.0017
	CH•••HC	H14,H21	2.273	–2.0	+0.1	–2.1	No	–0.0083	–0.0001
	CH•••HC	H11,H19	2.344	–1.4	+0.1	–1.5	No	–0.0083	–0.0004
	CH•••HC	H11,H21	2.553	–0.9	+0.0	–1.0	No	–0.0045	–0.0000

[a] Diatomic interaction energies and decomposed components, all in kcal/mol; [b] Values in au at GIP; [c]The deformation density in au at GIP.

All NH•••N interactions are characterized by large and negative interaction energies but H29•••N26 ($E_{int}^{A,B}$ = –61.4 kcal/mol in L2) does not have an AIL (most likely due to interatomic distance of 2.74 Å) and has a unique and somewhat unexpected set of NCI and deformation density indices, namely a local depletion in electron density ($\rho(\boldsymbol{r}) \times \text{sign}(\lambda_2)$ = +0.0116 au) and an outflow of density, $\Delta\rho(\boldsymbol{r}) < 0$. Furthermore, there is no significant inflow of density on the

formation of the H2•••N26 interaction in L1 even though it is the second strongest (in stabilizing manner), has large locally increased density ($\rho(\boldsymbol{r}) \times \text{sign}(\lambda_2)$ = –0.0239 au) and atoms involved are linked by the AIL. The data obtained for the NH•••N interactions show that (i) positive values of ($\rho(\boldsymbol{r}) \times \text{sign}(\lambda_2)$ and (ii) outflow or no change in the deformation density, $\Delta\rho(\boldsymbol{r}) \leq 0$ are not synonymous with destabilizing interaction; hence, former describes the resultant local density distribution and the latter explains the process of the resultant density formation, in- or outflow of density on an interaction formation and both these indices illustrate how a molecular system has minimized its energy in terms of density distribution.

Similar observations, related to significance of $\rho(\boldsymbol{r}) \times \text{sign}(\lambda_2)$ and $\Delta\rho(\boldsymbol{r})$, apply to CH•••HC interactions, all characterized by $E_{\text{int}}^{\text{A,B}} < 0$ with dominant $V_{\text{XC}}^{\text{A,B}}$ term and various combinations of resultant local density and its formation. For instance, even though an inflow of density is observed for both H8•••H15 and H14•••H19 in L1, only the former has $\rho(\boldsymbol{r}) \times \text{sign}(\lambda_2) < 0$ and AIL. In some other cases, $\Delta\rho(\boldsymbol{r})$ ~0 but density is locally increased with either AIL present or not. From the analysis of weaker intramolecular interactions, as identified by NCI, it would also appear that density is preferentially removed from peripheral or long-distance contacts (e.g. H14•••H19, H8•••H12) in favour of contacts with shorter distances which are localised within a ring formed by the leading NH•••N interaction.

From a chemist perspective it would be of importance to understand parameters controlling relative stability of conformers. A first attempt might be made by comparing the strength of the leading and 'truly' chemical in nature intramolecular H-bond. In this regard, analysis of the

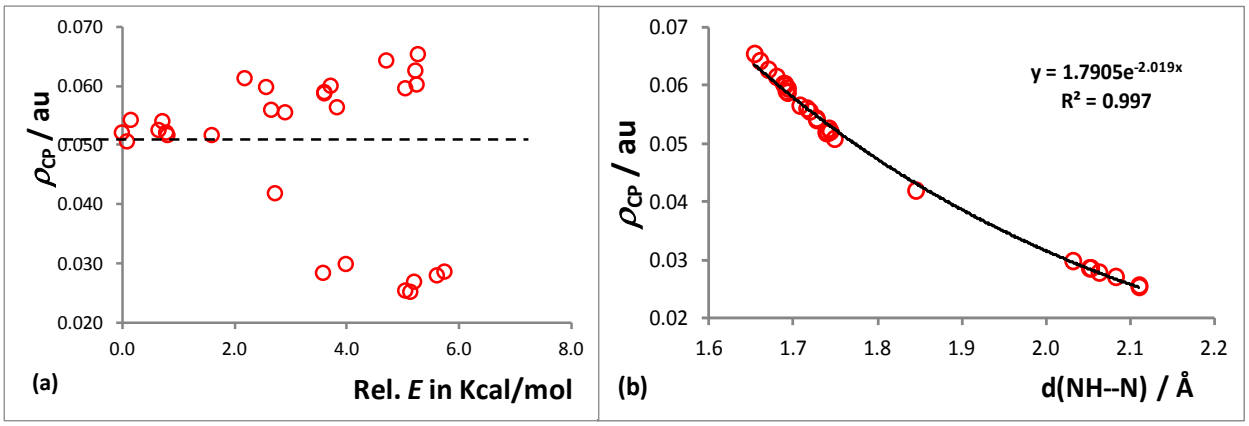


Figure 4.7. part (a) Relationship between electronic energy and ρCP ; part (b) Exponential decrease in ρ_{CP} with interatomic distance d(N,H), for leading NH•••N interactions in LECs of both $-\text{HL}_{\text{p}}$ and HL_{s} at MP2. Dashed line in part (a) is used to indicate ρCP of lowest energy

relationship between value of electron density at BCPs (ρCP) of leading NH•••N interactions for conformers of 2,2,2-tet shows, the strongest leading NH•••N bond is not necessarily found in the lowest energy conformer despite the fact that there is a good exponential relationship between the bond distance and ρCP of interactionsIt can be seen clearly from Figure 4.7. and Table B1 in Appendix B that the conformer which has the strongest NH•••N bond among the fifteen lowest in energy conformer of HL is ~5 kcal/mol higher in energy than the lowest energy conformer.

In addition several other NH•••N interactions exist in higher energy conformers which have much larger rho values than the leading NH•••N interaction of the lowest energy conformer (this is clearly indicated by the points above the dashed line in Figure 4.7a). This suggests that density at critical points of such AILs is a resultant effect of electron density distribution as determined by the physical forces i.e. Ehrenfest and Feynman responsible for the density distribution. Thus it is totally unrelated to electronic energy of such polyatomic molecules. This is not surprising as our previous attempt to use the sum of density contributions due to all intramolecular interaction identified by the presence of an AIL observed in a given conformer as a means of rationalizing its relative stability was also inconclusive.[45] We found out that there is no correlation whatsoever between the cumulative strength of AILs and conformer stability. Hence as pointed out previously, the assumption that each intramolecular interaction in a polyatomic molecule should provide a definite amount of stabilization and that these stabilizations are cumulative[46] should be taken cautiously especially in the case of polyatomic system such as 2,2,2-tet.

Furthermore, data in Table 4.1 reveals that for the two LECs of 2,2,2-tet considered here, all stability indices are in favour of the NH•••N interaction in the higher energy conformer for which we observe stronger by –14 kcal/mol interaction, much more significant covalent contribution, by about –5 kcal/mol, significantly larger density accumulation in the interatomic region (about –0.015 au) which resulted from a large inflow of density. The only reasonable explanation for the greater stability of L1 that we were able to come up with is the presence of three highly stabilizing NH•••N interactions in it whereas only two are observed in L2. However, if these were the only significant changes then L1 should be more stable, by ~–100 kcal/mol, than L2 but this is not the case. Hence, L1 must have paid some energy penalties (not reflected in Table 4.1) which largely reduced the decrease in the final energy of the L1 conformer. Clearly, any rigorous attempt to explain and quantify conformational preference is not an easy, if at all possible, task when polyatomic molecular structures are considered. In this regard, the NCI is very useful in identifying regions with increased density in the interatomic region from which additional and possibly significant interactions can be identified. However,

the appearance of blue regions in the NCI plots must be always accompanied by red ones (with depleted density) and interpretation of significance of the latter might be more difficult for chemists' purposes, in terms of stabilizing or unfavourable character of an interaction, as exemplified by, e.g., the highly stabilizing H29•••N26 interaction in L2. Likewise, deformation density analysis reveals that an inflow of electrons within an interatomic region would also be compensated for by an outflow from adjacent interatomic regions within the molecule. Therefore every stabilizing contributions in a given interatomic region might likely result in destabilization in another region of the molecule and vice-versa i.e. a molecule pays an energy penalty for every intramolecular interactions that occurs in the interatomic region between two atoms within it. As such, for molecules like ALPs which have numerous intramolecular interactions, the implication of these findings is that it is difficult if not impossible to sum up these competing energetic effects so as to be able to account for their conformational preferences.

4.4. Conclusions

Numerous NH•••N, CH•••HC and CH•••N intramolecular interactions in LECs of singly protonated 2,2,2-tet were investigated by exploring topology of electron density in the interatomic regions using standard protocols as implemented in QTAIM, IQA and NCI as well as density cross section along the eigenvector corresponding to the λ_2 eigenvalue of the Hessian matrix, starting from the geometric interaction point (the lowest density directly between two nuclei). All these techniques focus on properties of the resultant density distribution in a molecular system. To gain further insight, we have also analysed the deformation density in order to determine the process, inflow or outflow of density responsible for the formation of the various intramolecular interactions uncovered.

Data presented in Table 4.2 summarizes the results for 2,2,2-tet as well as other molecules studied in the full version of the published paper (i.e. water dimer arranged in different orientations, various protonated forms of cis- and trans- bpy, LEC and HEC of NTPA, see Figure B1-B3 in Appendix B for structures of these molecules). A full set of combined indices obtained for all interactions examined is shown, leading to the following final conclusions:

- an AIL can be observed for an interaction, regardless whether (i) it is highly attractive or repulsive as measured by the value and sign of $E_{\text{int}}^{\text{A,B}}$, (ii) it is a result of an inflow or outflow of density into the interatomic region.

- the sign of λ_2 cannot be used to predict that of $E_{\text{int}}^{\text{A,B}}$; the implication of this is that both highly repulsive and attractive, interactions might have locally depleted density and *vice versa*,

- locally accumulated density, with $\lambda_2 < 0$, does not necessarily originate from an inflow ($\Delta\rho(\boldsymbol{r}) > 0$) or outflow of density, and this equally applies to attractive and repulsive interactions either with or without an AIL.

Other important findings from a chemist's perspective are as follows:

- the first three interactions in Table 4.2 and their indices i.e. the presence of AIL, $E_{\text{int}}^{\text{A,B}} << 0$ dominated by the $V_{\text{cl}}^{\text{A,B}}$ term, $\lambda_2 < 0$ and $\Delta\rho(\boldsymbol{r}) > 0$ can be used to designate classical H-bonds (either inter- or intramolecular). Two other interactions, CH•••HC and CH•••HN have exactly the same set of indices (but their interaction energy is dominated by the $V_{\text{XC}}^{\text{A,B}}$) even though classically the former is classified as steric hindrance and the latter as another kind of a H-bond.

- the last two interactions in Table 4.2 are typical examples of classical non-bonded and repulsive contacts. As expected, they are characterized by the absence of AIL, $E_{\text{int}}^{\text{A,B}} >> 0$ dominated by the $V_{\text{cl}}^{\text{A,B}}$ term, locally depleted density ($\lambda_2 < 0$) and an outflow of density from the interatomic region ($\Delta\rho(\boldsymbol{r}) > 0$). It is important to note however, that surprisingly, there are also attractive interactions ($E_{\text{int}}^{\text{A,B}} < 0$) without AILs, for which also locally depleted density and outflow of density from the interatomic region is observed, and one of them, NH•••N in 2,2,2-tet, would easily be interpreted as an intramolecular H-bond.

- the O•••O interaction in *d*4 is highly repulsive and would be classified by any chemist as highly destabilizing a molecular system but, at the same time, is characterized by three identical topological features as found for classical H-bonds, namely (i) the presence of an AIL, locally increased density in and an inflow into the interatomic region.

- the character of an interaction can change radically in response to a change in its molecular environment as exemplified by CH•••HC for which the set of topological descriptors varies within the same molecule from that observed in the case of classical H-bonds to the set characterizing a destabilizing a molecule interaction, although the interaction energy between H-atoms involved is always negative.

- none of the indices (QTAIM, NCI, IQA, etc…), either separately or combined, can be used to predict the (de)stabilizing nature of an interaction except two limiting cases, the first and last interaction shown in Table 4.2.

Table 4.2. Comparative analysis of all interactions investigated in published work.[a]

Structure	Interaction	Dominant term	λ_2	$\Delta\rho(r)_{GIP}$
Attractive ($E_{int}^{A,B} < 0$) with AIL				
d1	O•••H	V_{cl}	neg	pos
L1 2,2,2-tet	NH•••N	V_{cl}	neg	pos
s-*cis* Hbpy	NH•••N	V_{cl}	neg	pos
s-*cis* bpy s-*cis* Hbpy s-*cis* H_2bpy L1 2,2,2-tet L2 2,2,2-tet	CH•••HC	V_{XC}	neg	pos
s-*trans* Hbpy HEC NTPA	CH•••HN	V_{XC}	neg	pos
L1 2,2,2-tet L2 2,2,2-tet	CH•••HC	V_{XC}	neg	neg
Repulsive ($E_{int}^{A,B} > 0$) with AIL				
d4	H•••H	V_{cl}	neg	pos
d4	O•••O	V_{cl}	neg	pos
s-*trans* H_2bpy	NH•••HN	V_{cl}	neg	pos
s-*trans* H_2bpy	CH•••HN	V_{cl}	neg	pos
d2, *d3*	O•••O	V_{cl}	neg	neg
d4	O•••O	V_{cl}	pos[b]	pos
Attractive ($E_{int}^{A,B} < 0$) without AIL				
d4	O•••H	V_{cl}	neg	pos
L2 2,2,2-tet	CH•••N	V_{XC}	neg	pos
s-*trans* bpy s-*trans* Hbpy	CH•••N	V_{cl}	neg	neg
d3 d4	O•••H	V_{cl}	neg	neg
LEC NTPA L2 2,2,2-tet	CH•••HC	V_{XC}	neg	neg
L1 2,2,2-tet L2 2,2,2-tet	CH•••HC	V_{XC}	neg	~0
L2 2,2,2-tet	CH•••HC	V_{XC}	pos	neg
HEC NTPA L1 2,2,2-tet L2 2,2,2-tet	CH•••HC	V_{XC}	pos	neg
L2 2,2,2-tet	NH•••N	V_{cl}	pos	neg
Repulsive ($E_{int}^{A,B} > 0$) without AIL				
d4	H•••H	V_{cl}	neg	pos
d4	O•••O	V_{cl}	pos	neg
s-*cis* bpy	N•••N	V_{cl}	pos	neg

[a] neg and pos stand for the negative and positive, respectively, signs of the λ_2 and $\Delta\rho(r)_{GIP}$ values; [b] this is at the GIP = RCP in this dimer where bifurcated AIL is observed

These observations demonstrate clearly that there is no correlation between (i) QTAIM-defined an atomic interaction line, AIL (presence or absence), (ii) IQA-defined interaction energy, $E_{\text{int}}^{\text{A,B}}$, and its components, classical $V_{\text{cl}}^{\text{A,B}}$ and exchange-correlation term $V_{\text{XC}}^{\text{A,B}}$, (iii) NCI-defined isosurfaces used to identify local regions of accumulated ($\lambda_2 < 0$) or depleted ($\lambda_2 > 0$) density relative to immediate environment, and (iv) deformation density for which $\Delta\rho(\boldsymbol{r}) > 0$ indicates an inflow and $\Delta\rho(\boldsymbol{r}) < 0$ indicates an outflow of density on the interaction formation.

The interpretation that the signs of λ_2 or $\Delta\rho(\boldsymbol{r})$ can be used as indications of "stabilizing", "attractive" or even "bonding" rests on the concept that an increase in density in the bonding region of an interaction is an indication of a bonding mechanism. In this work, we have applied two different techniques to measure an increase in density in the bonding region: the sign of λ_2 (as it is used in NCI and the interpretation of an AIL) and the sign of $\Delta\rho(\boldsymbol{r})$. The former indicates increased density relative to the local environment of an interaction, whereas the latter indicates increased density relative to non-interacting fragment states. It is important to note that the combination of the two methods gives much greater insight into the electron distribution of inter- and intramolecular interactions; particularly for all of the CH•••HC interactions investigated in this work. Even though the electron density distributions of these interactions show a wide range of different indices, we note that, in cases where a concentration of density or even an AIL is seen, a large outflow of density is observed between the neighbouring C-atoms. It appears that formation of an AIL does not have to be an output of the inflow of density into the interatomic region, as traced by the deformation density, but might be also an 'artefact' of density depletion between neighbouring atoms.

Clearly, one must exercise caution when using any *local* theoretical index (i.e. the value of the electron density or deformation density at a single coordinate), because the description of any interaction is highly influenced by its local environment. This is particularly true for congested systems such as 2,2,2-tet, with many intramolecular interactions present in the same space. It is obvious that to fully uncover the chemical character of an interaction it would be necessary and informative to include additional physical properties and expand on methodologies used thus far. Consequently, an attempt to explain the reason for conformational preference in molecules such as aliphatic linear polyamines might just remain an impossible task for now.

4.5. References

[1] P.L.A. Popelier, Quantum chemical topology: on bonds and potentials, *in Intermolecular Forces and Clusters I*, **2005**, Springer-Verlag, Berlin.

[2] N. Sukumar, C.M. Breneman, QTAIM in *drug discovery and protein modelling, in The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*, Wiley-VCH Verlag GmbH & Co. KGaA, **2007**, Weinheim.

[3] M. Vafaeezadeh, Z.B. Dizicheh, M.M. Hashemi, *Catal. Commun.*, **2013**, *41*, 96–100.

[4] C.D.M. Churchill, D.M. Cassandra, L.R. Rutledge, S.D. Wetmore, *Phys. Chem. Chem. Phys.*, **2010**, *12*, 14515–14526.

[5] S.M. LaPointe , S. Farrag , H.J. Bohórquez, R.J. Boyd, QTAIM study of an α-helix hydrogen bond network, *J. Phys. Chem. B*, **2009**, *113*, 10957–10964

[6] M. Mandado, A.M. Graña, R.A. Mosquera, *Phys. Chem. Chem. Phys.*, **2004**, *6*, 4391–4396.

[7] C.L. Firme, O.A.C. Antunes, P.M. Esteves, *Chem. Phys. Lett.*, **2009**, *468*,129–133.

[8] U. Koch, P.L.A. Popelier, *J. Phys. Chem.*, **1995**, *99*, 9747–9754.

[9] A.H. Pakiari, K. Eskandari, *J. Mol. Struct: THEOCHEM*, **2007**, *806*, 1–7.

[10] R. F. W Bader, *J. Phys. Chem. A*, **2009**, *113*, 10391–10396.

[11] L.J. Farrugia, C. Evans, M. Tegel, *J. Phys. Chem. A*, **2006**, *110*, 7952–7961.

[12] J. Cioslowski, S.T. Mixon, *Can. J. Chem.*, **1992**, *70* , 443-449.

[13] J. Cioslowski, S.T. Mixon., *J. Am. Chem. Soc.*, **1992**, *114*, 4382–4387.

[14] H.A. Jimenez-Vazquez, J. Tamariz, R.J. Cross, *J. Phys. Chem. A*, **2001**, *105*, 1315–1319.

[15] D. Moran, F. Stahl, E.D. Jemmis, H.F. Schaefer III, P.v.R. Schleyer, *J. Phys. Chem. A*, **2002**, *106*, 5144–5154.

[16] K. Kobayashi, S. Nagase, *Chem. Phys. Lett.*, **2002**, *362*, 373–379.

[17] Haaland, A.; Shorokhov, D. J.; Tverdova, N. V., *Chem. Eur. J.*, **2004**, *10*, 4416–4421.

[18] Jablonski, M., *J. Phys. Chem. A.*, **2012**, *116*, 3753–3764.

[19] Cerpa, E.; Krapp, A.; Flores-Moreno, R.; Donald, K. J.; Merino, G., *Chem. Eur. J.*, **2009**, *15*, 1985–1990.

[20] A.A. Popov, L. Dunsch, *Chem.–Eur. J.*, **2009**, *15*, 9707–9729.

[21] A.A. Popov, S.M. Avdoshenko, A.M. Pendás, L. Dunsch, *Chem. Commun.*, **2012**, *48*, 8031–8050.

[22] P. Demyanov, P. Polestshuk, *Chem. Eur. J.*, **2012**, *18*, 4982–4993.

[23] J. Poater, M. Solà, F.M. Bickelhaupt, *Chem.–Eur. J.*, **2006**, *12*, 2902–2905.

[24] Johnson, E. R.; Keinan, S.; Mori-Sánchez, P.; Contreras-García, J.; Cohen, A. J.; Yang, W., *J. Am. Chem. Soc.*, **2010**, *132*, 6498–6506.

[25] N. Gillet, R. Chaudret, J. Contreras-García, W. Yang, B. Silvi, J.P.J. Piquemal, *Chem. Theory Comput.*, **2012**, *8*, 3993−3997.

[26] J. Contreras-García, E.R. Johnson, S. Keinan, R. Chaudret, J.P. Piquemal, D. Beratan, W.J. Yang, *Chem. Theory Comput.*, **2011**, *7*, 625–632.

[27] R. Chaudret, B. de Courcy, J. Contreras-García, E. Gloaguen, A. Zehnacker-Rentien, M. Mons, J.P. Piquemal, *Phys. Chem. Chem. Phys.*, **2014**, *16*, 9876–9891.

[28] A. Armstrong, R.A. Boto, P. Dingwall, J. Contreras-García M.J. Harvey, N.J. Masona, H.S. Rzepa, *Chem. Sci.*, **2014**, *5*, 2057–2071.

[29] A. Otero-de-la-roza, R. Johnson, J. Contreras-garcía, *Phys. Chem. Chem. Phys.*, **2012**, *14*, 12165–12172.

[30] Bader, R. F. W., *J. Phys. Chem. A*, **1998**, *102*, 7314–7323.

[31] Bader, R. W. F. *Atoms in Molecules: A Quantum Theory*, **1990**, Oxford University Press: Oxford, U.K.

[32] J.C. Slater, *Phys. Rev.*,**1931**, *37*, 481–489.

[33] J.C. Slater, *Quantum theory of molecules and solids. Vol. 1.*, **1963**, New York: McGraw-Hill.

[34] R.P. Feynman, *Phys. Rev.*, **1939**, *56*, 340–343.

[35] W.L. Clinton, *J. Chem. Phys.*, **1960**, *33*, 1603–1606.

[36] J. Hernández-Trujillo, F. Cortés-Guzmán, D. Fang, R.F.W. Bader, *Faraday Disc.*, **2007**, *135*, 79–95.

[37] E. Agostinelli, M. P. M. Marques, R. Calheiros, F. P. S. C. Gil, G. Tempera, N. Viceconte, V. Battaglia, S. Grancara, A. Toninello, *Amino Acids.*, **2010**, *38*, 393–403.

[38] U. Bachrach, *Plant Phys Biochem.*, **2010**, *48*, 490–495.

[39] L. A. E. Batista de Carvalho, M. P. M. Marques, J. Tomkinson, J, *Phys Chem A*., **2006**, *110*, 12947–12954.

[40] G.J.S. Cooper. *Drugs*, **2011,** *71*, 1281–1320.

[41] I. Cukrowski, J. H. de Lange, A. S. Adeyinka, P Mangondo, *Comput Theoret Chem*., **2015**, *1053*, 60–76.

[42] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al., *Gaussian 09*, revision C.01; **2012**, Gaussian, Inc.: Wallingford, CT.

[43] Keith, T. A. AIMALL, version 13.05.06, **2013**, TK Gristmill Software, Overland Parks KS, USA, (aim.tkgristmill.com).

[44] Humphrey, W.; Dalke, A.; Schulten, K. *J. Molec. Graphics* **1996**, *14*, 33–38.

[45] A.S. Adeyinka, I. Cukrowski *J. Mol Model*., **2015**, *21:162*, 1–18.

[46] S. Gronert, R.A. J. O'Hair, *J Am Chem Soc*., **1995**, *117*, 2071–2081.

Chapter 5

Competition Reaction-Based Prediction of Polyamines' Stepwise Protonation Constants: a Case Study Involving 1,4,7,10-tetrazadecane (2,2,2-tet)

This chapter is essentially the published paper in *Theor. Chem. Acc.* **2016**, *135:139*, 1–17.

Summary

Theoretical prediction of four stepwise protonation constants of 1,4,7,10-tetraazadecane (2,2,2-tet) in correct order and with the smallest (largest) deviation of about 0.1 (–0.8) log unit from experimental values was achieved by an explicit application of a competition reaction (CRn) methodology in discrete-continuum solvation model involving four explicit water molecules. This methodology performs best when (i) tested ($L^{(1)}$) and reference ($L^{(2)}$) molecules are structurally similar, (ii) lowest energy conformers (LECs, selected from all possible tautomers) are used and (iii) a CRn, which assures a balanced charge distribution between reactants and products, $H_{n-1}L^{(1)} + H_nL^{(2)} = H_nL^{(1)} + H_{n-1}L^{(2)}$, is implemented. A 5-step *EEBGB*-protocol was developed to effectively and in shortest time possible select LECs (*E*, *B* and *G* stands for electronic-energy-, Boltzmann-distribution- and Gibbs-free-energy-based stepwise selection of conformers). The *EEBGB*-protocol (i) reduced (by 94%) the number of conformers subjected to the frequency calculations (to obtain *G*-values) from 420 MM-selected to 25 used to compute four protonation constants and (ii) is of general-purpose as it is applicable to any flexible and poly-charged molecules. Moreover, in search for LECs, a rapid pre-screening protocol was developed and tested; it was found efficient for the purpose of this study. Additional research protocols, aimed at even better prediction of protonation constants, are also suggested.

5.1. Introduction

Aliphatic polyamines (APs) are well known chelating ligands and extensive studies of their chemical properties have been carried out.[1–3] They are ubiquitous in cells and some of the biogenic ones can reduce proliferation of cells making them suitable drug candidates investigated by medicinal chemists for various therapeutic purposes.[2] As an example, their strong chelating ability has been utilized for the preparation of metal complexes which have been tested for anticancer properties.[3] Specifically, 1,4,7,10-tetraazadecane (2,2,2-tet) or *trien* is known to be a copper chelator used for the treatment of Wilson disease and is a possible drug candidate to prevent diabetic heart failure.[4] The biological activity of polyamines depends on their protonation state; hence, their protonation behaviour is of immense interest to both experimental and theoretical chemists.[5–7]

Proton transfer is one of the most important processes in chemical and biochemical systems.[8–13] Consequently, the ability of a molecule to accept or donate a proton is crucial and fundamental to our understanding of the pathways/mechanisms for several important reactions in living systems.[9,12] Several experimental techniques, such as mass spectrometry and ion-cyclotron resonance techniques in the gas phase as well as UV-visible spectroscopy, potentiometry, and NMR titration procedures in the solvent phase, have been used to obtain protonation constants.[8,10,12] Using most accurate experimental technique, glass electrode potentiometry, it is possible to obtain protonation constants with typical uncertainty on the second decimal place of the log unit. However, several experimental techniques (*e.g.*, ^{13}C NMR titration) are only capable of giving results to within a fraction of a log unit, in best cases with uncertainty on the first decimal place. In spite of the fact that experimental results for thousands of molecules are available (*e.g.*, those compiled by Martell and Smith[14] or IUPAC[15]), generating theoretical predictions is still of interest because (i) this would allow assessing biological activity of molecules yet to be synthesized, (ii) many biomolecules might be difficult to investigate due to solubility and stability issues and (iii) valuable insights one might gain from theoretical/computational modelling.[12,13]

Many papers have focused on theoretical prediction of protonation constants for diverse biologically important compounds such as amines, amides, carboxylic acids, bicarbonates and proteins, amongst others.[11,16–34] Most of these studies made use of various thermodynamic cycles (TCs) and mainly focused on neutral or singly charged molecules. These TCs involve a two-stage process; (i) full gas-phase energy minimization of components involved in the protonation reaction, followed by (ii) a single point calculation in solvent (water) from which

$\Delta G_{(aq)}$ is computed and used to calculate protonation constants at room temperature.[12,13,16] Typically, the TC-based methods are able to give protonation constants within ±2 log units of experimental value for neutral or singly charged molecules but this accuracy depreciates as the charge on the studied molecule increases.[9] There are several sources of errors which contribute to an inherent uncertainty of results obtained from TCs, such as (i) uncertainty in the solvation free energy of a proton, (ii) inaccuracy in evaluating the solvation free energy of ionic species by continuum solvation models (this might range between 0.5–1 kcal/mol for neutral molecules and 3–4 kcal/mol for ions)[8] and (iii) errors inherent in state-of-the-art quantum chemistry methods (about 1 kcal/mol)[8] used to compute free energies in the gas phase. In order to minimize errors, several modifications[10,12,13,35] have been developed, a prominent example of which is the incorporation of an isodesmic reaction within a TC. In other cases, results obtained from TCs have been empirically corrected using parameters obtained from linear regression analysis of experimentally measured protonation constants.[16] In some instances, these modifications have made the prediction of protonation constants to within 1–2 log units possible, though this is still far from what is obtainable experimentally.

It has been demonstrated that making use of a competition reaction (CRn) based methodology may result in more accurate predictions of protonation constants.[8,12,13,35–38] For example, it was used recently to predict four protonation constants, as $\log K_{\mathrm{H}}^{(n)}$ where *n* stands for the order of the stepwise protonation reaction, of highly negatively charged molecules, such as NTPA and NTA[12,13] in simulated solvent with DFT, a relatively low and cheap level among electronic structure methods. There are many factors which contributed to high quality computationally predicted four $\log K_{\mathrm{H}}^{(n)}$ values, among them (i) an inherent property of error cancellation which is typical for CRn (or isodesmic reaction in general), (ii) structural similarity of and charge distribution on the studied and reference molecules and (iii) simplicity of the continuum solvation model (CSM) used[12,13], which performs well when used at the level of theory for which it was parameterized.[9] Many existing computational methodologies can be seen as well established (or routine) now in the field and they are described in details in recent reviews by Ho and Coote[9,10] and Casasnovas et al.[8]

In spite of unquestionable successes in this area, it is somewhat surprising that there is still little[39] (in case of diamines) or no information about the theoretical prediction of protonation constants of polyamines, such as, *e.g.*, tetramines. This observation might be attributed to specific properties of polyamines, particularly with more than two N-atoms:

(i) They are extremely flexible, resulting in almost an infinite number of possible conformers; this makes discovery of required for computing protonation constants low energy conformers (LECs) a herculean task due to amount of computational resources required (energy optimisation and frequency calculations).

(ii) The difference between their first two stepwise protonation constants[14] is a fraction of a log unit in most cases and is well below typical 'resolution' reported from computational work.

(iii) Just considering the title compound, 2,2,2-tet, the HL^{+} and H_2L^{2+} protonated forms have two tautomers and their preferences in the gas and solvent phases differ. Hence, this rules out the use of commonly utilized thermodynamic cycles.

We reported recently the first extensive conformational analysis of protonated polyamines using *trien* as a case study[40] and showed that the developed protocol was able to identify representative sets of LECs. These were used to predict %-fraction of each tautomeric form of the singly and doubly protonated *trien* which were in good agreement with experimental data obtained from the ^{13}C NMR spectrometry. Usefulness of that conformational search protocol and the CRn methodology in predicting four protonation constants of highly negatively charged molecules[12,13] have motivated us to undertake this investigation where four stepwise protonation constants of 2,2,2-tet will be predicted using 1,5,8,12-tetraazadodecane (3,2,3-tet) as a reference molecule. As far as we could establish, this is the first attempt to combine the two methodologies we referred to above.[12,40] Hence, we have selected tetramines of similar structures for which experimental values are known as this allowed us verification of theoretically predicted $\log K_{\mathrm{H}}^{(\mathrm{n})}$ values and different protocols developed in this work.

5.2. Computational Methods

It has been emphasized[41–44] that an appropriate description of the solvation environment is critical for best theoretical prediction of protonation constants. Because continuum solvation models, CSMs, are known to suffer from errors due to their omission of discrete hydrogen bonding and inadequate treatment of short-range electrostatics[41,45,46], the so-called discrete-continuum solvation model (DCSM) was also used in this work. DCSM involves placing explicit solvent molecules around the solute to simulate the first solvation shell. The resultant 'supermolecule'[41,45] is immersed in a cavity that is surrounded by a dielectric continuum to model bulk solvent effects. Unfortunately, there is no generally applicable theoretical method to determine the appropriate number of explicit water molecules needed to represent the first solvation shell. Hence, four water molecules was used in order (i) to facilitate the formation of

maximum number of possible hydrogen bonding between water molecules and the solute with two pairs of –NH and $–NH_2$ groups and (ii) to keep the computational resources needed for this work affordable.

Conformational search was performed in Spartan[47] to generate a large set of representative conformers of the various protonated (H_nL^{n+}) forms of 2,2,2-tet and 3,2,3-tet using molecular mechanics with the MMFF force field. Furthermore, to account for the aqueous solvent effects, the Monte Carlo algorithm in combination with MMFF(aq) option, as implemented in Spartan, were utilised. It was necessary to employ MMFF(aq) because the sets of LECs discovered in the gas phase (using MMFF) were significantly different. This was done by a systematic variation of the torsional angle of each rotatable bond as described previously[40] with slight modifications implemented in the case of 3,2,3-tet (see PART 1 of Appendix C for a full description of the conformational search procedure used). We have also performed conformational search on the same ligands with explicitly added four water molecules which were placed (i) randomly in relation to their orientation toward a backbone structure of a ligand, but (ii) quite evenly along a molecule; for illustration, free ligands with water molecules are shown in Figure 5.1. Linear structures of all possible forms the 2,2,2,-tet ($L^{(1)}$) and 3,2,3-tet ($L^{(2)}$) ligands, shown in PART 2 of Appendix C as Figures C1-C2, were used as inputs for the MM-based conformational search.

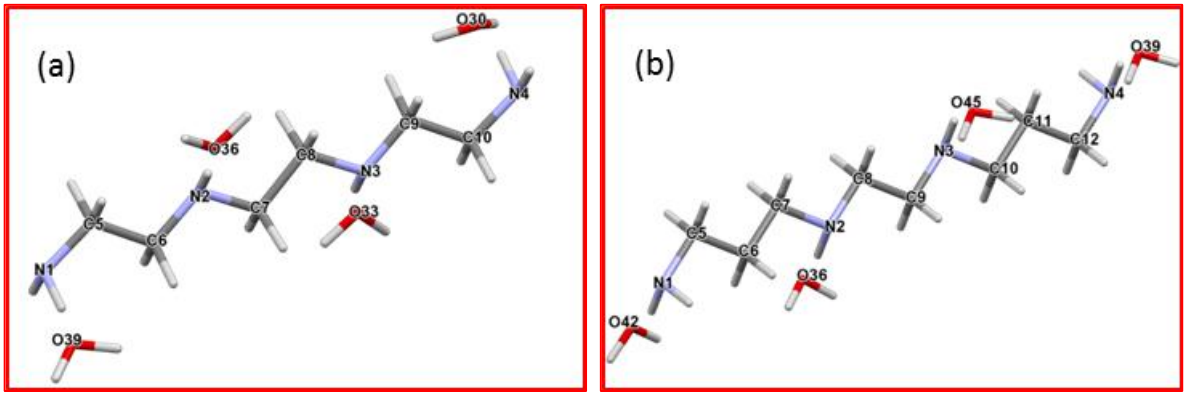


Figure 5.1. Capped-stick representation of free ligand linear input structures with explicit water molecules used for conformational search by MMFF(aq): 2,2,2-tet in part a; 3,2,3-tet in part b.

A maximum of thirty unique and lowest in energy conformers was retained after each conformational search; they were energy optimised in Gaussian 09, revision D01[48], at the RB3LYP/6-311++G(d,p) level of theory in conjunction with default settings of the Polarizable Continuum solvation Model (PCM) using water as solvent (ε = 78.3553). Vibrational frequencies were computed using the rigid rotor harmonic (RRHO) approximation, as implemented in Gaussian 09, in order to (i) obtain Gibbs free energies needed for computing

protonation constants and (ii) to verify that all minimum energy structures reported in this study were true stationary points on the potential energy surface. Furthermore, a tight gradient criterion was used along with an ultrafine integration grid to ensure acceptable convergence of frequencies computed.[49]

5.3. Competition Reaction based Protocol

A general concept of the CRn methodology was described previously when it was used to determine protonation constants of polycarboxylic acids[12,13] as well as formation constants.[50] As such, a competition for a proton between a polyamine (ligand) under investigation $L^{(1)}$ and a structurally analogous reference molecule $L^{(2)}$ is explored here to compute the free energy change $\Delta G_{\mathrm{CRn}}^{(n)}$ needed to calculate an *n*th protonation constant, as $\log K_{\mathrm{H}}^{(n)}$, in aqueous solution. An example of a CRn reaction for the first protonation reaction can be written as (for simplicity, the (aq) notation and charges were omitted throughout),

$$L^{(1)} + HL^{(2)} = HL^{(1)} + L^{(2)}. \qquad \Delta G_{\mathrm{CRn}}^{(1)} \qquad (5.1)$$

However, because this work is concerned with tetramines with several possible protonation sites, additional and important aspects had to be considered. As depicted in Scheme 5.1, which involves all possible protonation sites of 2,2,2-tet (here $L^{(1)}$) and 3,2,3-tet (here $L^{(2)}$) when they are singly protonated, one must consider a number of possible reactions which indeed might take place in a real competition experiment in a solution. To reflect this fact, a general form of competition reaction 1 can be expressed in the following forms

$$L^{(1)} + HL_{p}^{(2)} = HL_{p}^{(1)} + L^{(2)} \qquad \Delta G_{HL_{p}^{(1)},HL_{p}^{(2)}}^{(1)} \qquad (5.2)$$

$$L^{(1)} + HL_{p}^{(2)} = HL_{s}^{(1)} + L^{(2)} \qquad \Delta G_{HL_{s}^{(1)},HL_{p}^{(2)}}^{(1)} \qquad (5.3)$$

$$L^{(1)} + HL_{s}^{(2)} = HL_{p}^{(1)} + L^{(2)} \qquad \Delta G_{HL_{p}^{(1)},HL_{s}^{(2)}}^{(1)} \qquad (5.4)$$

$$L^{(1)} + HL_{s}^{(2)} = HL_{s}^{(1)} + L^{(2)} \qquad \Delta G_{HL_{s}^{(1)},HL_{s}^{(2)}}^{(1)} \qquad (5.5)$$

$$L^{(1)} + HL_{p}^{(2)} + HL_{s}^{(2)} = HL_{p}^{(1)} + L^{(2)} \qquad \Delta G_{HL_{p}^{(1)},HL_{p+s}^{(2)}}^{(1)} \qquad (5.6)$$

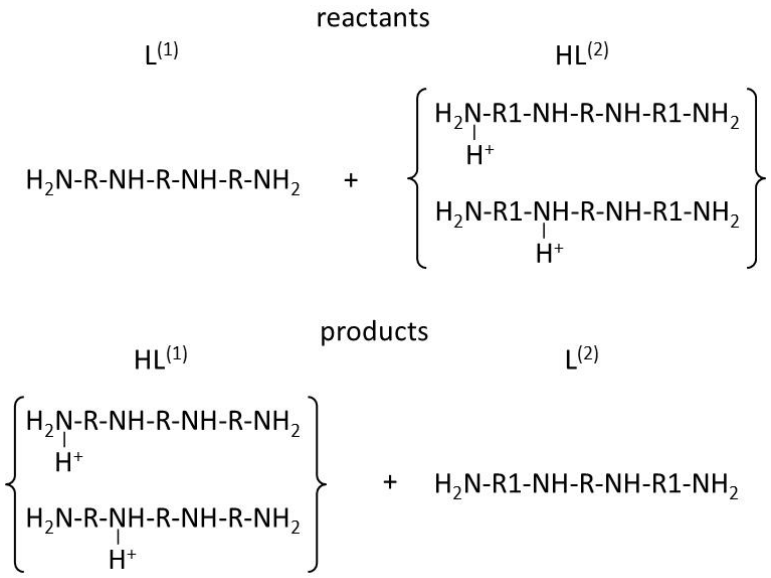
$$L^{(1)} + HL_{p}^{(2)} + HL_{s}^{(2)} = HL_{s}^{(1)} + L^{(2)} \qquad \Delta G_{HL_{s}^{(1)},HL_{p+s}^{(2)}}^{(1)} \qquad (5.7)$$

$$L^{(1)} + HL_{p}^{(2)} = HL_{p}^{(1)} + HL_{s}^{(1)} + L^{(2)} \qquad \Delta G_{HL_{p+s}^{(1)},HL_{p}^{(2)}}^{(1)} \qquad (5.8)$$

$$L^{(1)} + HL_{s}^{(2)} = HL_{p}^{(1)} + HL_{s}^{(1)} + L^{(2)} \qquad \Delta G_{HL_{p+s}^{(1)},HL_{s}^{(2)}}^{(1)} \qquad (5.9)$$

$$L^{(1)} + HL_p^{(2)} + HL_s^{(2)} = HL_p^{(1)} + HL_s^{(1)} + L^{(2)} \qquad \Delta G^{(1)}_{HL^{(1)}_{p+s},HL^{(2)}_{p+s}} \qquad (5.10)$$

where subscripts 's' and 'p' denote a primary and secondary N-atom of a ligand being protonated and the sequence of subscripts in the $\Delta G^{(1)}_{CRn}$ expressions shows tautomers of $HL^{(1)}$ and $HL^{(2)}$ involved; their sequence follows the way $\Delta G^{(1)}_{CRn}$ is calculated ($G_{products} - G_{reactants}$) and, *e.g.*, for $\Delta G^{(1)}_{CRn} = \Delta G^{(1)}_{HL^{(1)}_{p+s},HL^{(2)}_{p}}$ it means that two tautomers of a singly protonated polyamine under investigation, $HL^{(1)}_p$ and $HL^{(1)}_s$, and by default the free reference ligand $L^{(2)}$ were formed as a result of reaction between $L^{(1)}$ (by default) and the tautomer of the singly protonated reference polyamine with the primary N-atom being protonated, hence $HL^{(2)}_p$.



Scheme 5.1. Possible tautomers of 2,2,2-tet ($L^{(1)}$) and 3,2,3-tet ($L^{(2)}$) (R = $–C_2H_4–$; R1 = $–C_3H_6–$) which were considered in the competition reaction based protocol to compute $\log K^{(1)}_{H}$ of $L^{(1)}$.

It is obvious that each possible CRn generates different $\Delta G^{(1)}_{CRn}$ value because inequality $G(HL_p) \neq G(HL_s)$ holds for any polyamine. Furthermore, only in few instances one is able to predict most likely and the only one possible protonation site in polyamines with certainty. Hence, one must, in principle, use most general expression for a CRn,

$$L^{(1)} + HL_p^{(2)} + HL_s^{(2)} = HL_p^{(1)} + HL_s^{(1)} + L^{(2)} \qquad \Delta G^{(1)}_{CRn} \qquad (5.11)$$

where subscripts 's' and 'p' denote a primary and secondary N-atom of a ligand being protonated. This is an additional complication because to calculate $G_{products}$ and $G_{reactants}$ an exact %-fraction of the two possible tautomers, HL_p and HL_s, for both polyamines must be known. As

a matter of fact, this is still not a sufficient requirement to calculate $\Delta G^{(1)}_{\mathrm{CRn}}$; note that to compute, *e.g.*, $G_{\mathrm{reactants}}$ of reaction 5.2, all possible conformers (there are thousands of them when linear aliphatic polyamines are considered), or at least the LECs of L, $\mathrm{HL_p}$ and $\mathrm{HL_s}$ for both polyamines must be considered, hence

$$G_{\mathrm{reactants}} = \sum_{k=1}^{\mathrm{LEC}} w_k G_k(\mathrm{L}^{(1)}) + \sum_{m=1}^{\mathrm{LEC}} w_m G_m(\mathrm{HL}^{(2)}_{\mathrm{p}}) + \sum_{n=1}^{\mathrm{LEC}} w_n G_n(\mathrm{HL}^{(2)}_{\mathrm{s}}) \qquad (5.12)$$

or, in more general form when *t* tautomers are possible, one can write

$$G_{\mathrm{reactants}} = \sum_{k=1}^{\mathrm{LEC}} w_k G_k(\mathrm{L}^{(1)}) + \sum_{l=1}^{t} \sum_{m=1}^{\mathrm{LEC}} w_m G_m(\mathrm{HL}^{(2)}_{\mathrm{t}}) \qquad (5.13)$$

where *w* stands for the population fraction obtained from the Boltzmann distribution calculated for selected LECs of $\mathrm{L}^{(1)}$, $\mathrm{HL}^{(2)}_{\mathrm{p}}$ and $\mathrm{HL}^{(2)}_{\mathrm{s}}$ (or in general *t* tautomers of $\mathrm{HL}^{(2)}_{\mathrm{t}}$). Note that *w* can be seen and must be used as the weight factor which assures proportional (to this structure contribution to the entire population) free energy contribution to the computed free energy change of the competition reaction. The same considerations equally apply to products of the reaction 1, hence one can write

$$G_{\mathrm{products}} = \sum_{x=1}^{\mathrm{LEC}} w_x G_x(\mathrm{L}^{(2)}) + \sum_{y=1}^{\mathrm{LEC}} w_y G_y(\mathrm{HL}^{(1)}_{\mathrm{p}}) + \sum_{z=1}^{\mathrm{LEC}} w_z G_z(\mathrm{HL}^{(1)}_{\mathrm{s}}) \qquad (5.14)$$

or, as a general expression,

$$G_{\mathrm{products}} = \sum_{x=1}^{\mathrm{LEC}} w_x G_x(\mathrm{L}^{(2)}) + \sum_{k=1}^{t} \sum_{y=1}^{\mathrm{LEC}} w_y G_y(\mathrm{HL}^{(1)}_{\mathrm{t}}) \qquad (5.15)$$

As shown previously[12], the free energy change for the competition reaction of the first protonation step can be obtained from a general expression

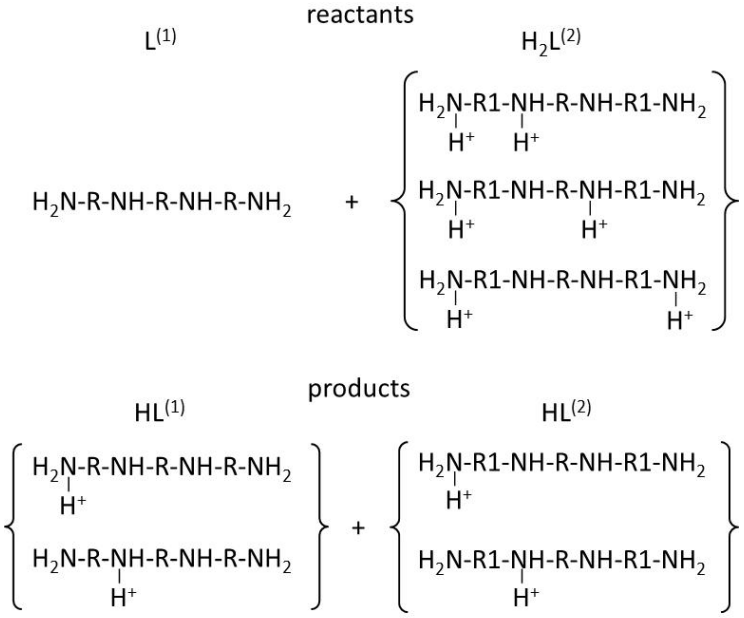
$$\Delta G^{(1)}_{\mathrm{CRn}} = G(\mathrm{HL}^{(1)}) + G(\mathrm{L}^{(2)}) - G(\mathrm{L}^{(1)}) - G(\mathrm{HL}^{(2)}) = \Delta G(\mathrm{L}^{(1)}) - \Delta G(\mathrm{L}^{(2)}) \qquad (5.16)$$

where $\Delta G(\mathrm{L})$ is calculated for a direct protonation reaction L + H = HL involving all tautomers. Our aim is to compute $\Delta G(\mathrm{L}^{(1)})$ which is needed to obtain the first protonation constant, as $\log K^{(1)}_{\mathrm{H}}$, of $\mathrm{L}^{(1)}$ from $\Delta G = -RT \ln K^{(1)}_{\mathrm{H}}$. Because protonation constants of $\mathrm{L}^{(2)}$ are known, the $\Delta G(\mathrm{L}^{(2)})$ term can be easily obtained (ΔG = –RTln*K*) and one is left, in general, with four *G* values for all reactants and products in reaction 1. However, when polyamines investigated here are considered, one must combine expressions 5.12 and 5.14 to compute $\Delta G^{(1)}_{\mathrm{CRn}}$,

$$\Delta G^{(1)}_{\mathrm{CRn}} = \sum_{x=1}^{\mathrm{LEC}} w_x G_x(\mathrm{L}^{(2)}) + \sum_{y=1}^{\mathrm{LEC}} w_y G_y(\mathrm{HL}^{(1)}_{\mathrm{p}}) + \sum_{z=1}^{\mathrm{LEC}} w_z G_z(\mathrm{HL}^{(1)}_{\mathrm{s}})$$

$$- \sum_{k=1}^{\mathrm{LEC}} w_k G_k(\mathrm{L}^{(1)}) + \sum_{m=1}^{\mathrm{LEC}} w_m G_m(\mathrm{HL}^{(2)}_{\mathrm{p}}) + \sum_{n=1}^{\mathrm{LEC}} w_n G_n(\mathrm{HL}^{(2)}_{\mathrm{s}}) . \qquad (5.17)$$

As it is seen from expression 5.17, computing $\Delta G^{(1)}_{\mathrm{CRn}}$ is a formidable and almost an impossible task when, at least, time and computational resources needed to achieve our main goal are considered. Because of that, we explored different options (they will be discussed in sections that follow) to simplify the protocol without compromising the quality of computed protonation constants.



Scheme 5.2. Possible tautomers of 2,2,2-tet and 3,2,3-tet (R = –C_2H_4–; R1 = –C_3H_6–) which were considered in the competition reaction based protocol when $L^{(1)}$ and $H_2L^{(2)}$ were employed to compute $\log K^{(1)}_{\mathrm{H}}$ of $L^{(1)}$.

The added advantage of using a competition reaction is that a significant cancellation of different errors inherent in solvation model and electronic structure method used to optimise molecules should take place. Also, one expects that structural similarity of reactants and products, as is the case here, should result in errors minimization (cancellation). To this effect and knowing that computational optimisation of structures with multiple charges is still a challenge when accuracy goes; one should also consider the selection of reference molecules in terms of resultant placement of charges among reactants and products. For instance, can one

obtain better theoretical estimate of, *e.g.*, (i) the first protonation constant of $L^{(1)}$ using $HL^{(2)}$ or rather $H_2L^{(2)}$ or (ii) the third protonation constants of $H_2L^{(1)}$ involving in the competition reaction $HL^{(2)}$, $H_2L^{(2)}$, $H_3L^{(2)}$, or rather $H_4L^{(2)}$? There is no easy way to predict this and, as an example, see Scheme 5.2 where (i) both products of the protonation competition reaction have exactly the same charge, (ii) the same number of tautomers for $L^{(1)}$ and $L^{(2)}$ is involved, but (iii) charges on reactants differ. This aspect has also been investigated in the present work in terms of accuracy in the computed protonation constants.

5.4. Results and discussion

5.4.1. Pre-optimisation protocol

Firstly, it is important to realize how enormous computational task this kind of study requires when all conformers were to be optimized with frequency calculations; retaining 30 MM-selected lowest energy conformers of each tautomer results in 210 structures for each ligand, $L^{(1)}$ and $L^{(2)}$. When structures with explicit water molecules are also considered, as is the case here, then the starting minimum number of conformers one must consider is 840. Because we wanted to develop a feasible protocol, we decided to seek alternative avenues. To this effect, we took advantage of having a large data bank from previous work[40] where hundreds of 2,2,2-tet conformers were fully optimized in Gaussian. A thorough examination of the optimization profiles revealed that there are common patterns when a relationship between electronic energy of a conformer after each optimization cycle and a step number was analysed: (i) a sharp decrease in *E* is observed in the first 3-5 steps in case of conformers which optimise within 10-15

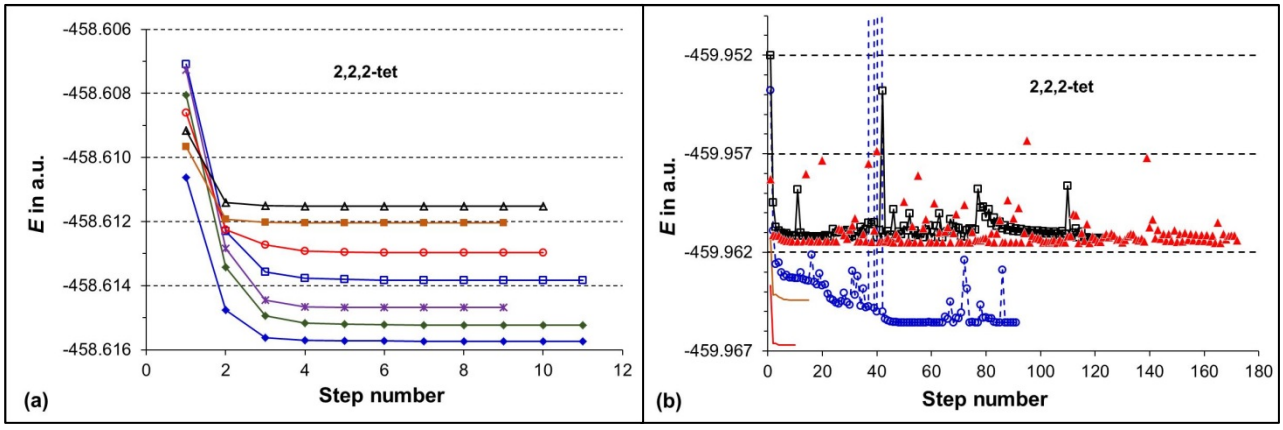
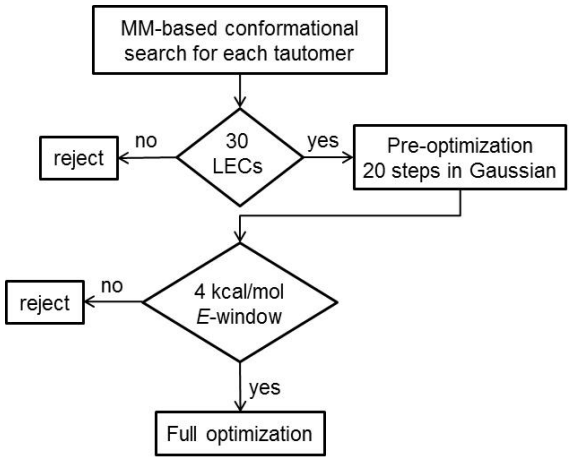


Figure 3.2. Examples of optimization profiles for selected conformers of 2,2,2-tet in CSM showing the change in the electronic energy E with the optimization step in cases of: part (a) - convergence reached within 15 optimization cycles, part (b) - convergence reached after large number of optimization cycles without (red triangles and black squares) and insignificant (blue circles) relative change of conformers' placement in their energy spectrum.

steps; in such a case the relative energies of conformers generally do not change (Figure 5.2a), (ii) a major decrease in *E* is observed within first 10-15 optimization steps for each conformer and (iii) in many instances a large number of steps was required to reach convergence with insignificant energy change after 20-50*th* step (Figure 5.2b). These observations provide a useful hint when one is interested only in the set of the lowest energy conformers needed to compute, *e.g.*, protonation constants. A thorough inspection of the optimisation profiles generated for all tautomers of protonated forms of 2,2,2-tet revealed that in order to predict 'safely' the set of lowest energy conformers needed for the purpose of this study it would be sufficient to implement a pre-optimization operation which involves terminating the optimization process after 20 steps. We decided to implement this finding, as shown in Scheme 5.3, in the optimization of 3,2,3-tet conformers in both solvation models.

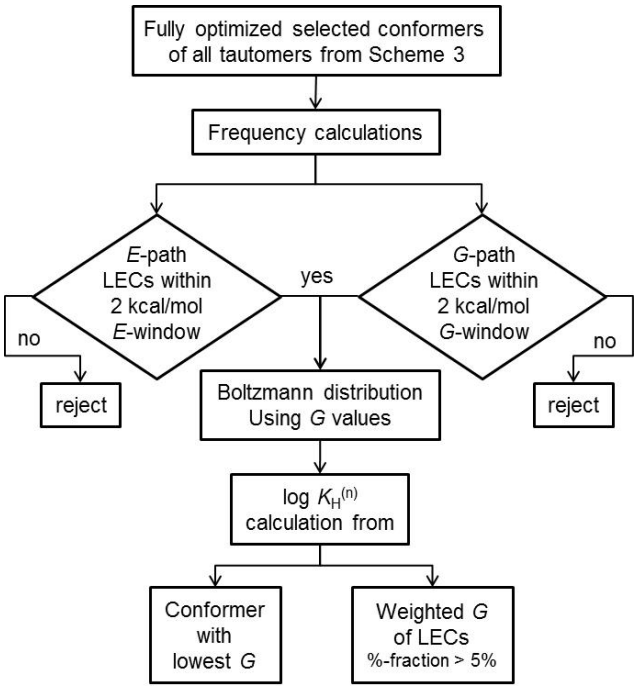


Scheme 5.3. Protocol used to select structures for full energy

5.4.2. General purpose protocol

Implementation of the pre-optimisation step allowed us to preliminarily reject about 60-70% of conformers. The remaining conformers had to be fully energy optimised and subjected to frequency calculations as *G* values are required to predict protonation constants. Knowing that frequency calculations are extremely time consuming and this is particularly true when explicit water molecules are included, we decided to explore additional two selection paths with a hope that maybe it would be possible to reduce the number of necessary conformers even further, hence reduction in computational time should result too. The general purpose protocol developed in this work is shown in Scheme 5.4. It incorporates a step-wise elimination of 'redundant' conformers and specific strategies tested in computing protonation constants.

Examples of 2,2,2-tet and 3,2,3-tet LECs for all protonated forms of each tautomer selected after full optimisation in the continuum and discrete-continuum solvation model are shown in PART 3 (as Figures C3–C12) and PART 4 (as Figures C13–C22), respectively, of Appendix C. The lowest energy HL, H_2L and H_3L conformers discovered for each ligand in DCSM are shown in Figure 5.3. To ensure easy identification and differentiation between conformers of the various protonated (H_nL^{n+}) forms of 2,2,2-tet and 3,2,3-tet, we have consistently labelled them as:

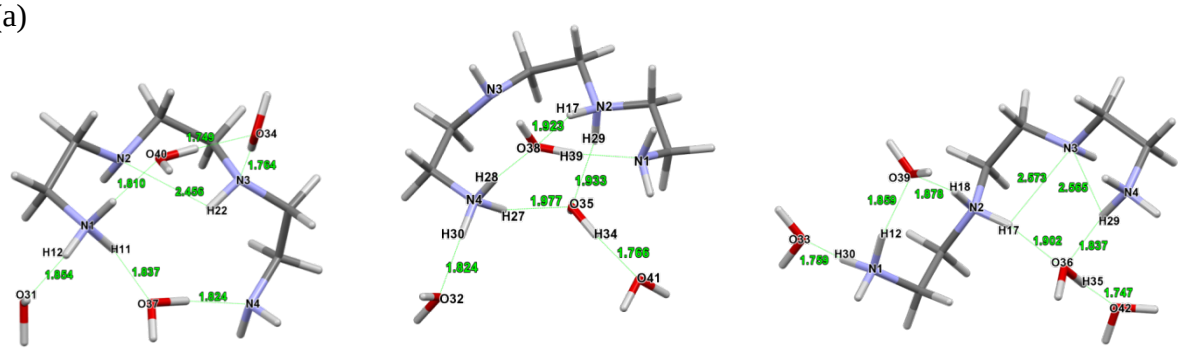


Scheme 5.4. General purpose approach used in testing different methodologies in search of time (cost) most-effective protocol for computational determination of protonation constants.

- C_Ln for structures of the free ligand (L),
- C_pn and C_sn for the primary (HL_p) and secondary (HL_s) forms of monoprotonated structures where the primary and secondary N-atoms are protonated, respectively,
- $C_{ps}n$ and $C_{pp}n$, for the diprotonated structures, where $C_{ps}n$ is used to denote conformers of the tautomeric form where one primary and one secondary nitrogen atoms are protonated (H_2L_{ps}) whereas $C_{pp}n$ is used for a structure in which the two terminal nitrogen atoms are protonated (H_2L_{pp}),
- $C_{tp}n$ for structures of the triply protonated form in which both primary nitrogen atoms and one secondary nitrogen atom are protonated; in the case of ligands studied here, there is only one stable tautomer according to physical charge separation requirements, and
- $C_{fp}n$ is used to denote structures of the fully protonated form.

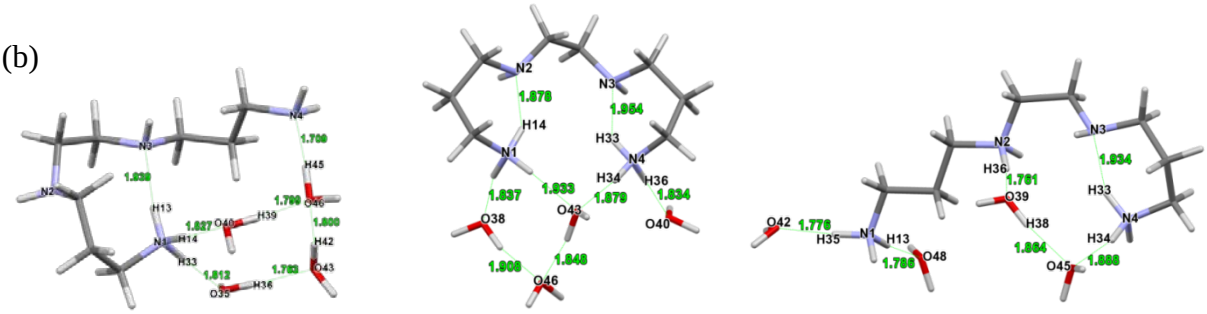
It is easy to establish, using Boltzmann distribution, that conformers with energies greater than 3 kcal/mol (relative to the lowest energy conformer when using either *E*- or *G*-path) contribute insignificantly to the total population, typically below 0.1 %-fraction. Furthermore, when the weighted energy of conformers was used to compute the overall *G* value of all selected conformers with %-fraction either above 1 or 5%, it became clear that incorporation of conformers characterized by 1 < %-fraction of the total population < 5 had no significant impact on the computed $\log K_{\mathrm{H}}^{(\mathrm{n})}$ values.

(a)



C_p01 $C_{ps}02$ $C_{tp}10$

(b)



C_p26 $C_{pp}18$ $C_{tp}01$

Figure 5.3. Lowest energy HL, H_2L and H_3L conformers with explicit water molecules of 2,2,2-tet in part (a) and 3,2,3-tet in part (b) found from the *E*-path shown in Scheme 4.4.

Analysis of Boltzmann distributions obtained for all protonated forms of both ligands (when applicable, the combined tautomers were used to generate a population of conformers, *e.g.*, HL_p plus HL_s) revealed that selecting conformers with %-fraction > 5% always resulted in the *E*- or *G*-window < 2 kcal/mol within which LECs were found. For illustration purposes, Table 5.1 shows five lowest in energy conformers (with explicit water molecules) selected by the electronic and Gibbs free energy based paths for 2,2,2-tet and 3,2,3-tet; relevant data for the implicit solvation model are included in Table 5.2. The *E*-path shown in Scheme 5.4 was implemented to test whether the selection of LECs within 2 kcal/mol *E*-window would retain conformers which, after frequency calculations, would give *G* values suitable for protonation constant calculations in terms of quality (accuracy) of computed values. This approach was

Table 5.1 Five lowest energy conformers of all H_nL with explicit water molecules, using either *E* or *G* values, for: part (a) - 2,2,2-tet and part (b) - 3,2,3-tet.[a]

(a)

	L			HL			H_2L			H_3L			H_4L	
Conf	ΔE	%	Conf	ΔE	%	Conf	ΔE	%	Conf	ΔE	%	Conf	ΔE	%
C_L03	0.00	58.4	C_p01	0.00	69.5	$C_{ps}02$	0.00	91.7	$C_{tp}10$	0.00	52.4	$C_{fp}01$	0.00	76.5
C_L01	0.66	19.1	C_s07	0.49	30.5	$C_{ps}03$	1.44	8.0	$C_{tp}02$	0.08	45.8	$C_{fp}02$	0.88	17.3
C_L05	1.00	10.9	–	–	–	$C_{ps}12$	4.13	0.1	$C_{tp}06$	2.40	0.9	$C_{fp}04$	2.09	2.2
C_L04	1.55	4.3	–	–	–	$C_{pp}09$	4.24	0.1	$C_{tp}01$	2.94	0.4	$C_{fp}06$	2.10	2.2
C_L06	1.59	4.0	–	–	–	$C_{pp}04$	4.37	0.1	$C_{tp}11$	3.14	0.3	$C_{fp}07$	2.24	1.7
Conf	ΔG	%	Conf	ΔG	%	Conf	ΔG	%	Conf	ΔG	%	Conf	ΔG	%
C_L09	0.00	39.7	C_p01	0.00	88.7	$C_{pp}08$	0.00	84.1	$C_{tp}02$	0.00	44.1	$C_{fp}01$	0.00	89.0
C_L03	0.03	37.9	C_s07	1.22	11.3	$C_{ps}02$	1.38	8.1	$C_{tp}06$	0.44	21.1	$C_{fp}02$	1.34	9.3
C_L01	0.48	17.7	–	–	–	$C_{pp}06$	1.65	5.1	$C_{tp}01$	0.60	16.1	$C_{fp}06$	2.56	1.2
C_L05	1.39	3.8	–	–	–	$C_{ps}03$	2.44	1.4	$C_{tp}10$	0.79	11.7	$C_{fp}04$	3.31	0.3
C_L06	2.36	0.7	–	–	–	$C_{pp}07$	2.52	1.2	$C_{tp}09$	1.78	2.2	$C_{fp}07$	3.61	0.2

(b)

	L			HL			H_2L			H_3L			H_4L	
Conf	ΔE	%	Conf	ΔE	%	Conf	ΔE	%	Conf	ΔE	%	Conf	ΔE	%
C_L21	0.00	96.3	C_p26	0.00	62.5	$C_{pp}18$	0.00	92.0	$C_{tp}01$	0.00	64.2	$C_{fp}01$	0.00	35.4
C_L 19	2.45	1.5	C_p24	0.38	33.1	$C_{ps}06$	2.05	2.9	$C_{tp}25$	0.57	24.3	$C_{fp}08$	0.59	13.01
C_L 01	2.54	1.3	C_p19	2.06	1.9	$C_{pp}19$	2.15	2.4	$C_{tp}02$	1.26	7.6	$C_{fp}09$	0.97	6.9
C_L 09	3.30	0.4	C_p12	2.61	0.8	$C_{ps}21$	2.64	1.1	$C_{tp}10$	2.33	1.2	$C_{fp}04$	1.03	6.2
C_L 11	3.68	0.2	C_p07	2.85	0.5	$C_{ps}09$	2.83	0.8	$C_{tp}15$	2.58	0.8	$C_{fp}25$	1.06	6.0
Conf	ΔG	%	Conf	ΔG	%	Conf	ΔG	%	Conf	ΔG	%	Conf	ΔG	%
C_L19	0.00	46.1	C_p07	0.00	64.3	$C_{ps}13$	0.00	59.8	$C_{tp}21$	0.00	42.5	$C_{fp}06$	0.00	28.8
C_L09	0.31	27.3	C_p24	0.96	12.7	$C_{ps}10$	0.29	36.9	$C_{tp}23$	0.07	37.4	$C_{fp}04$	0.04	27.0
C_L01	0.61	16.3	C_s07	0.99	12.0	$C_{ps}28$	2.06	1.9	$C_{tp}04$	0.97	8.3	$C_{fp}01$	0.23	19.7
C_L21	1.50	3.7	C_p27	1.42	5.8	$C_{ps}01$	2.50	0.9	$C_{tp}09$	1.23	5.3	$C_{fp}05$	0.89	6.3
C_L07	1.76	2.4	C_p26	1.82	3.0	$C_{ps}06$	3.23	0.3	$C_{tp}01$	1.28	4.9	$C_{fp}25$	0.94	5.9

[a]Conf stands for conformer; ΔE (ΔG) was calculated relative to the lowest *E* (*G*) energy conformer; % is the %-fraction of the total population from Boltzmann distribution.

Table 5.2. Five lowest energy conformers of all H_nL in CSM, using either *E* or *G* values, for: part (a) 2,2,2-tet and part (b) 3,2,3-tet.[a]

(a)

L			HL			H_2L			H_3L			H_4L		
Conf	ΔE	%	Conf	ΔE	%	Conf	ΔE	%	Conf	ΔE	%	Conf	ΔE	%
$C_{fL}01$	0.00	18.83	C_s01	0.00	41.41	$C_{ps}01$	0.00	60.29	$C_{tp}01$	0.00	81.22	$C_{fp}02$	0.00	92.70
$C_{fL}02$	0.21	13.30	C_s02	0.29	25.38	$C_{ps}02$	0.33	34.45	$C_{tp}02$	1.23	10.14	$C_{fp}05$	1.50	7.30
$C_{fL}03$	0.29	11.54	C_s03	0.46	19.16	$C_{ps}09$	1.79	2.95	$C_{tp}03$	1.81	3.85			
$C_{fL}04$	0.43	9.15	C_s04	0.82	10.34	$C_{ps}10$	1.93	2.31	$C_{tp}06$	2.36	1.51			
$C_{fL}05$	0.77	5.14	C_p01	2.07	1.27				$C_{tp}07$	2.48	1.23			
Conf	ΔG	%	Conf	ΔG	%	Conf	ΔG	%	Conf	ΔG	%	Conf	ΔG	%
$C_{fL}01$	0.00	18.57	C_s01	0.00	30.94	$C_{ps}01$	0.00	33.04	$C_{tp}01$	0.00	75.05	$C_{fp}02$	0.00	72.57
$C_{fL}02$	0.21	13.12	C_s02	0.29	18.96	$C_{ps}02$	0.33	18.88	$C_{tp}02$	1.23	9.37	$C_{fp}03$	0.72	21.34
$C_{fL}03$	0.29	11.38	C_s03	0.46	14.32	$C_{ps}03$	0.49	14.35	$C_{tp}03$	1.81	3.56	$C_{fp}05$	1.50	5.72
$C_{fL}04$	0.43	9.02	C_s04	0.82	7.73	$C_{ps}04$	0.88	7.42	$C_{tp}04$	1.99	2.59	$C_{fp}06$	3.40	0.23
$C_{fL}05$	0.77	5.06	C_s05	0.88	6.97	$C_{ps}05$	1.40	3.13	$C_{tp}05$	2.28	1.59	$C_{fp}07$	3.69	0.14

(b)

L			HL			H_2L			H_3L			H_4L		
Conf	ΔE	%	Conf	ΔE	%	Conf	ΔE	%	Conf	ΔE	%	Conf	ΔE	%
$C_{fL}01$	0.00	25.74	C_s03	0.00	42.00	$C_{ps}14$	0.00	28.34	$C_{tp}02$	0.00	40.47	$C_{fp}01$	0.00	96.41
$C_{fL}05$	0.21	17.95	C_s01	0.19	30.72	$C_{ps}03$	0.04	26.59	$C_{tp}07$	0.30	24.41	$C_{fp}02$	1.95	3.59
$C_{fL}02$	0.40	13.08	C_s02	0.74	11.95	$C_{pp}01$	0.60	10.21	$C_{tp}01$	0.47	18.40			
$C_{fL}06$	0.81	6.54	C_s04	0.75	11.78	$C_{pp}05$	0.73	8.21	$C_{tp}09$	0.67	12.97			
$C_{fL}03$	0.83	6.36	C_p01	1.72	2.29	$C_{pp}02$	0.99	5.29	$C_{tp}05$	1.62	2.60			
Conf	ΔG	%	Conf	ΔG	%	Conf	ΔG	%	Conf	ΔG	%	Conf	ΔG	%
$C_{fL}01$	0.00	25.52	C_s03	0.00	35.93	$C_{ps}14$	0.00	27.69	$C_{tp}02$	0.00	39.93	$C_{fp}01$	0.00	96.32
$C_{fL}05$	0.21	17.80	C_s01	0.19	26.29	$C_{ps}03$	0.04	25.98	$C_{tp}07$	0.30	24.09	$C_{fp}02$	1.95	3.59
$C_{fL}02$	0.40	12.97	C_s02	0.74	10.23	$C_{pp}01$	0.60	9.98	$C_{tp}01$	0.47	18.15	$C_{fp}04$	4.15	0.09
$C_{fL}06$	0.81	6.49	C_s04	0.75	10.07	$C_{pp}05$	0.73	8.02	$C_{tp}09$	0.67	12.79			
$C_{fL}03$	0.83	6.31	C_s10	0.99	6.73	$C_{pp}02$	0.99	5.17	$C_{tp}05$	1.62	2.57			

[a]Conf stands for conformer; ΔE (ΔG) was calculated relative to the lowest *E* (*G*) energy conformer; % is the %-fraction of the total population from Boltzmann distribution.

taken because it might result in smaller number of conformers subjected to frequency calculations. Note, that in the case of the *G*-path shown in Scheme 5.4, frequency calculations were performed for all conformers falling within 4 kcal/mol electronic energy window of pre-optimisation.

5.4.3. Computed protonation constants.

We tested numerous competition reaction types, such as shown in Scheme 5.1 and Scheme 5.2, but whenever the reference molecule $L^{(2)}$ had (i) more than one proton relative to molecule under investigation (*e.g.*, $L^{(1)} + H_2L^{(2)}$ or $HL^{(1)} + H_3L^{(2)}$), (ii) the same number of protons (*e.g.*, $HL^{(1)} + HL^{(2)}$) or (iii) smaller number of protons (*e.g.*, $H_3L^{(1)} + H_2L^{(2)}$) results obtained were of poor quality - some examples are provided in Table C1 of Appendix C . This is in full agreement with previous reports.[12,13] Therefore, we would only be discussing results obtained from competition reactions where two competing for a proton ligands are involved in the protonation reaction of the same order, $H_{n-1}L^{(1)} + H_nL^{(2)} = H_nL^{(1)} + H_{n-1}L^{(2)}$, where n = 1,2,…,NPS and NPS stands for the number of protonation steps a ligand can be involved in, here four. The computed protonation constants obtained in different solvation models are presented in Table 5.3 where either a single conformer with the lowest *G* value (shown under column heading '*G* of LEC') or the weighted *G* values of selected conformers with the %-fraction > 5 in *G* (under column heading 'Weighted *G*') were used.

To assess quality of computed protonation constants one must consider two important aspects, namely (i) the error in computed protonation constant relative to the relevant experimental $\log K_H^{(n)}$ values (9.75, 9.07, 6.58 and 3.27 for the consecutive, from first to forth, stepwise protonation constant[14] of 2,2,2-tet; for the reference molecule, 3,2,3-tet, $\log K_H^{(n)}$ values of 10.53, 9.77, 8.30 and 5.59 for the first to fourth stepwise protonation constant[14] were used) and (ii) theoretically predicted sequence in values of protonation constants. The second criterion is also of an utmost importance because the experimental first and second protonation constants of 2,2,2-tet differ only by less than 0.7 log unit (a typical feature among polyamines; note also that 0.76 log unit difference is observed for 3,2,3-tet) which, in principle, can be seen as hardly achievable when typical accuracy obtained from computational work reported to date is considered. The analysis of the data in Table 5.3 demonstrates that, indeed, it is possible to predict all stepwise protonation constants of 2,2,2-tet in correct sequence and with errors smaller than 1 log unit but only when structures with explicit water molecules were used and *E*-path was followed. Interestingly and importantly, results obtained from a single and weighted *G* values

(data under the '*G* of LEC' and 'Weighted *G*' headings in *E*-path) are comparable as they differ by about ±0.1 log unit (see Δ_1 –Δ_2 values in Table 5.3).

The first protonation constant can be seen as of analytical quality as it differs from the experimental $\log K_H^{(1)}$ value by –0.01 and 0.08 log unit when a single or weighted *G* value was used, respectively, whereas the second and third protonation constants we regard as excellent prediction as they reproduced experimental values just to within –0.3±0.1 log units.

Table 5.3. Computed from *E*- and *G*-paths protonation constants, as $\log K_H^{(1)}$, for 2,2,2-tet using data from a discrete-continuum solvation model (DCSM) in part (a) and continuum solvation model (CSM) in part (b)[a]

(a)

	***E*-path**				
DCSM	***G* of LEC**		**Weighted *G***		$\Delta_1 - \Delta_2$
Reaction	$\log K_H^{(n)}$	Δ_1	$\log K_H^{(n)}$	Δ_2	
$L^{(1)} + HL^{(2)} = HL^{(1)} + L^{(2)}$	9.74	–0.01	9.83	0.08	–0.09
$HL^{(1)} + H_2L^{(2)} = H_2L^{(1)} + HL^{(2)}$	8.87	–0.20	8.75	–0.32	0.13
$H_2L^{(1)} + H_3L^{(2)} = H_3L^{(1)} + H_2L^{(2)}$	6.12	–0.46	6.19	–0.39	–0.07
$H_3L^{(1)} + H_4L^{(2)} = H_4L^{(1)} + H_3L^{(2)}$	2.41	–0.86	2.50	–0.77	–0.09
	***G*-path**				
	***G* of LEC**		**Weighted *G***		
$L^{(1)} + HL^{(2)} = HL^{(1)} + L^{(2)}$	10.14	0.39	10.21	0.46	–0.07
$HL^{(1)} + H_2L^{(2)} = H_2L^{(1)} + HL^{(2)}$	7.37	–1.70	7.15	–1.92	0.22
$H_2L^{(1)} + H_3L^{(2)} = H_3L^{(1)} + H_2L^{(2)}$	7.40	0.82	7.42	0.84	–0.02
$H_3L^{(1)} + H_4L^{(2)} = H_4L^{(1)} + H_3L^{(2)}$	3.34	0.07	3.44	0.17	–0.09

(b)

	***E*-path**				
CSM	***G* of LEC**		**Weighted *G***		$\Delta_1 - \Delta_2$
Reaction	$\log K_H^{(n)}$	Δ_1	$\log K_H^{(n)}$	Δ_2	
$L^{(1)} + HL^{(2)} = HL^{(1)} + L^{(2)}$	8.95	–0.80	8.88	–0.87	0.07
$HL^{(1)} + H_2L^{(2)} = H_2L^{(1)} + HL^{(2)}$	7.71	–1.36	7.80	–1.27	–0.09
$H_2L^{(1)} + H_3L^{(2)} = H_3L^{(1)} + H_2L^{(2)}$	6.34	–0.24	6.34	–0.24	0.00
$H_3L^{(1)} + H_4L^{(2)} = H_4L^{(1)} + H_3L^{(2)}$	–0.96	–4.23	–1.13	–4.40	0.17
	***G*-path**				
	***G* of LEC**		**Weighted *G***		
$L^{(1)} + HL^{(2)} = HL^{(1)} + L^{(2)}$	8.95	–0.80	8.85	–0.90	0.10
$HL^{(1)} + H_2L^{(2)} = H_2L^{(1)} + HL^{(2)}$	7.71	–1.36	7.72	–1.35	–0.01
$H_2L^{(1)} + H_3L^{(2)} = H_3L^{(1)} + H_2L^{(2)}$	6.34	–0.24	6.45	–0.13	–0.11
$H_3L^{(1)} + H_4L^{(2)} = H_4L^{(1)} + H_3L^{(2)}$	–0.96	–4.23	–1.23	–4.50	0.27

[a]Weighted *G* values were obtained using each conformers fraction of the total population (from Boltzmann distribution) as a weight for their *G* contribution ($w \times G$); Δ_n = computed – experimental $\log K_H^{(n)}$.

used, respectively, whereas the second and third protonation constants we regard as excellent prediction as they reproduced experimental values just to within –0.3±0.1 log units.

Furthermore, one observes a unidirectional error obtained for the second, third and fourth protonation constants (computed values are consistently smaller relative to experimental values) and the departure increases with the increase in the protonation constant number. This, most likely, might be attributed to somewhat poorer performance of energy optimisation in case of highly charged molecules; the larger the positive charge on molecules involved in the CRn, the larger difference between experimental and computed values.

To this effect, it has been noted previously that aliphatic polyamines are difficult to model using most quantum chemical solvation models[41]; hence, we consider results reported here as highly satisfactory and significant improvement relative to data reported for amines previously. It is also possible to assume that closer to experimental values third and fourth protonation constants could be obtained by placing larger number of explicit water molecules to 'better' disperse charges on the macro-molecular assembly (*e.g.*, $H_nL + 8H_2O$) immersed in a simulated water environment.

Our focus now is on *G*-path for data obtained with explicit water molecules – part (a) in Table 5.3. Except for the second protonation constant ($\log K_H^{(2)}$ was underestimated by about 1.7–1.9 log units) results obtained could be seen as satisfactory because they fall within or below the typically reported error ranges when TCs are used for neutral or singly charged molecules.[8–10] Unfortunately, the overall quality of data obtained from the *G*-path in DCSM must be seen as unacceptable. This is because the experimental sequence of protonation constants, $\log K_H^{(n)} > \log K_H^{(n+1)}$, is not reproduced. To illustrate this, performance of different methodologies tested in this work is depicted in Figure 5.4 as differences between successive protonation constants, $\Delta \log K_H^{(n,n+1)} = \log K_H^{(n)} - \log K_H^{(n+1)}$, where such values obtained for experimental data are also included. Clearly, most accurate protonation constants were computed from *E*-path in DCSM – see top left graph in Figure 5.4.

To gain some insight on the origin of the observed disparity in accuracy between *E*- and *G*-paths, we compared structures of relevant conformers; their *E* and/or *G* values were used to select conformers for computing the second protonation constant. Figure 5.5 shows the lowest energy conformers of diprotonated 2,2,2-tet and 3,2,3-tet obtained from the *G*-path whereas those for *E*-path are shown in Figure 5.3 (additional structures are shown in Figures C3–C22 in PARTS 3 and 4 of Appendix C). Structural comparison revealed that conformers selected from

E-path have a compact structure with water molecules in the first solvation shell being arranged such that (i) polyamines form a ring closed by water molecules and (ii) each protonated site is involved in interactions with several water molecules. In contrast, the *G*-path produced conformers with extended configurations of polyamines with explicit water molecules (i) distributed unevenly between two terminal functional groups and (ii) not interacting with all protonated sites. Hence, these structures tend to have increased entropic contributions to their

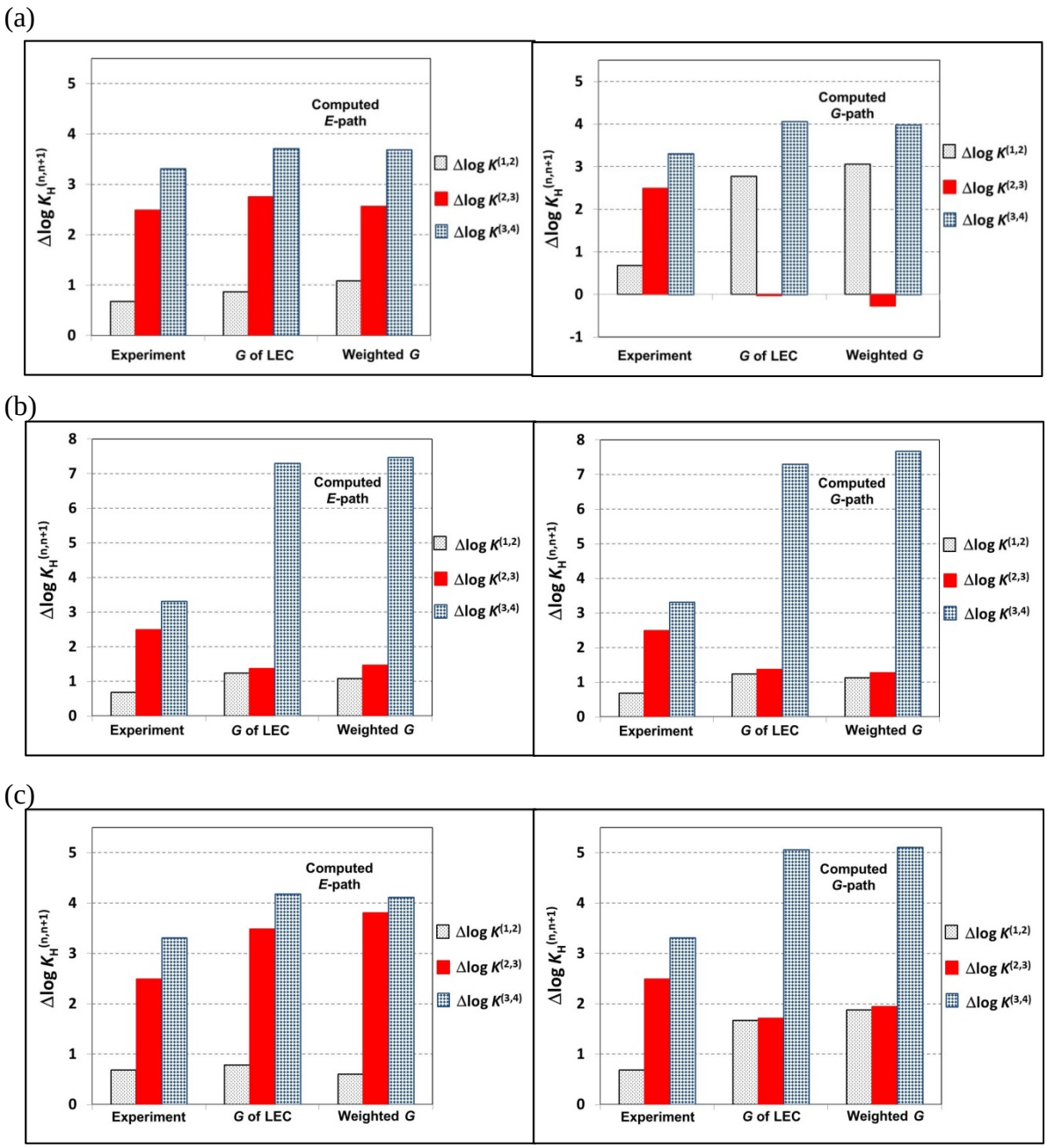


Figure 5.4. Graphical presentation of differences between successive stepwise protonation constants, $\Delta \log K_{H}^{(n,n+1)} = \log K_{H}^{(n)} - \log K_{H}^{(n+1)}$, for experimental and computed data in: part (a) – DCSM at the B3LYP level, part (b) – CSM at B3LYP level, and part (c) – DCSM with dispersion corrected B97D level of theory, all with the 6-311++G(d,p) functional.

free energy compared to those selected from the *E*-path. Their increased entropic correction is due to arrangements of explicit water molecules which may artificially result in a greater number of low frequency (*i.e.* soft) vibrational modes.

Consequently, these low frequency vibration modes contribute significantly to increased thermal entropy and lower ZPVE contributions to the Gibbs free energy of a molecular system.[51–53] Also, the inability of the RRHO model to correctly evaluate vibrational frequencies, especially for such low vibrational modes, may compound this problem since thermal corrections to electronic energies depend on computed vibrational frequencies.[51] To correct for this effects, one would have to introduce the anharmonic correction, specifically for those identified at low frequency modes (*i.e.* using the so-called quasi-harmonic model) and this is not a trivial task especially when the DCSM is utilized to describe solvent environment.

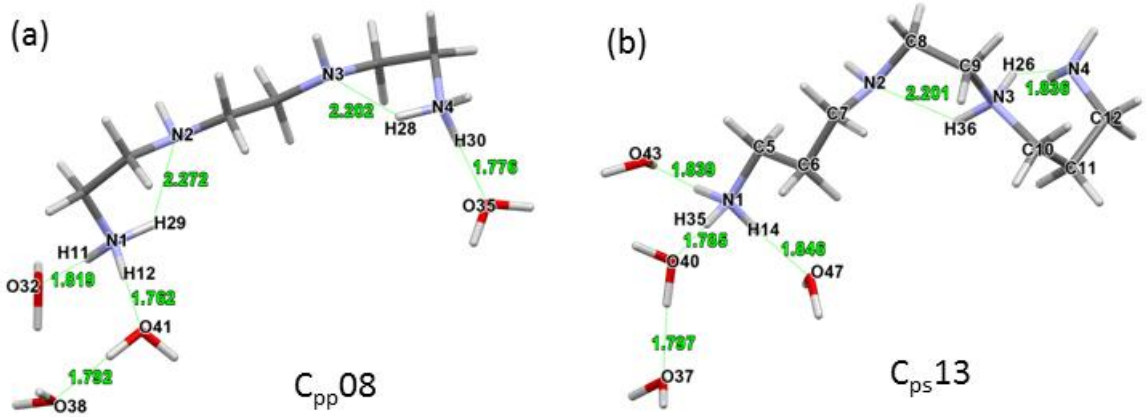


Figure 5.5. LECs selected from the *G*-path for H_2L of (a) 2,2,2-tet and (b) 3,2,3-tet.

Furthermore, other factors such as the coupling of rotational and vibrational modes in solution, particularly when discrete solvent molecules are included, may also affect the accuracy of computed thermal corrections.[9] The overall effect of all these factors is such that, in most cases, the LEC selected by the *G*-path dominates the conformer population even though it has a higher electronic energy (*i.e.* it is less stable) than the one selected from the *E*-path. A similar discrepancy between Gibbs free energy and electronic energy based selection of LECs has been reported by Salehzadeh et al.[16] in their study of micro protonation constants of spermine.

Considering results obtained in computationally least expensive medium, CSM, data in Table 5.3 and Figure 5.4(b) shows that although the sequence in protonation constants has been reproduced correctly by both, *E*- and *G*-paths, the results obtained for (i) the fourth protonation

constant which was underestimated by more than four log units and (ii) the difference between the second and third protonation constants, which is much too small, must be seen as unacceptable. Moreover, the overall mean absolute deviation (1.72) of predicted stepwise protonation constants in CSM is over four times larger than that obtained in DCSM (0.37). This is not entirely surprising as (i) proper modelling of the solvation environment determines to a large extent the accuracy of computed protonation constants in general[9,41–44] and (ii) reasonable computational evaluation of the solution free energy for highly charged ionic species, such as H_3L and H_4L forms of aliphatic polyamines, is usually prone to large errors unless explicit water molecules are incorporated to describe the first solvation layer.[9,41]

Interestingly, we noted that in the case of the *G*-path implemented in CSM, the second protonation constant was correctly predicted to be larger than the third one whereas the same protocol failed in this respect when implemented in DCSM. In search for possible origin of this observation we examined relevant conformers which were selected from these two solvation models. Considering DCSM, we found that, in line with experimental observations[5,7], the LECs of 2,2,2-tet were mainly those of HL_p and H_2L_{pp} tautomers of the HL and H_2L forms, respectively. In contrast, the LECs of 3,2,3-tet with largest %-fraction of the total population were those of HL_p and H_2L_{ps} tautomers. Therefore, differences in charge distribution on conformers of H_2L_{pp} and H_2L_{ps} tautomers selected for 2,2,2-tet and 3,2,3-tet, respectively, in combination with uneven water molecules' distribution might be responsible for inaccuracy of the second protonation constant of 2,2,2-tet when *G*-path was followed in DCSM. This correlates well with previous studies where it has been pointed out that similarity of charge and its distribution between a reference molecule and the molecule of interest appears to be of utmost importance in accurate prediction of protonation constants using the CRn methodology.[12,13]

It is also important to note that with CSM (part (b) in Table 5.3), there is no apparent difference in predicted protonation constants using either *E*- or *G*-paths. This is due to the fact that the computed electronic and Gibbs free energies of the LECs followed exactly the same trends in relative values; hence, the selected sets of LECs with %-fraction > 5% from *E*- and *G*-paths were very much the same (a feature which is not observed in DCSM). Clearly, the absence of explicit water molecules eliminated all the above mentioned complications and uncertainties in computed *G* values.

Finally, we also tested whether accuracy of predicted protonation constants could be improved by accounting for dispersion interactions as their importance in obtaining accurate thermochemical parameters has been emphasized recently.[54] To accomplish this, we re-

optimised all conformers with %-fractions > 5 found in DCSM at the B97D and B3LYP-gD3 levels of theory, both with 6-311++G(d,p) functional. For both levels of theory protonation constants was predicted correctly from both, *E*- and *G*-paths – see Table 5.4 but overall results obtained at B97D are much better than those at B3LYP-gD3 – see Figure 5.6; hence, we will

Table 5.4. Computed from *E*- and *G*-paths protonation constants, as $\log K_H^{(n)}$, for 2,2,2-tet using data from dispersion corrected DFT in a discrete-continuum solvation model using B3LYP-gD3 in part(a) and B97D in part (b).[a]

(a)

	***E*-path**				
	***G* of LEC**		**Weighted *G***		
Reaction	$\log K_H^{(n)}$	Δ_1	$\log K_H^{(n)}$	Δ_2	$\Delta_1 - \Delta_2$
$L^{(1)} + HL^{(2)} = HL^{(1)} + L^{(2)}$	11.45	1.70	11.34	1.59	0.11
$HL^{(1)} + H_2L^{(2)} = H_2L^{(1)} + HL^{(2)}$	10.30	1.23	10.47	1.40	0.17
$H_2L^{(1)} + H_3L^{(2)} = H_3L^{(1)} + H_2L^{(2)}$	5.63	–0.95	5.37	–1.21	0.26
$H_3L^{(1)} + H_4L^{(2)} = H_4L^{(1)} + H_3L^{(2)}$	1.98	–1.29	2.14	–1.13	0.16
	***G*-path**				
	***G* of LEC**		**Weighted *G***		
$L^{(1)} + HL^{(2)} = HL^{(1)} + L^{(2)}$	11.45	-1.70	11.62	-1.87	0.17
$HL^{(1)} + H_2L^{(2)} = H_2L^{(1)} + HL^{(2)}$	9.27	-0.20	9.21	-0.14	0.06
$H_2L^{(1)} + H_3L^{(2)} = H_3L^{(1)} + H_2L^{(2)}$	6.66	-0.08	6.68	-0.10	0.02
$H_3L^{(1)} + H_4L^{(2)} = H_4L^{(1)} + H_3L^{(2)}$	1.98	1.29	2.00	1.27	0.02

(b)

	***E*-path**				
	***G* of LEC**		**Weighted *G***		
Reaction	$\log K_H^{(n)}$	Δ_1	$\log K_H^{(n)}$	Δ_2	$\Delta_1 - \Delta_2$
$L^{(1)} + HL^{(2)} = HL^{(1)} + L^{(2)}$	10.56	–0.81	11.34	–1.59	0.78
$HL^{(1)} + H_2L^{(2)} = H_2L^{(1)} + HL^{(2)}$	9.78	–0.71	10.47	–1.40	0.69
$H_2L^{(1)} + H_3L^{(2)} = H_3L^{(1)} + H_2L^{(2)}$	6.30	0.28	5.37	1.21	0.93
$H_3L^{(1)} + H_4L^{(2)} = H_4L^{(1)} + H_3L^{(2)}$	2.12	1.15	2.14	1.13	0.02
	***G*-path**				
	***G* of LEC**		**Weighted *G***		
$L^{(1)} + HL^{(2)} = HL^{(1)} + L^{(2)}$	10.56	–0.81	10.85	–1.10	0.29
$HL^{(1)} + H_2L^{(2)} = H_2L^{(1)} + HL^{(2)}$	8.89	0.18	8.97	0.10	0.08
$H_2L^{(1)} + H_3L^{(2)} = H_3L^{(1)} + H_2L^{(2)}$	7.18	–0.60	7.03	–0.45	0.15
$H_3L^{(1)} + H_4L^{(2)} = H_4L^{(1)} + H_3L^{(2)}$	2.12	1.15	1.92	–1.35	0.20

[a]Weighted *G* values were obtained using each conformers fraction of the total population (from Boltzmann distribution) as a weight for their *G* contribution ($w \times G$); Δ_n = computed – experimental $\log K_H^{(n)}$.

focus on the former. In general, one could consider B97D-predicted protonation constants as reasonable as, on average, the departure in absolute terms from experimental $\log K_H^{(n)}$ values for all protonation constants from *E*- and *G*-paths combined was 0.9 ± 0.5 log units with the largest deviations found for the first protonation constant which was overestimated by about 1.6 log

units. However, it has been pointed out[55] that in certain instances, addition of empirical dispersion correction accounts properly for short-range (intramolecular) but fails for long-range (intermolecular) dispersion effects. This results in imbalance between intra- and intermolecular dispersion effects on electronic structure which might be responsible for larger errors in computed protonation constants when compared with dispersion-uncorrected B3LYP functional. In addition, accuracy in $\log K_{\mathrm{H}}^{(\mathrm{n})}$ values obtained using the B3LYP functional might be also due to hidden error cancellations[56,57], a unique situation for 'electronically simple' molecules (such as aliphatic polyamines) and, as such, our results do not preclude the use of dispersion corrected functionals when carrying out this kind of investigation on other molecules.

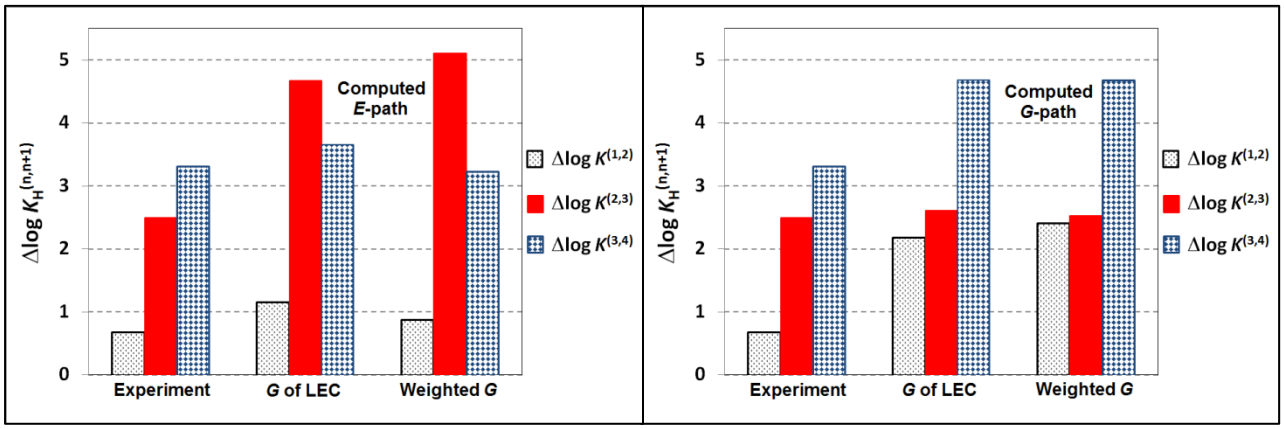


Figure 5.6. Graphical presentation of differences between successive stepwise protonation constants, $\Delta\log K^{(\mathrm{n,n+1})} = \log K_{\mathrm{H}}^{(\mathrm{n})} - \log K_{\mathrm{H}}^{(\mathrm{n+1})}$, for computed data at the B3LYP-gD3 level of theory with DCSM (values obtained for experimental data is also included for comparison).

However, one must also realize that full-scale comparative studies on entire sets of all conformers found from MM-search using several functionals would be even more time demanding and because results obtained here are very satisfactory, one would have to justify if it is really worthwhile to strive for a small, a fraction of a log unit, just possible but not guaranteed improvement for the third and fourth protonation constant.

5.4.4. Testing reliability of pre-optimisation protocol.

Even though the developed general purpose protocol provided theoretically predicted protonation constants (i) of outstanding quality relative to typically reported data in the computational field, (ii) appears to work well for molecules with multiple positive charges and (iii) can be seen as reliable to provide a valuable insight on relevant properties of 2,2,2-tet for a solution chemist, we decided to test it further. Clearly, prior to recommending any protocol as of

general purpose, it is important to find out whether some LECs were missed and, if this was the case, what impact on quality of computed $\log K_H^{(n)}$ values that would have.

To this effect, we have focused on structures with explicit water molecules as the best and reliable results were obtained only in DCSM and decided to fully energy optimise the 'redundant' conformers which were rejected after the pre-optimisation. Firstly, we wanted to find out whether (i) a new and the lowest in electronic energy conformer could be discovered and secondly (ii) new conformers would have to be included in the LECs sets, within 2 kcal/mol window, which had to be used in computing $\log K_H^{(n)}$ in case of 'Weighted *G*' strategy. Data in Table 5.5 shows that:

a) In all cases the lowest electronic energy conformer has been identified from the pre-optimisation protocol. This is gratifying finding because, as pointed out above (Table 5.3) it is sufficient to use the single *G* value from *E*-path to obtain excellent prediction in the $\log K_H^{(n)}$ values (recall that they hardly differ from those obtained using computationally more expensive weighted *G* values of the selected LECs).

b) Only in one case we found an additional conformer, that of H_2L_{ps}, which was within the 2 kcal/mol window of conformers to be selected for computing protonation constants. It is important to stress here that this conformer was not the lowest in electronic energy; hence, it could only influence results obtained from *E*-path involving weighted *G*.

Table 5.5. Summary of identified (yes) and missed (no) lowest energy conformers of 2,2,2-tet and 3,2,3-tet in DCSM from proposed the pre-optimization protocol involving selection of conformers after 20 optimization steps with 4 kcal/mol *E*-window showing also an impact on computed protonation constants.[a]

	2,2,2-tet		3,2,3-tet		F-set	R-set	F-set	R-set
Prot. form	**The LEC**	**LECs**	**The LEC**	**LECs**	**log $K_H^{(2)}$**	**log $K_H^{(2)}$**	**log $K_H^{(3)}$**	**log $K_H^{(3)}$**
L	yes	yes	yes	yes	–	–	–	–
HL_p	yes	yes	yes	yes	–	–	–	–
HL_s	yes	yes	yes	yes	–	–	–	–
H_2L_{ps}	yes	no (1)*	yes	yes	8.68	8.75	6.25	6.19
H_2L_{pp}	yes	yes	yes	yes	–	–	–	–
H_3L	yes	yes	yes	yes	–	–	–	–
H_4L	yes	yes	yes	yes	–	–	–	–

[a]Prot. form stands for protonated form of a polyamine, The LEC stands for the lowest in electronic energy conformer, F-set = full set of LECs found after full optimization, R-set = reduced set of LECs as found from pre-optimization protocol. *(1) indicates that one new conformer was added to the set of LECs after full optimization.

The newly discovered H_2L_{ps} conformer of 2,2,2-tet has changed the number of LECs from two (from pre-optimisation) to three (after full optimisation of all conformers) – see Figure 5.7. These three conformers were combined with four H_2L_{pp} LECs of 2,2,2-tet (this set has not changed after full optimisation) and those with %-fraction above 5% (from Boltzmann distribution done on seven combined conformers) were used to compute protonation constants. In other words, the protocol developed here and described in details in proceeding sections was fully followed and we found that this conformer was predicted to contribute 6% to the total population when free energies of seven LECs were used.

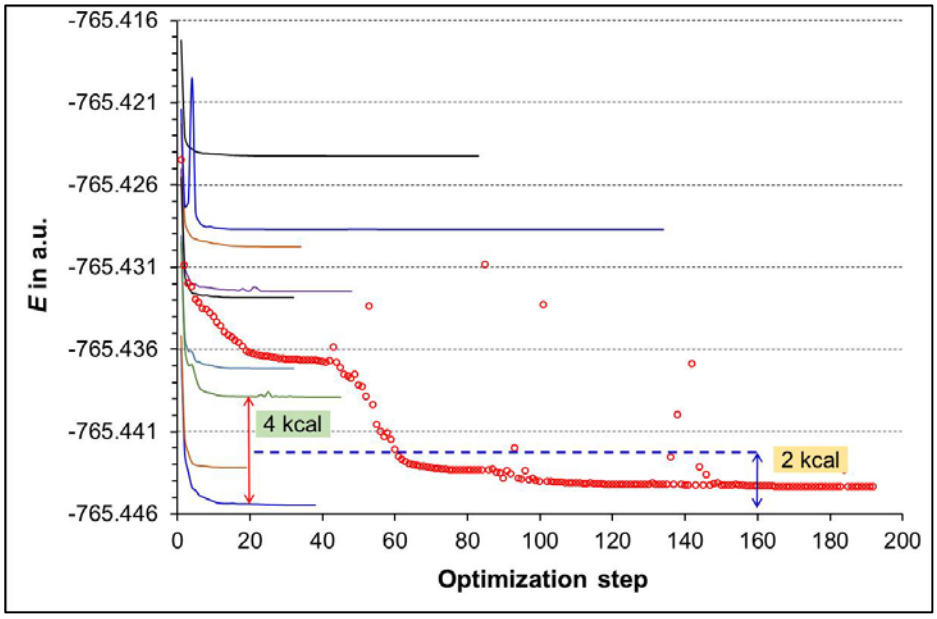


Figure 5.7. Optimisation profile for conformers of 2,2,2-tet in DCSM showing the 4 kcal/mol E-window at twentieth step used to select structures for full optimisation when the pre-optimisation protocol was implemented and 2 kcal/mol E-window to select conformers required to compute protonation constants

One must note that two protonation constants, for H_2L and H_3L forms of 2,2,2-tet, had to be re-computed because $H_2L^{(1)}$ is involved in two protonation reactions, $HL^{(1)} + H_2L^{(2)} = H_2L^{(1)} + HL^{(2)}$ (for $\log K_H^{(2)}$) and $H_2L^{(1)} + H_3L^{(2)} = H_3L^{(1)} + H_2L^{(2)}$ (for $\log K_H^{(3)}$). The values of the re-calculated protonation constants changed by ±0.07 log unit; the second decreased by ~0.08 log unit whereas the third protonation constant increased by ~0.06 log unit. Clearly, this had no effect on the overall quality of $\log K_H^{(n)}$ values as well as the sequence of stepwise protonation constants.

Just for completeness, we have also found one new conformer, this time for HL_p of 2,2,2-tet, when full optimisation data in CSM was analysed and it is seen in Figure 5.2 as empty circles. In this instance, it had no influence on predicted protonation constants as its energy was at the border line of the 2 kcal/mol *E*-window.

5.4.5. Recommended protocols for protonation constants calculations

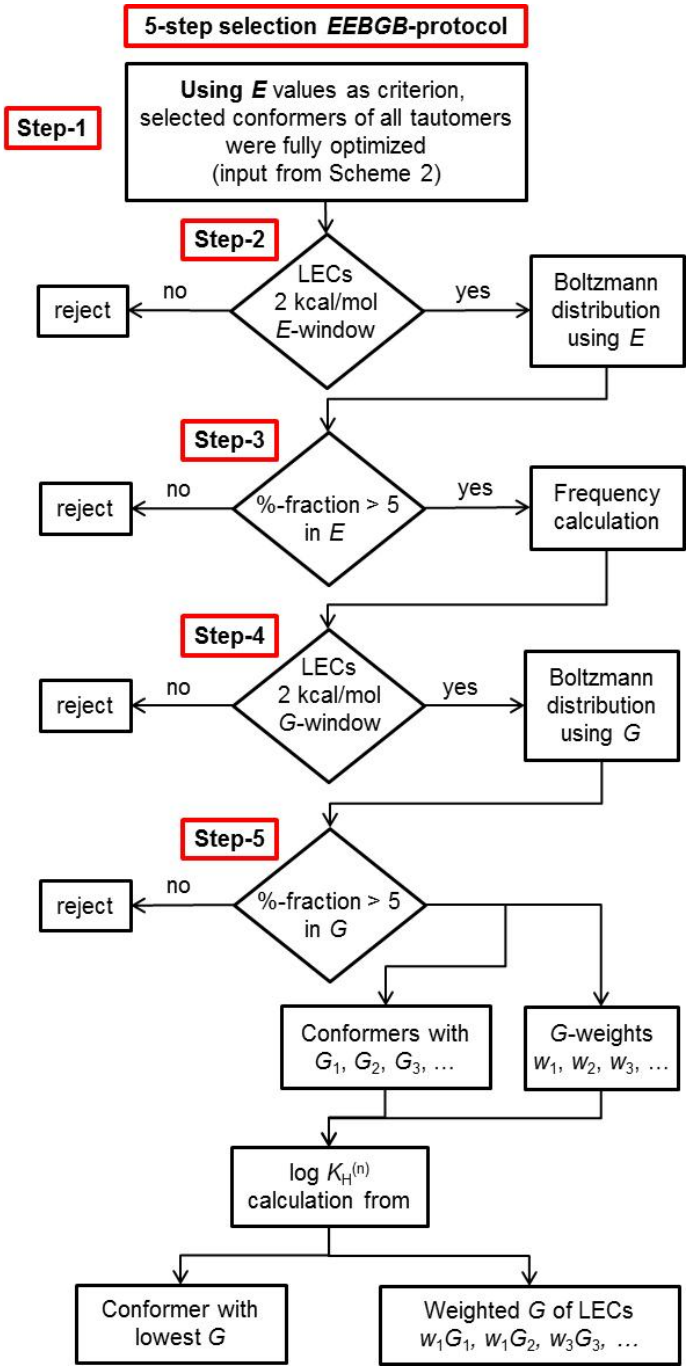
Data in Table 5.5 gave us confidence to reanalyse the entire set of optimised structures in DCSM as we wanted to come up with the final protocol(s) which would generate excellent results with minimum computational time. Hence, our focus was on fully optimised conformers which were selected from Scheme 5.3 and fine-tuning of steps implemented in *E*-path shown in Scheme 5.4. Analysis in variation in electronic energies and their influence on selected conformers lead to the final protocol shown in Scheme 5.5. We call it a 5-step selection *EEBGB*-protocol because it incorporates two steps involving *E*-based selection (one in the pre-optimisation operation and the second after full optimisation), Boltzmann distribution using *E* values to select conformers for frequency calculations, followed by *G*-based selection of conformers within 2 kcal/mol window of the free energy, and the final step from which, based on Boltzmann distribution using *G*-values of retained conformers, only those with %-fraction > 5 were retained. As an example, we will illustrate the performance of the *EEBGB*-protocol, in terms of reduction of the number of conformers after each step, using data obtained for the H_4L of 3,2,3-tet:

- **Step-1 (*E*-based selection)**: out of 30 MM-identified LECs, 28 were within 4 kcal/mol *E*-window after the pre-optimisation operation and they were fully optimised.
- **Step-2 (*E*-based selection)**: out of 28 fully optimised conformers, 14 were selected which were within 2 kcal/mol *E*-window. Note that in case of HL (HL_p and HL_s) and H_2L (H_2L_{pp} and H_2L_{ps}) tautomers are combined after full optimisation and only conformers within the 2 kcal/mol *E*-window are selected.
- **Step-3 (*B*-based selection)**: using *E* values of 14 selected conformers, 7 met the criterion of Boltzmann distribution %-fraction > 5; they were submitted for frequency calculations.
- **Step-4 (*G*-based selection)**: Only conformers within 2 kcal/mol *G*-window were selected; the set of 6 structures with their G_1, G_2, G_3, …G_n values (G_1 is the lowest in the free energy) were submitted for the Boltzmann distribution calculation.
- **Step-5 (*B*-based selection)**: using G_n values of retained 6 conformers, a selection criterion of Boltzmann distribution generated %-fraction > 5 was applied to obtain the FINAL set of conformers. In this particular case all 6 were retained and their fraction contributions were used as weights w_n.

Finally, the selected 6 conformers were used to compute protonation constants using either the single value of *G* of the lowest in the free energy conformer or weighted *G* value obtained by

pairing G_n and w_n values ($w_1 \times G_1 + w_2 \times G_2 + \ldots + w_n \times G_n$). This protocol has decreased the initial number of H_4L conformers from 30 MM-generated in Step-1 to 7 which were submitted for frequency calculations and 6 for protonation constants calculations.

The same *EEBGB*-protocol (Scheme 5.5) was implemented for each tautomer in the DCSM. A complete set of data for each tautomer is presented in Table 5.6 and this resulted in:



Scheme. 5.5. Recommended and time most-effective 5-step selection *EEBGB*-protocol for protonation constants calculations of polyamines.

Table 5.6. Step-wise selection of conformers needed and sufficient for protonation constants calculations using the 3-step (*EGB*), 4-step (*EEGB*) and most time-effective 5-step (*EEBGB*) protocol.* Data obtained for each tautomers are shown in PART (a) for 2,2,2-tet and PART (b) for 3,2,3-tet.

PART (a)			***EGB*-protocol**			***EEGB*-protocol**				***EEBGB*-protocol**				
2,2,2-tet + 4H_2O	**MM confs**	**Step-1**	**Freq. calc.**	**Step-2**	**Step-3**	**Step-2**	**Freq. calc**	**Step-3**	**Step-4**	**Step-2**	**Step-3**	**Freq. calc**	**Step-4**	**Step-5**
L	30	7	7	4	3	6	6	4	3	6	3	3	3	3
HL_p	30	1	2	2	2	2	2	2	2	2	2	2	2	2
HL_s	30	1												
H_2L_{ps}	30	3	12	3	3	2	2	2	2	12	2	2	2	2
H_2L_{pp}	30	9												
H_3L	30	8	8	8	4	2	2	2	2	8	2	2	2	2
H_4L	30	5	5	3	3	3	3	2	2	4	2	2	2	2
Sum:	210	34	**34**	20	15	15	**15**	12	11	32	11	**11**	11	11

PART (b)			***EGB*-protocol**			***EEGB*-protocol**				***EEBGB*-protocol**				
3,2,3-tet $4H_2O$	**MM confs**	**Step-1**	**Freq. calc.**	**Step-2**	**Step-3**	**Step-2**	**Freq. calc.**	**Step-3**	**Step-4**	**Step-2**	**Step-3**	**Freq. calc.**	**Step-4**	**Step-5**
L	30	10	10	6	3	1	1	1	1	1	1	1	1	1
HL_p	30	8	21	6	4	3	3	2	2	3	2	2	2	2
HL_s	30	13												
H_2L_{ps}	30	11	13	3	2	1	1	1	1	3	1	1	1	1
H_2L_{pp}	30	2												
H_3L	30	12	12	5	5	3	3	2	2	3	3	3	2	2
H_4L	30	28	28	9	5	14	14	7	5	14	7	7	6	6
Sum:	210	84	**84**	29	19	22	**22**	13	11	24	14	**14**	12	12
Total:	420	118	**118**	49	34	37	**37**	25	22	56	25	**25**	23	23
%-total:		**28**	**28**	**12**	**8**	**9**	**9**	**6**	**5**	**13**	**6**	**6**	**5**	**5**

*MM confs = MM-generated conformers; **Step-1** is common to all protocol and it involves selection of lowest energy conformers (LECs) falling within 4 kcal/mol *E*-window from pre-optimization in Gaussian after 20 cycles; **Step-2** in ***EGB*-protocol** involves selection of LECs falling within 2 kcal/mol *G*-window after frequency calculation; **Step-3** in ***EGB*-protocol** involves selection of conformers with %-fraction > 5 after Boltzmann distribution on selected LECs in Step-2; **Step-2** in ***EEGB*-protocol** involves selection of LECs falling within 2 kcal/mol *E*-window after full optimization of conformers selected in Step-1; **Step-3** in ***EEGB*-protocol** involves selection of LECs falling within 2 kcal/mol *G*-window after frequency calculation; **Step-4** in ***EEGB*-protocol** involves selection of conformers with %-fraction > 5 after Boltzmann distribution on selected LECs in Step-3; **Step-2** in ***EEBGB*-protocol** is as Step-2 in *EEGB*-protocol; **Step-3** in ***EEEGB*-protocol** involves selection of conformers with %-fraction > 5 after Boltzmann distribution on selected LECs in Step-2; **Step-4** in ***EEEGB*-protocol** involves selection of LECs falling within 2 kcal/mol *G*-window after frequency calculation; **Step-5** in ***EEEGB*-protocol** involves selection of conformers with %-fraction > 5 after Boltzmann distribution on selected LECs in Step-4.

a) 94% reduction in the number of conformers submitted for frequency calculations, from initial 420 MM-generated to 25 in the Step-5;

b) 75% reduction relative to the *G*-path which can be seen as a 3-step selection *EGB*-protocol – see Scheme C1 in Appendix C where 118 conformers were submitted for frequency calculations;

c) An additional 30% reduction of frequency calculations relative to the *E*-path shown in Scheme 5.3, which is a 4-step selection *EEGB*-protocol – see Scheme C2 in Appendix C. Importantly, this had no detrimental effect on the computed protonation constants at all - see Table 5.7.

Table 5.7. PART (a) Comparison of theoretically computed four stepwise protonation constants using the recommended and time most-efficient 5-step selection EEBGB-protocol and, for comparison shown in brackets, second time-efficient 4-step selection *EEGB*-protocol. PART (b) Averaged values from two methods (*G* of LEC and weighted *G*) of *EEGB*- and *EEBGB*-protocols.[a]

PART (a)

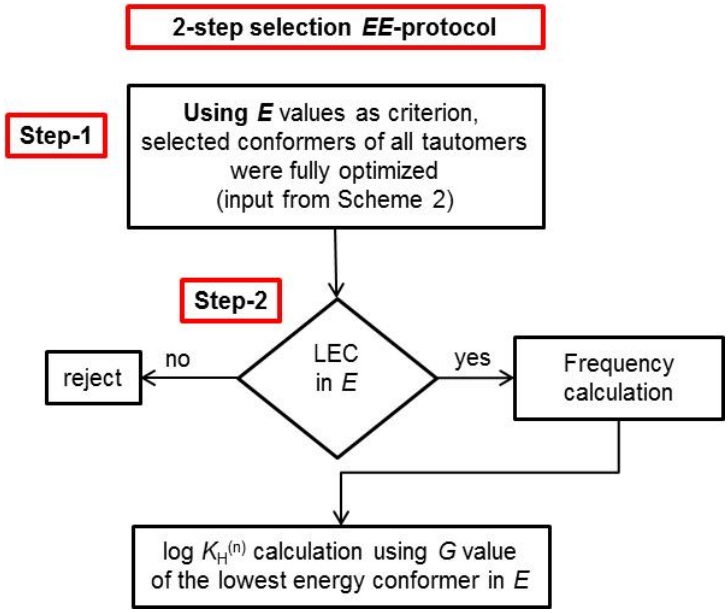
Step-wise protonation constant	**5-step selection *EEBGB*-protocol (4-step selection *EEGB*-protocol)**			
	***G* of LEC**	**Δ**	**Weighted *G***	**Δ**
$\log K_{\mathrm{H}}^{(1)}$	9.76 (9.74)	0.01 (–0.01)	9.94 (9.83)	0.19 (0.08)
$\log K_{\mathrm{H}}^{(2)}$	8.87 (8.87)	-0.20 (–0.20)	8.75 (8.75)	-0.32 (–0.32)
$\log K_{\mathrm{H}}^{(3)}$	6.12 (6.12)	-0.46 (–0.46)	6.19 (6.19)	-0.39 (–0.39)
$\log K_{\mathrm{H}}^{(4)}$	2.44 (2.41)	-0.83 (–0.86)	2.65 (2.50)	-0.62 (–0.77)
	Average $\lvert\Delta\rvert$	0.37 (0.38)	Average $\lvert\Delta\rvert$	0.38 (0.39)

PART (b)

Step-wise protonation constant	**Average from 2 methods of *EEGB*- and *EEBGB*-protocols**	
	Value	**Δ**
$\log K_{\mathrm{H}}^{(1)}$	9.82	0.07
$\log K_{\mathrm{H}}^{(2)}$	8.81	-0.26
$\log K_{\mathrm{H}}^{(3)}$	6.16	-0.42
$\log K_{\mathrm{H}}^{(4)}$	2.50	-0.77
Average $\lvert\Delta\rvert$		0.38
Standard deviation $\lvert\Delta\rvert$		0.30

[a]Δ = computed – experimental $\log K_{\mathrm{H}}^{(n)}$ value.

We have also tested a protocol where the only selection criterion was variation in electronic energy - it can be seen as a 2-step selection *EE*-protocol shown in Scheme 5.6.



Scheme 5.6. 2-step selection EE-protocol tested for selection of conformers for protonation constants calculations.

Interestingly, the computed protonation constants, as stepwise $\log K_H^{(n)}$ values, resulted in 10.39, 8.24, 5.55, and 3.15 (note that they follow protonation sequence correctly) with differences from experimental values of 0.64, –0.83, –1.03 and –0.12, respectively (0.7±0.3 log units for absolute differences from experimental values). From this follows that to compute preliminary but still reasonable estimates of stepwise protonation constants (they are at least as good, if not better, when commonly reported from thermodynamic cycles) it is sufficient to find the lowest in electronic energy conformer among possible tautomers and use their *G*-values as components in each stepwise competition reaction, $H_{n-1}L^{(1)} + H_nL^{(2)} = H_nL^{(1)} + H_{n-1}L^{(2)}$. To appreciate simplicity of the latter *EE*-protocol, one can write a general expression for the free energy change of competition reaction applicable to each *n*th protonation step;

$$\Delta G^{n}_{\mathrm{CRn}} = {}^{*}G(\mathrm{H_nL_t^{(1)}}) + {}^{*}G(\mathrm{H_{n\text{-}1}L_t^{(2)}}) - {}^{*}G(\mathrm{H_{n\text{-}1}L_t^{(1)}}) - {}^{*}G(\mathrm{H_nL_t^{(2)}}) \qquad (5.18)$$

where **G* stands for the free energy of the lowest electronic energy conformer found among tautomers of each protonated form of two polyamines, the one under investigation ($L^{(1)}$) and that used as a reference molecule ($L^{(2)}$). In contrast, when one would need to use weighted *G* values

$$\Delta G^{n}_{\mathrm{CRn}} = \sum_{k=1}^{t}\sum_{x=1}^{\mathrm{LEC}} w_x G_x(\mathrm{H_nL_t^{(1)}}) + \sum_{m=1}^{t}\sum_{y=1}^{\mathrm{LEC}} w_y G_y(\mathrm{H_{n-1}L_t^{(2)}}) - \sum_{n=1}^{t}\sum_{z=1}^{\mathrm{LEC}} w_z G_z(\mathrm{H_{n\text{-}1}L_t^{(1)}}) - \sum_{o=1}^{t}\sum_{s=1}^{\mathrm{LEC}} w_s G_s(\mathrm{H_nL_t^{(2)}})$$

(5.19)

of LECs found from Scheme 5.4 then the following expression for $\Delta G^{n}_{\mathrm{CRn}}$ applies where symbols are as described for expressions 5.12 and 5.13. Note that expression 5.19 is equally applicable to *E*- and *G*-paths in Scheme 5.4 as well as the refined *EGB*-, *EEGB* and *EEBGB*- protocols with the only, but significant, difference in the decreasing number of LECs obtained from the final third, fourth and fifth selection step, respectively.

5.5. Conclusions

This work has demonstrated that it is possible as well as time-wise and computationally feasible to theoretically predict stepwise protonation constants, as $\log K^{(n)}_{\mathrm{H}}$ values, of polyamines with the largest error smaller than 1 log unit, relative to the experimentally determined values from glass electrode potentiometry (GEP) which is most accurate among all analytical techniques in the field. In this particular case, where 2,2,2-tet was investigated, the predicted first protonation constant can be seen as of GEP-analytical quality as it differs from the experimental $\log K^{(1)}_{\mathrm{H}}$ value by less than ±0.1 log unit whereas the second and third might be seen as of NMR-analytical quality because they were predicted to within 0.2–0.4 log units of the GEP experimental values. Deviation from experimental values was systematic and unidirectional when going from the second $\log K^{(2)}_{\mathrm{H}}$ to the fourth $\log K^{(4)}_{\mathrm{H}}$ value (they all were underestimated) and the largest deviation, of about –0.8 log unit, was observed for the $\log K^{(4)}_{\mathrm{H}}$ value. It is important to stress that these results (which we see as of excellent overall quality) were obtained even though aliphatic linear polyamines are characterised by (i) numerous tautomers, (ii) almost an infinite number of possible conformers for each tautomer and (iii) very small, often well below 1 log unit, differences between consecutive protonation constants. Regarding the latter point, protocols developed here were also able to predict the values in correct order, in each case $\log K^{(n)}_{\mathrm{H}} > \log K^{(n+1)}_{\mathrm{H}}$ was reproduced as observed from experimental data.

Considering the quality of computed protonation constants we attribute this to successful implementation of the competition reaction (CRn) based methodology which requires (i) a polyamine under investigation, here $\mathrm{L}^{(1)}$ = 2,2,2-tet, and reference molecule, here $\mathrm{L}^{(2)}$ = 3,2,3-tet, to be structurally similar, (ii) correct selection of lowest energy conformers of all possible $\mathrm{H_nL}^{(1)}$ and $\mathrm{H_nL}^{(2)}$ tautomeric forms and (iii) balanced charge distribution between reactants and products, which in this case translates to $\mathrm{H_{n-1}L}^{(1)}$ and $\mathrm{H_nL}^{(2)}$ to be involved in a stepwise CRn, $\mathrm{H_{n-1}L}^{(1)} + \mathrm{H_nL}^{(2)} = \mathrm{H_nL}^{(1)} + \mathrm{H_{n-1}L}^{(2)}$, used to compute the $\log K^{(n)}_{\mathrm{H}}$ values. Furthermore, this

work has shown that it is not only sufficient to select lowest in electronic energy conformers (their Gibbs free energy values, *G*, are used in computing protonation constants) but it has resulted in higher quality of the $\log K_{H}^{(n)}$ values relative to *G*-based selection of LECs. However inspite of this success, the competition reaction based methodology has the following obvious limitations (i) It can only be used when an appropriate reference molecule (i.e. structurally similar to the molecule of interest) is available and (iii) Its use is dependent on the availability of experimental protonation constants for the identified reference molecule if it exists in the first place .

Regarding highly improved time and computational feasibility of theoretically predicting stepwise protonation constants, this has been achieved by implementing thoroughly investigated selection protocols developed in this work. The proposed *EEBGB*-protocol (*E*, *B* and *G* stand for electronic-energy-, Boltzmann-distribution- and Gibbs-free-energy-based stepwise selection of conformers – see Scheme 5.5) resulted in the 94% reduction of conformers submitted for frequency calculations from which four protonation constants were calculated, from initial 420 conformers selected from MM-based conformational search, to 25 in the final Step-5 of this protocol. Further reduction in time has been achieved by selecting conformers from an accelerated 'optimisation' operation, *i.e.*, instead of fully energy optimise all 420 MM-selected conformers, they were subjected to pre-optimisation involving only first 20 optimisation steps in Gaussian. Two important comments are in order here: (i) although we have verified validity of the accelerated 'optimisation' protocol by full optimisation of all, 420 2,2,2-tet and 3,2,3-tet structures, there is no guarantee that for larger polyamines (like penta- or hexamines) initial 20 optimisation steps will work perfectly well (one would have to consider either increasing the number of initial steps or perform full optimisation) and (ii) the pre-optimisation step will only influence time required for the first selection step in the developed *EEBGB*-protocol; the overall efficiency in the reduction of the number of conformers subjected to the frequency calculation remains intact. The reduced number of time-demanding frequency calculations is beneficial because, as this work shown, involving explicit water molecules significantly improves predictions in protonation constants.

Let us now comment on the systematic departure of computed $\log K_{H}^{(n)}$ values from experimental ones. We attribute this to intrinsic errors in computed energies when charges on molecules increase. The possible solution is to implement a stepwise increase in the number of explicit water molecules to dissipate the charge throughout the macromolecular assembly, (L + nH_2O) from four H_2O molecules for the singly protonated tetramines (this resulted here in

excellent prediction of $\log K_{\mathrm{H}}^{(1)}$ value) to, *e.g.*, seven H_2O molecules when H_3L and H_4L are involved (to compute the $\log K_{\mathrm{H}}^{(4)}$ value). It is also reasonable to assume that in case of polyamines with a larger number of protonation sites one should also need to increase the number of explicit water molecules.

Finally, it would be of fundamental importance to investigate an impact the functionals, such as B97D or the latest B3LYP-gD3, can make on the quality of computed protonation constants. To achieve that one would have to use them from the very beginning of the proposed protocol, namely all conformers selected from MM-based search would have to be (pre)optimised using dispersion-included functional. In our opinion, however, regardless of all the above comments related to feature studies, the proposed *EEBGB*-protocol can be successfully used and we are also of an opinion that its applicability is not restricted to polyamines.

5.6. References

[1] L. A. E. Batista de Carvalho, M. P. M. Marques, J. Tomkinson, *J. Phys. Chem. A.*, **2006**, *110*, 12947–12954.

[2] E. Agostinelli, M. P. M. Marques, R. Calheiros, F. P. S. C. Gil, G. Tempera, N. Viceconte, V. Battaglia, S. Grancara, A. Toninello, *Amino Acids*, **2010,** *38*, 393–403.

[3] M. P. M. Marques, L. A. E. Batista de Carvalho, *Biochem. Soc. Trans.*, **2007**, *35*, 374–380.

[4] G. J. S. Cooper, *Drugs*, **2011,** *71*, 1281–1320.

[5] M. Borkovec, D. Cakara, G. J. M. Koper, *J. Phys. Chem. B.*, **2012,** *116*, 4300–4309.

[6] I. Cukrowski, C. F. Matta, *Comput. Theoret. Chem.*, **2011**, *966*, 213–219.

[7] D. N. Hague, A. D. Moreton, *J. Chem. Soc. Perkin. Trans.*, **1994**, *2*, 265–270.

[8] R. Casasnovas, J. Ortega-Castro, J. Frau, J. Donoso, F. Munoz, *Int. J. Quantum Chem.*, **2014**, *114*, 1350–1363.

[9] J. Ho, *Aust. J. Chem*., **2014**. *67*, 1441–1460.

[10] J. Ho, M. L. Coote, *Comput. Mol. Sci.*, **2011,** *1*,649–660, doi:10.1002/WCMSW.43.

[11] M. Namazian, S. Halvani, *J. Chem. Thermodyn.*, **2006,** *38*, 1495–1502.

[12] K. K. Govender, I. Cukrowski, *J Phys. Chem. A.*, **2009**, *113*, 3639–3647.

[13] K. K. Govender, I. Cukrowski, *J. Phys. Chem. A.*, **2010,** *114*, 1868–1878.

[14] 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, U.S. Department of Commerce, National Institute of Standards and Technology.

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

[16] S. Salezadeh, Y. Gholiee, M. Bayat, *Int. J. Quantum Chem*., **2011**, *111*, 3608–3615.

[17] G. A. A. Saracino, R. Improta, V. Barone, Chem *Phys. Lett*, **2003**, *373*, 411–415.

[18] G. Schuurmann, M. Cossi, V. Barone, J. Tomasi, *J. Phys. Chem. A.*, **1998**, *102*, 6706–6712.

[19] M. Namazian, H. Heidary, *J. Mol. Struct*. (THEOCHEM), **2003**, *620*, 257–263.

[20] M. Namazian, S. Halvani, M. R. Noorbala, *J. Mol. Struct*. (THEOCHEM), **2004**, *711*, 13–18.

[21] M. Namazian, F. Kalantary-Fotooh, M. R. Noorbala, D. J. Searles, M. L. Coote, *J. Mol. Struct.* (THEOCHEM), **2006**, *758*, 275–278.

[22] M. Namazian, M. Zakery, M. R. Noorbala, M. L. Coote, *Chem. Phys. Lett*., **2008**, *451*, 163–168.

[23] I. E. Charif, S. M. Mekelleche, D. Villemin, N. J. Mora-Diez, *Mol. Struct*. (THEOCHEM), **2007**, *818*, 1–6.

[24] M. D. Liptak, G. C. Shields, *Int. J. Quantum Chem*., **2001**, *85*, 727–741.

[25] M. D. Liptak, G. C. Shields, *J. Am. Chem. Soc*., **2001**, *123*, 7314–7319.

[26] C. O. Silva, E. C. Silva, M. A. C. Nascimento, *J. Phys. Chem. A.*, **2000**, *104*, 2402–2409.

[27] D. M. Chipman, *J. Phys. Chem. A*., **2002**, *106*, 7413–7422.

[28] S. Sastre, R. Casasnovas, F. Munoz, J. Frau, *Theor. Chem. Acc.*, **2013**, *132*, 1310.

[29] A. G. Riojas, A. K. Wilson, *J. Chem. Theory Comput*., **2014**, *10*, 1500–1510.

[30] J. R. Pliego, J. M. Riveros, *J. Phys. Chem*. A., **2002**, *106*, 7434–7439.

[31] J. J. Klicic, R. A. Friesner, W. C. Liu, S. Guida, *J. Phys. Chem. A*., **2002**, *106*, 1327–1335.

[32] K. R. Adam, *J. Phys. Chem. A*., **2002**, *106*, 11963–11972.

[33] K. B. Wiberg, S. Clifford, W. L. Jorgensen, M. J. Frisch, *J. Phys. Chem. A.*, **2000**, *104*, 7625–7628.

[34] W .Sang-Aroon, V. Ruangpornvisuti, *Int. J. Quantum Chem.*, **2008**, *108*, 1181–1188.

[35] J. Ho, A. Klamt, M. L. Coote, *J. Phys. Chem. A.*, **2010**, *114*, 13442–13444.

[36] J. Ho, *Phys. Chem. Chem. Phys*., **2015**, *17*, 2859–2868.

[37] C. R. Sutton, G. V. Franks, G. da Silva, *J. Phys. Chem. B.*, **2012**, *116*, 11999–12006.

[38] A. T. Afaneh, G. Schreckenbach, F. Wang, *J. Phys. Chem. B.*, **2014**, *118*,11271–11283.

[39] V. S. Bryantsev, M. S. Diallo, W. A. Goddard III, *J. Phys. Chem. A.*, **2007**, *111*, 4422–4430.

[40] A. S. Adeyinka, I. Cukrowski, *J. Mol. Model*., **2015**, *21:162*, 1–18.

[41] F. Eckert, M. Diedenhofen, A. Klamt, *Mol. Phys*., **2010**, *108*, 229–241.

[42] A. V. Marenich, W. Ding, C. Cramer, D. G. Truhlar, *J. Phys. Chem. Lett.*, **2012**, *3*, 1437–1442.

[43] A. Klamt, F. Eckert, M. Diedenhofen, M. E. Beck, *J. Phys. Chem. A.*, **2003**, *107*, 9380–9386.

[44] C. P. Kelly, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem. A.*, **2006**, *110*, 2493–2499.

[45] B. Mennucci *J. Phys. Chem. Lett.*, **2010**, *1*, 1666–1674.

[46] J. Tomasi, M. Persico, *Chem. Rev*., **1994**, *7*, 2027–2094.

[47] Spartan'10, version 1.1.0, **2010**, Wavefunction, Inc., 18401 Von Karmen Ave., Suite 370, Irvine, CA92612, USA.

[48] Frisch M. J, Trucks G. W, Schlegel H. B et al., 2009, Gaussian 09, Revision D.1, Gaussian, Inc., Wallingford CT.

[49] A. E. Frisch , **1998**, Gaussian 09 User's Reference, Gaussian Inc., Pittsburgh, PA.

[50] I. Cukrowski, K. K. Govender, *Inorg. Chem.*, **2010**, *49*, 6931–6941.

[51] B. Temelso, G. C. Shields, *J. Chem. Theory Comput*., **2011**, *7*, 2804–2817.

[52] B. Njegic, M. S. Gordon, *J. Chem. Phys*., **2006**, *125:224102*, 1-12.

[53] A. Jinich, D. Rappoport, I. Dunn, B. Sanchez-Lengeling, R. Olivares, E. Amaya Noor, A. Bar Even, A. Aspuru-Guzik, *Sci. Rep*., **2014**, *4:7022*, 1–6. DOI: 10.1038/srep07022.

[54] S. Grimme, *Mol. Sci*., **2011**, *1*, 211–228, DOI i: 10.1002/WCMSW.30.

[55] S. N. Steinmann, G. Csonka, C. Cominboeuf, *J. Chem. Theory Comput*., **2009**, *5*, 2950–2958.

[56] S. Kozuch, S. M. Bachrach, J. M. L. Martin, *J. Phys. Chem. A*., **2014**, *118*(1), 293–303.

[57] A. D. Boese, *ChemPysChem*., **2015**, *16*, 978–985.

Chapter 6

Protonation Sequence of Aliphatic Linear polyamines; A theoretical ^{13}C NMR Study

Summary

Theoretical modelling of the ^{13}C NMR–pH titration curves for the various carbon atoms in *trien* was attempted using a combination of DFT electronic structure method, appropriate conformational search techniques and the GIAO method for computing NMR chemical shifts. A comparison of theoretical plots with experimental ones suggests that a mixture of the two possible monoprotonated forms would most likely co-exist in solution. COSMO solvation model performed better than the PCM solvation model for computing NMR chemical shifts of highly charged polyamine intermediates. Taking the isotropic shift of dioxane as reference gave chemical shifts that were more accurate than using that of TMS which is widely accepted as the reference molecule of choice for computing NMR shifts. Our results also show that in the absence of appropriate scaling factors, the use of external reference results in a significant improvement in the accuracy of computed ^{13}C NMR shifts.

6.1. Introduction

Ever since polyamines were discovered in 1678 by Antoni van Leeuwenhoek, there has been increasing interest and intense research efforts aimed at understanding and elucidating their chemical behaviour and the mechanism responsible for the effects they have on biological processes in living systems – both plants and animals.[1–8] Among the various biological functions attributed to them, their protonation sequence has been the subject of great attention since it is a well-known fact that their interaction with DNA and RNA is dependent on their proton distribution patterns. Hence several experimental and theoretical techniques has been applied in order to determine their protonation sequence but until now there exists in literature, a controversy about the preferred site of protonation particularly for monoprotonated forms.[8–24] Furthermore the difficulty encountered up to date in isolating the singly protonated form by some physical techniques such as X-ray crystallography[25,26] has made a conclusive resolution of this situation almost if not impossible.

NMR spectroscopy is one of the most promising methods to resolve this controversy experimentally because ^{1}H, ^{13}C and ^{15}N chemical shifts are all sensitive to the degree of protonation of individual basic sites in any polybasic molecule.[9,14–16] However, using NMR technique as a tool is not devoid of challenges such as the similarity in the behaviour of α-methylene protons in LAPs which makes the resolution of individual changes in their ^{1}H NMR shifts impossible without the aid of ^{1}H–^{13}C two-dimensional correlation NMR spectroscopy.[15] Secondly, the need for isotopic enrichment or a very high sample concentration places a severe limitation on the use of ^{15}N NMR even though it would have been the most effective of these methods.[15] Consequently among all the possible types of NMR spectroscopic techniques, ^{13}C NMR technique is the one that has been mostly used to study protonation sequences of LAP in literature because of its better spectral resolution when compared to ^{1}H NMR.

However another difficulty arises as a result of the fact that at any given pH, all the probable protonated forms of a given polyamines coexist in solution in various proportions, hence it is not straightforward to interpret experimental spectra for a given polyamine without the introduction of certain theoretical assumptions. Specifically, the use of nonlinear least square analysis involving the fitting of up to fifteen parameters in some cases to analyse experimental NMR data casts a doubt as to the validity of conclusions from such studies.[22] All the above mentioned factors weakens the validity of the conclusions reached based on various experimental NMR studies even though there is some measure of agreement in their conclusions.[14–16]

Computing NMR shifts is a fairly routine process nowadays; hence in this work, we decided

to investigate the protonation sequence of aliphatic linear polyamines from a computational perspective. Our aim in this work was to obtain directly, the ^{13}C NMR chemical shifts of the various carbon atoms in the various protonated intermediates (and tautomeric forms) of *trien* (or 2-2-2-tet), generate its ^{13}C NMR–pH titrations curves using this data and then compare theoretical ^{13}C NMR–pH titrations curves with experimental ones in order to determine its protonation sequence in solution. This approach differs from what has been done previously where chemical shifts of carbon atoms in the various protonated intermediate forms of polyamines (i.e. **HL**, $\mathbf{H_2L}$, $\mathbf{H_3L}$) were obtained from linear least square regression analysis of experimental NMR data because they cannot be accessed experimentally. Furthermore various parameterization schemes are used to estimate the chemical shifts of carbon atoms in each possible tautomeric form for each of partially protonated species (e.g. $\mathbf{HL_p}$ and $\mathbf{HL_s}$ forms of **HL** for example). The most popular of these schemes is the one by Sarneski and co-workers which utilizes data from experimental NMR shifts of model compounds such as monoamines to determine empirical parameters that can be used to calculate the chemical shifts of carbon atoms in partially protonated species of polyamines. The average fraction of time each basic site is protonated (i.e. Proton distribution pattern) can then be determined from all these data using equations of the following general form[9,14]:

$$\Delta\delta_i^c = \sum_{j=1}^{N} C_{ij} f_j \qquad (6.1)$$

Where $\Delta\delta_i^c$ is the total calculated protonation shift (change in chemical shift value caused by protonation) of the *i*th methylene carbon resonance, C_{ij} is the protonation shift constant of the *i*th resonance for the total protonation of the *j*th basic site and f_j is the average fraction of time the *j*th site is protonated. The use of such equations to determine protonation sequence is based on two assumptions; (i) the contribution of a given basic site to the change in chemical shift of an adjacent carbon atom is linearly related to the fraction of time that basic site is protonated and (ii) second, that the contributions due to the protonations of different sites are perfectly additive. The first assumption is based on the time-averaging effect of rapid proton exchange which has been well substantiated by Grunwald, Loewenstein and Meiboom.[9] Similarly, Shoolery and co-workers have demonstrated that the second assumption is reasonably applicable to methylenic proton chemical shifts.[9] A protonation sequence is then derived using the values of f_j obtained as the solutions of such equations. However, such schemes do not account for conformational effects on the chemical shifts of carbon atoms in polyamines and as such may be neglecting its effect on proton distribution pattern. Hague and co-workers have developed amine shift

parameters from the experimental NMR shift data of a number of polyamines which they studied. These have been more successful in predicting the chemical shifts of carbon atoms in aliphatic polyamines. In their approach, the chemical shift of each carbon atom is determined principally by the types (α,β or δ) of the two nearest amino groups to it and their protonated state which is designated as π for the unprotonated amino group or π^+ for the protonated amino group. It has however been pointed out that conformational effects in a molecule such as polyamines might be very important in determining the chemical shift of its constituent nuclei.[26] Hence it was noted that considerable caution should be exercised when using such empirical protonation shifts to determine the protonation sequence of polyamines. More importantly, protonation shift obtained from monoamines may not provide a reliable estimate because (among other reasons) polyamines may form hydrogen bonded species in solution. This can alter the protonation shifts experienced by carbon atoms in polyamines due to resultant conformational changes in the structure of polyamines.

Our aim in this work was to completely eliminate the use of empirical parameters for determining the chemical shifts and protonation shifts of carbon atoms in polyamines. Rather, using the conformational search protocol (CSP) developed previously[27], low energy conformers (LECs) and tautomers of each partially protonated form, the free ligand and fully protonated forms were generated and the ^{13}C chemical shifts of their carbon atoms computed directly. This kind of approach should account adequately for any conformational and hydrogen bonding effects on chemical shifts usually neglected by the use of empirical methods. The neglect of all these subtle but important factors is the reason why cross terms (terms which arises due to the interaction between two or more protonated base sites in any given polyamine) must be subtracted in certain cases from estimated chemical shifts of partially protonated species obtained from parameterization schemes.

Trien was selected as a prototype polyamine molecule based on the observation that the basicity of nitrogen atoms in symmetrical polyamines is of the same order of magnitude hence making it difficult to determine a priori, their preferred site of protonation especially in the first step. Furthermore, the nitrogen atoms in *trien* are separated by the smallest possible ethylene chain ($–CH_2–CH_2–$) thus making this difference in basicity of its primary and secondary nitrogen atoms even smaller. This therefore makes it a suitable model compound to carry out a computational investigation of this kind. The kind of theoretical approach used in this study should make it possible to verify whether only one or a mixture of the two possible tautomeric forms ($\mathbf{HL_p}$ and $\mathbf{HL_s}$) of monoprotonated *trien* is formed in solution when a single equivalent of acid is added.

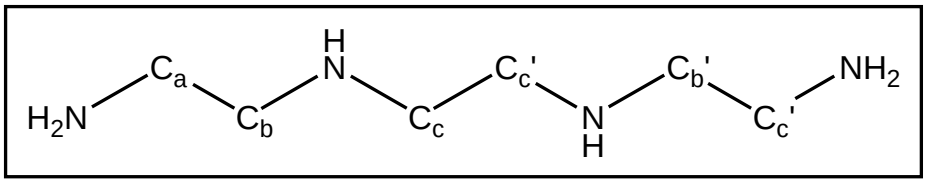


Figure 6.1. Structure of triethylenetetramine showing carbon atoms labelling.

In addition, being able to theoretically predict the variation in chemical shift of various carbon atoms as a function of pH might serve as a reliable aid for proper interpretation of the various experimental NMR–pH titration data acquired for several polyamines over the years.

6.2 Computational Methods

As described previously[27], conformational analysis was done using molecular mechanics and DFT methods in order to identify representative LECs of all H_nL^{n+} forms of *trien*. The search was carried out with and without explicit water molecules. All the conformers retained initially from the MM-based search were fully energy optimized in solvent. Subsequently, isotropic shielding constants of carbon atoms in conformers having an energy within 3kcal energetic window of the lowest energy conformer (LEC) were computed at the B3LYP/6-311++G(d,p) level of theory using the GIAO method for relevant tautomers at thermodynamic equilibrium. The isotropic shielding constants of carbon atoms in dioxane were used as reference in our calculations of NMR shifts since it has been demonstrated that using alternative reference compounds with nuclei having similar characteristics as that in the molecule of interest results in a better error cancellation and consequently more accurate chemical shift values.[28] Furthermore almost all previous experimental NMR spectra of aliphatic linear polyamines were recorded using dioxane as an internal standard. Hence this approach is analogous to recording experimental spectra with the aid of an internal standard. This has been shown to improve the accuracy of computed NMR shifts significantly.[28] All electronic structure calculations were performed with the aid of Gaussian 09, revision B01.[8] Gauss View 4.1.2[9] was utilized as a molecular builder and for visualization. All structures have been fully optimized without constraints at the RB3LYP/6-311++G(d,p) level of theory. Calculations in ADF were carried out using the B3LYP DFT, aug-ATZP basis set and COSMO solvation model.

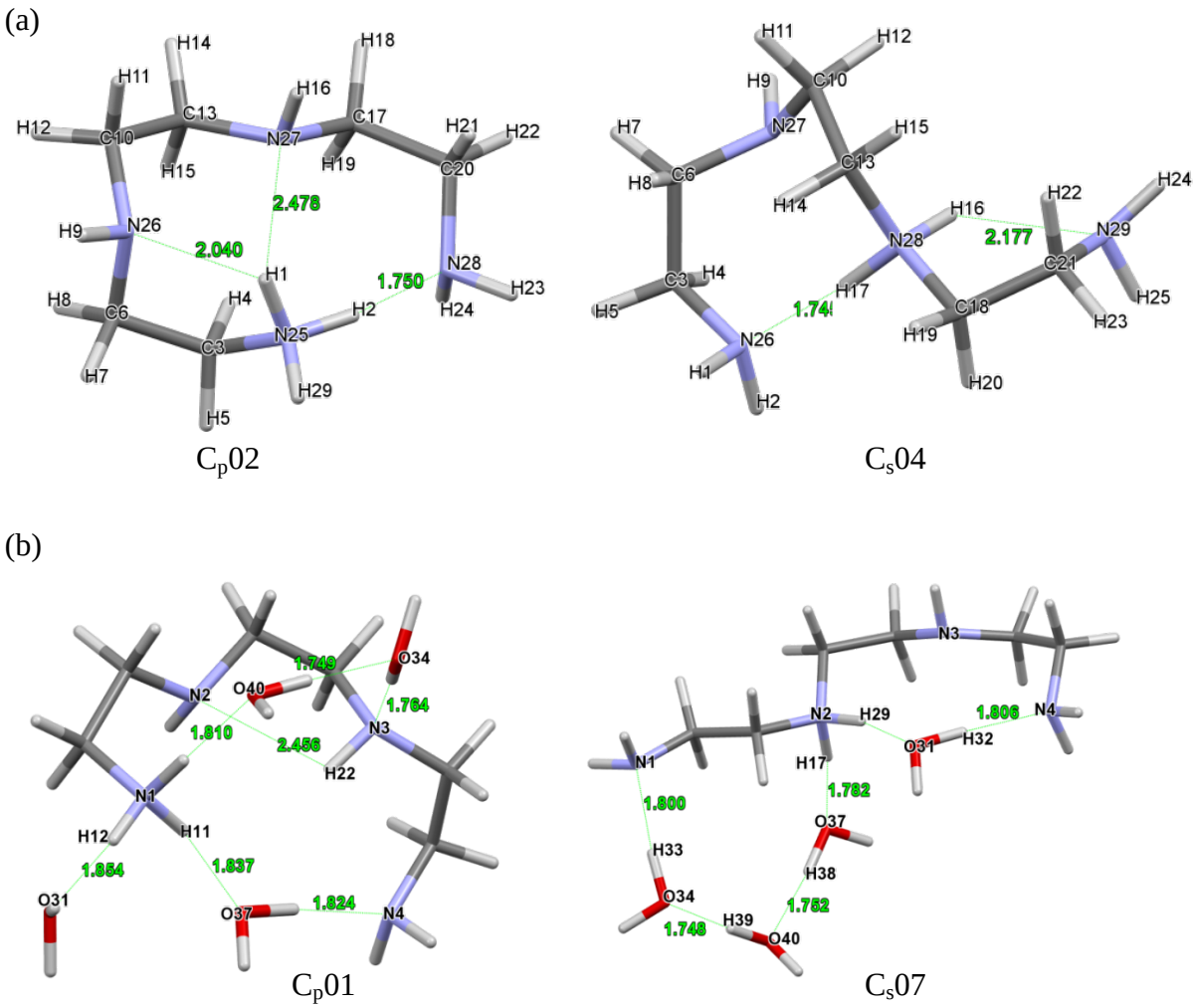
6.3. Results and Discussion

6.3.1. G(aq) and the protonation sequence in solvent

In the first step of protonation, *trien* has two possible tautomeric forms, $\mathbf{HL_p}$ and $\mathbf{HL_s}$ in which the primary and secondary nitrogen atoms are involved in protonation respectively.

Optimized structures of LEC for these two possible monoprotonated forms with and without explicit solvent molecules is shown in Figure 6.2. From the analysis of their Gibbs free energy data obtained from electronic structure and frequency calculations, and based on the assumption that at thermodynamic equilibrium the lowest energy form of H_nL^{n+} will be predominant in solution, , the ratio of the $\mathbf{HL_p}$: $\mathbf{HL_s}$ form is ~ 86 to 14% in the presence of four explicit water molecules. This suggests clearly that the primary nitrogen atom of trien will be preferentially protonated and predominant in solution in the first stage of protonation. This is in contrast to the results obtained in our previous study[27] where PCM solvation model alone was utilized and it was found that $\mathbf{HL_s}$ form is far more favoured energetically over the $\mathbf{HL_p}$ form. Similarly for the diprotonated form of *trien* our prediction with hybrid solvation model (we found a ratio of 23:76% for $\mathbf{H_2L_{ps}}$:$\mathbf{H_2L_{pp}}$) contradicts that based on the PCM solvation model alone (which predicts the formation of only $\mathbf{H_2L_{ps}}$ tautomer). The result obtained here for the diprotonated form is in close agreement with the general consensus in literature (i.e. $\mathbf{H_2L_{pp}}$ form where the two protons are separated by the maximum possible distance is generally accepted as the preferred tautomer). In comparison to other methods, the cluster expansion technique[29–30] seems to be the most consistent and reliable means of predicting the protonation sequence of polyamines from experimental ^{13}C NMR data. This is because of its underlying idea that, if we know the free energy of protonation for a specific basic site or combination of sites in a polyamine, then we can obtain directly, microscopic protonation constants and relative %-contributions of all possible microspecies of the polyamine. Knowledge of these makes it possible for one to determine their proton sequence directly.

However since these species cannot be isolated experimentally talk less of measuring their free energy, Borkovec and co-workers utilized the principles of statistical mechanics to develop a means of parameterizing their microscopic protonation constants. Both Gibbs free energy and % contribution of microspecies could then be obtained from the microscopic protonation constants. It is important to point out that as theoretically sound as their approach seems, it is totally dependent on data from experimental techniques such as NMR or macroscopic titration. Because it is dependent on being able to evaluate Gibbs free energies of all possible microspecies, it is therefore not entirely unexpected that our results are in excellent agreement with theirs.[30] Hence our result demonstrates that, using a representative set of low energy conformers and appropriate solvation models in conjunction with a high level electronic structure method, it is possible to predict entirely from first principles, the protonation sequence of any given aliphatic linear polyamine including those that are yet to be discovered/synthesized.



C_p02 C_s04

C_p01 C_s07

Figure 6.2. Optimized lowest energy conformers of HL_p and HL_s forms of trien respectively, in part (a) PCM solvation model and (b) with explicit water molecules and PCM solvation

This should make the preliminary assessment of such molecules for their suitability in medicinal chemistry applications more accessible and cost effective. More importantly, this kind of approach will also eliminate the need for parameterization while also taking care of conformational and structural effects that were not explicitly accounted for with such schemes. There has been an attempt to use this approach for predicting microscopic constants (and consequently protonation sequence) of spermine.[31] But the authors had to empirically scale their predicted micro protonation constants in order to reach close agreement with experimental values of macroconstants obtained by combining predicted microconstants in the usual way to obtain macroconstants measured from experiment. This might be due to errors inherent in their conformational search method or solvation model utilized in their study.

Therefore, it might be worthwhile to carry out a more extensive study to determine the appropriate solvation model and cost-effective electronic structure method that would be

sufficient for an accurate determination of the relative energy of all possible microspecies for aliphatic linear polyamines, hence their protonation sequence. If such a theoretical model chemistry can be "discovered", it would be the best means of predicting the protonation sequence of polyamines in general. This is due to its inherent advantage of being a direct method in contrast to all other indirect methods where information about protonation sequence is obtained from a nonlinear least square regression analysis of experimental ^{13}C NMR data.

6.3.2. Analysis of ^{13}C NMR Shifts

Having obtained the absolute isotropic shielding constants for both *trien* and dioxane we calculated the NMR chemical shifts of the various carbon atoms in trien relative to TMS using the following expression:

$$\delta_i = \sigma_{ref} - \sigma_i + \delta_{ref} \qquad (6.2)$$

where δ_i is the chemical shift relative to TMS for each nucleus in the molecule of interest

σ_i is the computed shielding constants for each nucleus in the molecule of interest and

σ_{ref} is the computed isotropic shielding constants computed for the same type of nucleus in reference compound.

δ_{ref} is the known experimental chemical shift for the reference compound.

By averaging the chemical shift of each carbon atom in a given tautomer over all relevant conformations (those contributing ≥ 5% according to their relative Boltzmann distribution population), we were able to calculate its overall chemical shift in that tautomeric form. It should be noted that due to symmetry constraints, the chemical shifts of symmetrical carbon atoms were averaged in this work. From this information we obtained theoretical chemical shift of a carbon atoms in any partially protonated form H_nL^{n+} of *trien* by taking a weighted average of its shift in all relevant tautomers (according to various suggested relative percentage composition in literature). Finally, we used species distribution data which can be obtained from either experimental or theoretical stepwise protonation constants reported elsewhere[32–34] to calculate chemical shift of the *k*th carbon atom in *trien* at any given pH using the following expression:

$$\delta_k = \sum_{i=0}^{N} f_i \delta_k^i \qquad (6.3)$$

where δ_k^i is the intrinsic chemical shift for the *k*th signal and f_i is the fraction of the *i* protonated species. The summation is over all protonated forms *N* in which the polyamine exists in solution.

To determine preferred protonation sites in *trien*, we generated a theoretical plot of its ^{13}C NMR shifts vs. pH and compared it to that obtained from our own experimental NMR data.

In a series of papers, D.N Hague et al have tried to explain protonation sequence for LAPs by comparing experimental ^{13}C NMR shifts to those calculated from empirical parameters derived from experimental NMR measurements of model compounds. In their study, they assumed various possible compositions of relevant tautomers (e.g. $\mathbf{HL_p}$ and $\mathbf{HL_s}$) that may determine the observed experimental chemical shift of the carbon atoms in the singly protonated form of a given polyamine. Using amine shift parameters, they computed the chemical shifts that would be observed for carbon atoms in the **HL** form resulting from different compositions proposed. They suggested that the composition whose resulting chemical shifts deviates the least from experimental values indicates the protonation sequence in solution. These authors however pointed out that strictly speaking, their approach is ambiguous in the sense that in some instances it might be difficult to discriminate between a 2:1 or 1:1 ratio of $\mathbf{HL_p}$ and $\mathbf{HL_s}$ tautomers if they give the same average computed chemical shift. Our approach in this work even though different, closely follows their strategy. In computing the chemical shifts of the singly protonated form of *trien* for example, we used various percentage ratios of the $\mathbf{HL_p}$ and $\mathbf{HL_s}$ form suggested in literature (e.g. we tried a. 1:1 and 2:1 mixture of $\mathbf{HL_p}$ to $\mathbf{HL_s}$ tautomers respectively). We kept the percentage ratio of H_2L forms constant since it is widely accepted[9, 11–22] that the $\mathbf{H_2L_{pp}}$ form is preferred due to the minimum electrostatic repulsion between charged groups resulting from its formation as opposed to the $\mathbf{H_2L_{ps}}$ form. Our hypothesis was that the ratio of these tautomeric forms that best reproduces the change in ^{13}C NMR shifts as a function of pH as that which we observed from experiment is the one that actually represents their true abundance in solution and consequently we can determine their protonation sequence from such information.

Deviations from experiment in the computed ^{13}C NMR shift of carbon atoms in the various protonated forms of *trien* using either the PCM or COSMO solvation model is shown in Table 6.1. A comparison of mean absolute deviation of chemical shifts obtained using these two solvation models reveals that COSMO is a better solvation model for computing NMR chemical shifts of polyamines. This is especially true for highly charged states of polyamines (e.g. H_3L and H_4L species) involved in this study and could be due to a better description of the solvation of a charged solute by COSMO model. As pointed out earlier in our discussion, this observation also indicates the need to carry out a more extensive investigation of the effect of solvation model on the energetic and geometrical properties of polyamines. This will go a long way in the development of a direct protocol for obtaining the free energies and consequently, the microscopic equilibrium constants for the protonation of specific basic sites (and or combination

of sites) in aliphatic linear polyamines. Analysis of data in Table 6.1 also reveals that using another reference molecule other than TMS, containing similar nuclei as the molecule of interest in theoretical evaluation of chemical shifts improves the accuracy of chemical shifts prediction significantly. The huge error in computed NMR shifts when the isotropic shifts of TMS is used as reference value and a density functional theory method utilized for computing chemical shifts has been attributed to the neglect of relativistic effects on the electronic structure of silicon which then affects the chemical shifts of the carbon atoms attached to it.[28] Hence in the absence of appropriate scaling factors, this approach is highly recommended as a means of improving the accuracy of computed chemical shifts. This is clearly pointed out in the excellent review by Tantillo et al.[28] on computing theoretical chemical shifts.

Table 6.1. Computed NMR shifts for carbon atoms in L, H_3L and H_4L forms of *trien* using a discrete-continuum solvation model (DCSM)

H_nL^{n+}	C-atom	Exp.	$\Delta\delta_{PCM}$	$\Delta\delta^{a}_{COSMO}$	$\Delta\delta^{b}_{COSMO}$
	C_a	41.00	3.42	2.77	7.75
L	C_b	51.80	2.55	2.06	6.89
	C_c	48.80	3.23	2.79	7.57
	C_a	38.20	3.53	1.88	7.87
H_3L	C_b	45.90	1.49	0.36	5.83
	C_c	47.00	1.31	0.45	5.65
	C_a	36.50	4.08	1.97	8.42
H_4L	C_b	45.90	2.38	1.23	6.72
	C_c	44.20	3.37	1.54	7.70
		MAD	***2.82***	***1.67***	***7.15***

[a]Dioxane used as reference, [b]TMS used as reference.

Comparison of experimental NMR–pH titration curves with theoretical plots obtained using LECs generated with four explicit water molecules (Figure 6.4) indicates clearly that the shape of titration curves for individual carbon atoms in our own theoretical plots differs from that of experiment. Also all the various theoretical NMR titration curves look qualitatively the same in shape and trend followed by chemical shifts of carbon atoms. This is irrespective of the variation in the relative percentage composition of $\mathbf{HL_p}$ and $\mathbf{HL_s}$ tautomers employed in making these plots hence one cannot visually identify any significant difference between these plots. This implies that it might be difficult to decide theoretically whether HL_{N1} or HL_{N2} is the preferred monoprotonated form. Furthermore, the shape of the theoretical NMR titration curves only matches that of experiment in the pH ranges of 0–4 and 10–14 (see Figure. 6.3 and also Figure. D1 of Appendix D). Between pH 4 and 10, there is a drastic difference in the trend followed by theoretical chemical shifts of carbon atoms as compared to experiment. In contrast, theoretical

^{13}C NMR–pH titration curve plots for the fully linear conformers of *trien* shown in Figure 6.5 surprisingly follows the trend and shape of experimental ones despite the fact that the mean absolute deviation of chemical shifts of various H_nL^{n+} forms of the fully linear conformers from experimental values is significantly larger than those of the LECs generated in the presence of explicit water molecules.

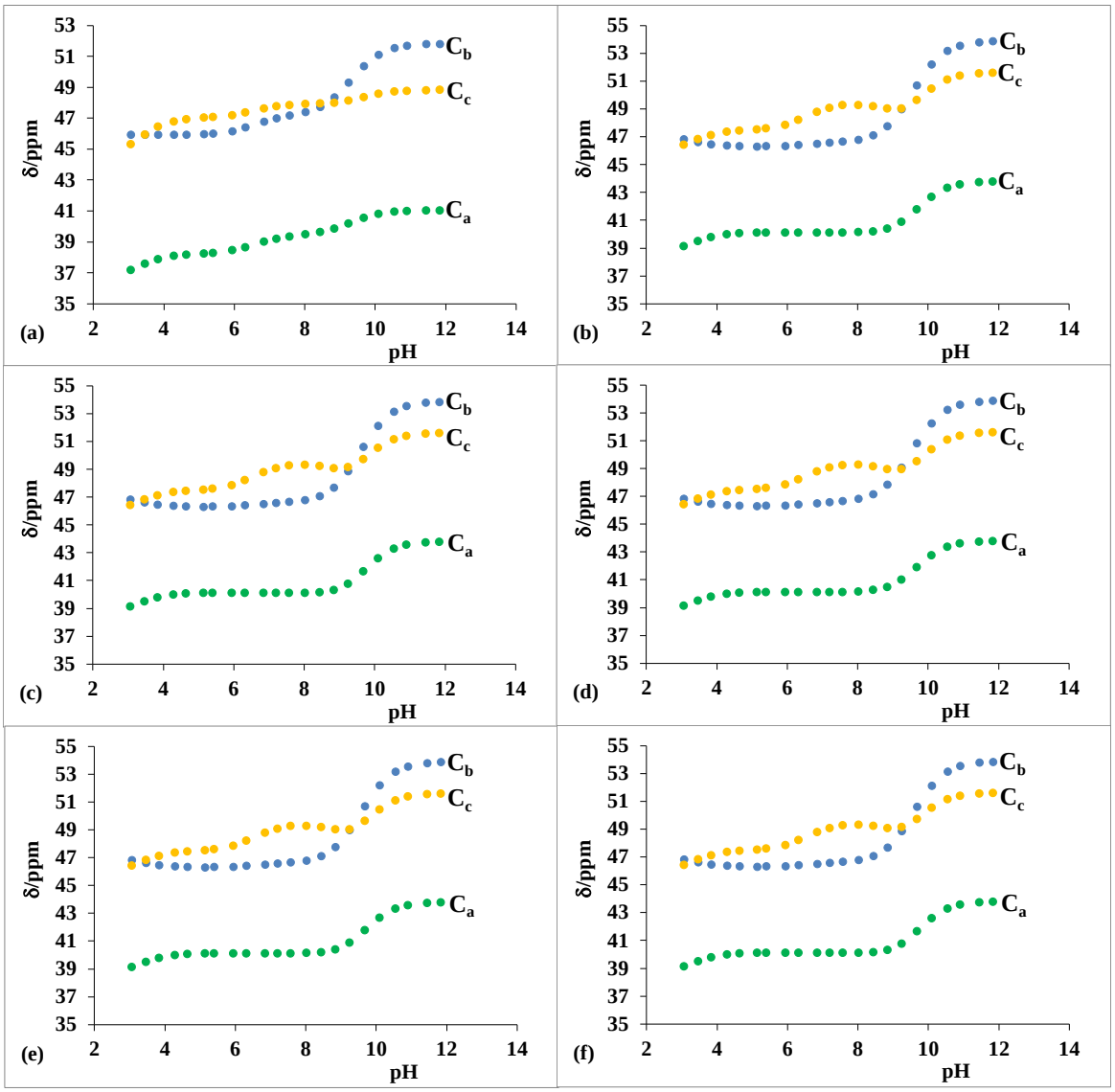


Figure 6.3. ^{13}C NMR chemical shifts of *trien* as a function of pH (a) experiment (b) HL_p: HL_s (1:1) (c) HL_p: HL_s (2:1) (d) HL_p: HL_s (1:2) (e) HL_p only (f) HL_s only. using low energy conformers, B3LYP/4H2O+COSMO.

An in-depth examination and comparison of the geometry of LECs generated in the presence of explicit solvent molecules to the fully linear conformers revealed that there is a huge variation in the chemical shifts of equivalent carbon atoms in the explicit solvation situation contrary to when we use fully linear conformers to compute ^{13}C NMR chemical shifts. This is especially true between pH 5–10 where the H_2L (in particular H_2L_{pp} form), H_3L and H_4L forms are the

predominant protonated species in solution. This coincides with pH the region where we observed significant deviation from experimental trends in the shape and trend of theoretical ^{13}C NMR–pH titration curves (see Figure 6.4.). Our quest to understand the origin of this unexpected result showed that it is a consequence of non-uniform distribution of explicit water molecules around protonated polyamine molecules in such cases. This resulted in huge differences in the chemical environment of the two equivalent carbon atoms (e.g. C_a and C_a'according to symmetry constraints) of those H_nL^{n+} forms when explicit solvation was utilized (see Figure D3–D4 in Appendix D). The carbon atom bonded to a protonated nitrogen atom which has more water molecules interacting with it will experience a deshielding effect due to the fact that the water molecules will disperse the charge on the protonated nitrogen atom and this will in turn reduce the influence of protonation on the chemical shift of such carbon atom i.e. it will be artificially more deshielded hence its chemical shift will shift downfield i.e. high δ values. The converse is true for a carbon atom attached to a nitrogen atom which artificially interacts with lesser number of explicit water molecules than its equivalent counterpart. All these results in significantly larger deviation of the chemical shifts of carbon atoms C_a, C_b and C_c form experiment particularly in the H_3L and H_4L forms which one would intuitively expect to be more linear structurally due to the presence of multiple positive charges within the molecule. This is however not the case when fully linear conformers are used in an implicit aqueous environment hence there is smaller difference between the chemical shifts of equivalent carbon atoms in such molecules due to a uniform solvation environment. The implication of this is that the structure of H_3L and H_4L LECs generated in the presence of explicit water molecules differs significantly from their real structure in solution while those of the fully linear forms of these H_nL^{n+} forms matches their real geometry in solution hence a better prediction of ^{13}C NMR values was observed using these conformers.

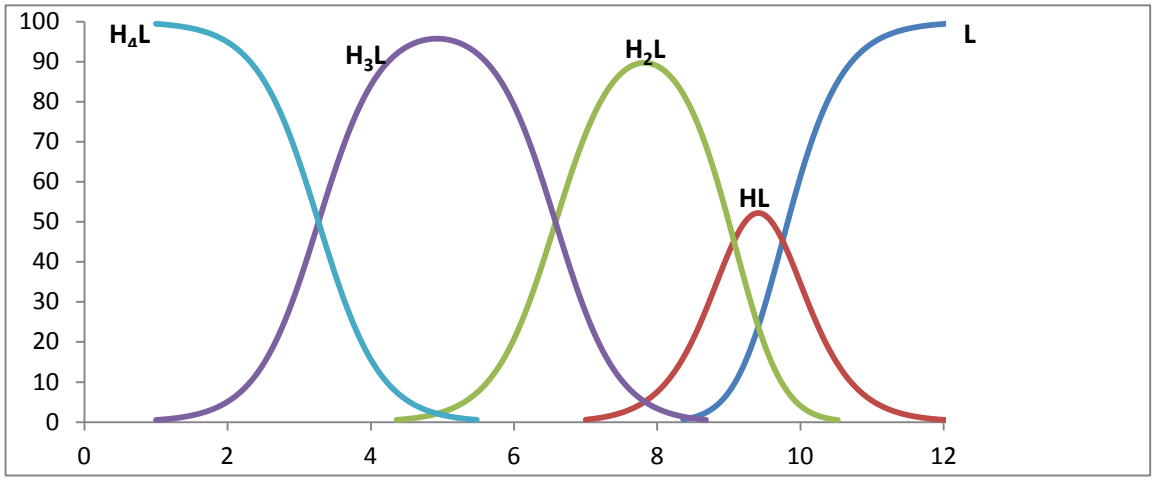


Figure 6.4. Species distribution diagram of trien between pH 0-12

It is important to reiterate however that, in contrast to results obtained using LECs with four explicit molecules, we could not predict relative percentages of each tautomeric HL form in agreement with previous reports by comparing *G*(aq) values of fully linear HL_{N1} and HL_{N2} tautomeric forms when these structures were optimized either using only the PCM solvation model or in the presence of explicit solvent molecules.

Previous studies suggested that fully protonated polyamines exist in the fully linear conformation in aqueous solution due to the possibility of hydrogen bonding interactions with water molecules when protonated.[6–8] Hence, the reason why experimental trends in the change of ^{13}C NMR shifts as a function of pH was not reproduced with LECs in the presence of explicit solvent could be that, although our optimization with four water molecules gave representative conformers of tautomeric structures with relative free energies that was accurate enough to predict their relative distribution in solution in better agreement with previous reports , the geometries of structures generated in the case of H_3L and H_4L does not actually match exactly the one adopted by polyamines in bulk solution. Hence as pointed out earlier, one may have to increase the number of explicit water molecules used to represent the first solvation shell of highly charged forms e.g. H_3L and H_4L in order to ensure uniform charge distribution and consequently better accuracy of predicted ^{13}C NMR shifts in such situations. In agreement with observation for theoretical NMR titration curves obtained using chemical shifts of low energy conformers, there is no noticeable difference between the various theoretical plots using fully linear conformers despite variation in percentage of $\mathbf{HL_p}$ and $\mathbf{HL_s}$ tautomers used in the different plots. A closer examination however shows that the NMR titration curve obtained when it is assumed that only the secondary nitrogen atom is protonated in the monoprotonated form of *trien* is different from the curve obtained with all other compositions (see in Figure 6.5, and also Figure D2 of Appendix D, note the trend for C_c in this plot as compared to other theoretical plots). This suggests that the $\mathbf{HL_s}$ form is not the only monoprotonated form in solution. The only conclusion that can be safely reached from an examination of these plots is that both monoprotonated forms most likely co-exist in solution. However it is impossible to ascertain their relative proportions as there is hardly any noticeable change in the various theoretical plots when their relative ratio is varied.

A totally different scenario plays out when we assumed that both **HL** forms are present in solution in 1:1 ratio and then varied the relative proportions of the two most likely diprotonated

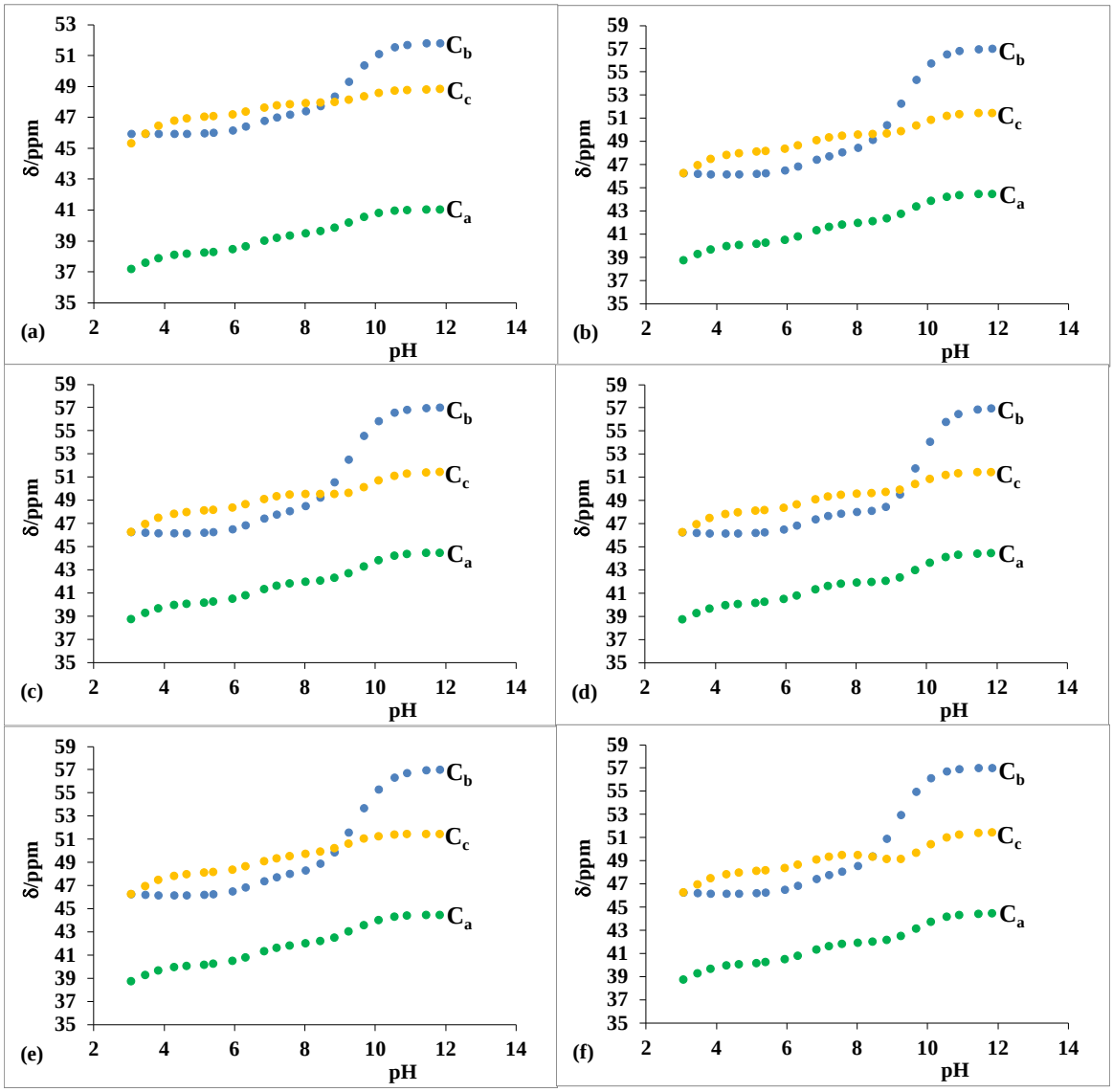


Figure 6.5. ^{13}C NMR chemical shifts of *trien* as a function of pH (a) experiment (b) HL_p: HL_s (1:1) (c) HL_p: HL_s (2:1) (d) HL_p: HL_s (1:2) (e) HL_p only (f) HL_s only, using fully linear conformers, B3LYP/COSMO.

forms that should exist in solution (**H_2L_{ps}** and **H_2L_{pp}**). Figure. 6.6 reveals that as the relative proportions of these diprotonated forms are varied there is a noticeable change in the shapes of the theoretical NMR titration curves. The theoretical plot that closely follows experimental trend of change in chemical shifts as a function of pH is one in which the proportion of **H_2L_{ps}** to **H_2L_{pp}** form is 1:2. This is in good agreement with majority of previous studies where it has been emphasized the **H_2L_{pp}** form is preferred in solution. The implication of such conclusion is that only the primary nitrogen atoms are involved in the first stage of protonation.

Only few authors[15,30] have suggested that both of these **H_2L** forms will coexist in solution. Although it is difficult, if not impossible to determine the exact proportions of the various

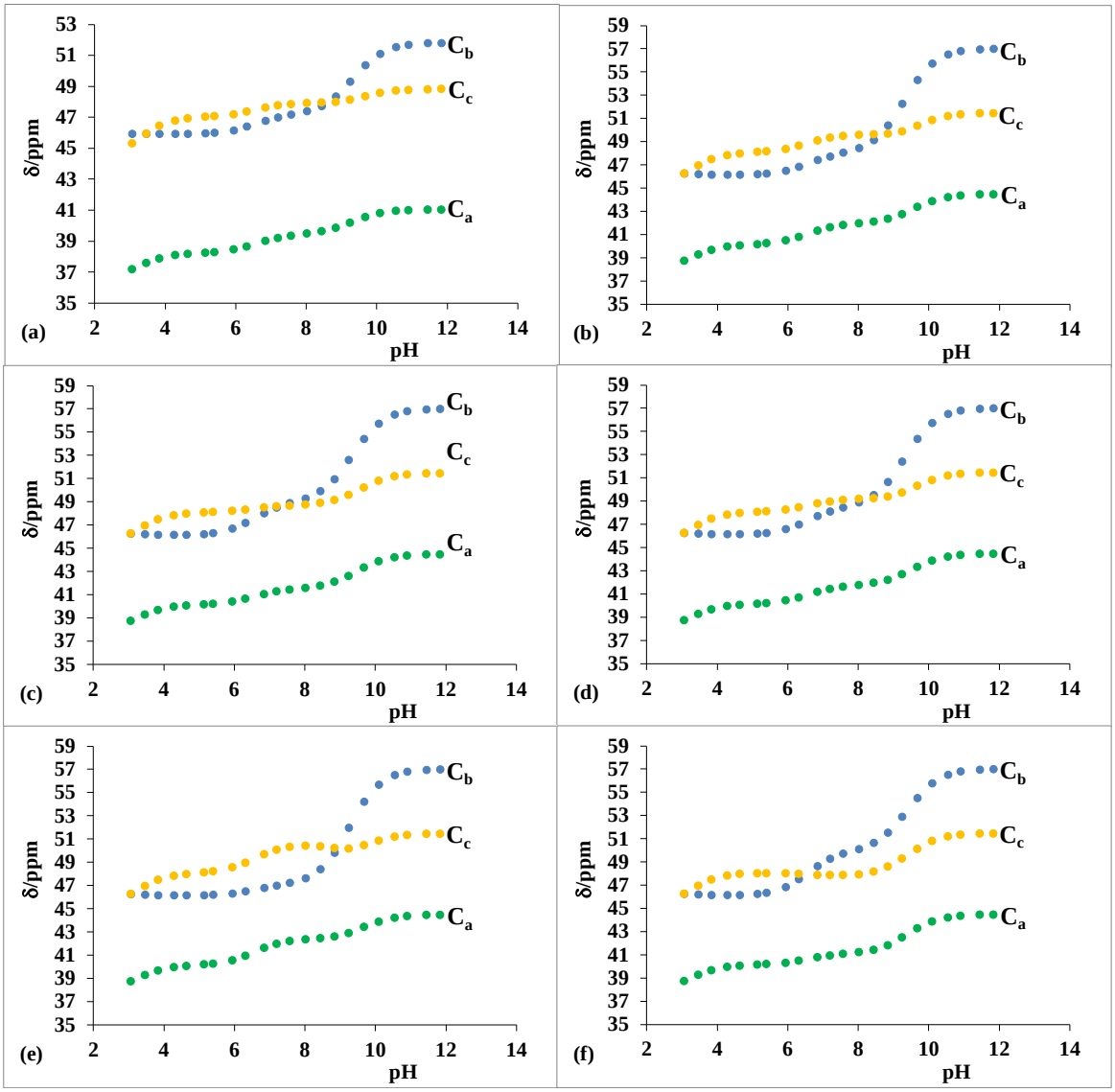


Figure 6.6. ^{13}C NMR chemical shifts of *trien* as a function of pH (a) experiment (b) H_2L_{ps}: H_2L_{pp} (1:2) (c) H_2L_{ps}: H_2L_{pp} (2:1) (d) H_2L_{ps}: H_2L_{pp} (1:1) (e) H_2L_{pp} only (f) H_2L_{ps} only, using fully linear conformers. B3LYP/COSMO.

protonated forms relative to one another from this kind of analysis, our results suggests that both primary and nitrogen atoms are involved in the first and second stage of protonation. Our approach also reveals that the diprotonated forms of *trien* are more abundant than the monoprotonated forms in solution hence the shape of the theoretical NMR titration curves can be used with confidence as an indicator of the relative proportions of the diprotonated forms. The same thing cannot be said of the monoprotonated forms. Because they are only present as minor species in solution, information about their relative proportions in obtained from theoretical NMR plots or analysis of their experimental NMR data may not be reliable. But one can at least infer that since a mixture of both diprotonated forms (H_2L_{ps} and H_2L_{pp}) gave a better representation of experimental observation, then both primary and secondary nitrogen atoms must be involved in the first stage of protonation.

6.3.3. Estimating degree of protonation from theoretical NMR shifts.

As stated previously, it is possible to determine the degree of protonation of various basic sites in a molecule from its NMR data using equation (6.1) where $\Delta\delta_i^c$ is the total change in chemical shift value caused by protonation of the *i*th methylene carbon resonance, C_{ij} is the protonation shift constant of the *i*th resonance for the total protonation of the *j*th basic site and f_j is the average fraction of time the *j*th site is protonated. This kind of expression can be written for various carbon atoms C_a, C_b, and C_c in singly protonated form of trien and the sets of equations solved simultaneously to obtain the degree of protonation of each of the basic sites (N_p and N_s). This kind of approach has been used with experimental ^{13}C NMR data to obtain the %-fraction of time (f_j) a proton spends at each basic site for monoprotonated triamines. Typically, this kind of approach proceeds as follows: Firstly since one cannot obtain the chemical shift of particular carbon atoms in each partially protonated form of polyamines, one has to find a way of obtaining this values. Previous experimental studies have utilized linear least square regression analysis to accomplish this. Using equation (6.3) that shows how chemical shift of each carbon atom varies as a function of pH, one can write a series of such equations for all unique carbon atoms in a given polyamine at each pH where ^{13}C NMR data was recorded. Subsequently, a linear least square minimization method is used to estimate the chemical shifts of HL, H_2L, H_3L and H_4L forms of the polyamine. From these values, it is then possible to calculate $\Delta\delta_i^c$ for the addition of proton(s). Furthermore C_{ij} values are obtained as the difference between experimental chemical shift of a carbon atom with similar chemical environment as C_a in the unprotonated and protonated form of a model monoamine. To obtain values of C_{ij} for the protonation of primary or secondary nitrogen atoms, one has to use a monoamine with primary and secondary nitrogen atoms respectively. Infact in certain cases where there is no monoamine compound that has carbon atoms similar to the polyamine under investigation, the most similar monoamine is normally used to obtain approximate C_{ij} values which were used in equation (6.1). In case of the diprotonated forms, the model compounds used were diamines. Having obtained C_{ij} values for each carbon atom in a given polyamine of interest, a set of equations is then formulated and solved simultaneously to obtain f_i and f_j values, from which one can ultimately determine the protonation sequence of the polyamine. Due to the series of approximation made and the fact that in several cases there is no available model compound to estimate C_{ij} values, the sum of f_i and f_j deviated a little bit from unity for monoprotonated triamines in all cases.[9,14]

Since we now have a means of computing the ^{13}C NMR chemical shifts of carbon atoms in various conformers and tautomers of **HL**, **H_2L**, and **H_3L** forms of aliphatic polyamines

theoretically, this should able to eliminate the errors associated with the use of approximate C_{ij} values in these equation when computing f_i and f_j values using the approach described above. But we would still have to rely on application of linear least square minimization techniques to experimental ^{13}C NMR shift over a given pH range in order calculate the chemical shift of **HL**, $\mathbf{H_2L}$, and $\mathbf{H_3L}$ forms from which we can then calculate the degree of protonation of each particular basic site in the molecule using the expression described above. In essence being able to compute theoretical chemical shifts eliminates the need to also obtain experimental ^{13}C NMR shifts for model compounds so as to obtain C_{ij} values. This definitely results in a huge simplification of the task of obtaining f_i values and might likely result in better estimation of the degree of protonation of each basic site.

Consequently, we tried to use this kind of approach to determine the degree of protonation at individual basic sites of *trien* and compared our results with what has been reported previously. Unfortunately due to the symmetrical nature of *trien*, the simultaneous equations that resulted (i.e. 3 sets of equation with two unknowns) did not have unique solutions and as such we would proceed further. As an alternative, we have also explored various models that describe the relationship between experimental chemical shifts and our theoretically computed chemical shifts as a means of calculating the relative % fractions of the $\mathbf{HL_p}$ and $\mathbf{HL_s}$ monoprotonated forms using linear least square regression analysis but preliminary statistical analysis reveals that there are no unique solutions due to the high degree of correlation between these two monoprotonated forms. This is not entirely surprising as %-fraction of $\mathbf{HL_p}$ and $\mathbf{HL_s}$ is pH independent. Therefore, without making certain statistical assumptions such as allowing each of them to vary within a permissible error range, it might be impossible to obtain unique solution from a linear least square regression analysis of such models. Therefore even if solutions are obtained from such analysis, they may not have any reliable physical meaning.

6.4. Conclusions.

Modelling the change in ^{13}C NMR shifts as a function of pH has only been done experimentally up till date. This is the first attempt to do it theoretically in order to determine protonation sequence. Our preliminary analysis suggests, although not clearly, that for the mono- and diprotonated forms, a mixture of the two possible tautomers is most likely to exist in solution. Furthermore, we have shown that relative percentages of the various tautomeric forms reported based on experimental observations should not be taken literarily as slight to medium variations in the percentages of the various tautomeric forms relative to one another did not result in a noticeable change in the NMR spectrum predicted theoretically. We showed that

correct % fraction of the HL_{N1} and HL_{N2} forms was only reproduced by explicit solvation, but accurate trends in ^{13}C NMR shifts as a function of pH was only reproduced when we used NMR shifts of fully linear conformers. Use of external reference such as dioxane goes a long way to improve the accuracy of computed chemical shifts in situations where appropriate scaling factors are unavailable. Importantly, we note that the best way to resolve this longstanding controversy in literature might be to carry out high accuracy electronic structure calculations e.g. CCSD(T) which has been referred to as the gold standard in computational chemistry to obtain the relative free energy of each possible tautomeric form from which accurate Boltzmann distribution population of each tautomeric form and their protonation sequence. However, there is a long way to go in the development of computing power for such calculations to be feasible computationally time wise. For now such problems can only be tackled using DFT methods. But as we have shown previously[27], using different density functionals resulted in varying % populations of the singly protonated tautomers being predicted to form. However a quick comparison of the structure of the two possible HL forms of *trien* when optimized in PCM and PCM+4H2O shows that protonation results in the formation of hydrogen bonding which could easily serve as a means of fast proton transfer between primary and secondary nitrogen atoms hence making it impossible for one of these tautomers to exist in solution without the other. One can also conclude that each of them might play important biological roles.

6.5. References

[1] U. Bachrach, *Plant Phys. Biochem.*, **2010,** *48*, 490–495.

[2] B. Ganem, *Acc. Chem. Res.*, **1982**, *15*, 290–298.

[3] M. D. Bratek-Wiewiorowska, M. Alejska, M. Figlerowicz, J. Barciszewski, M. Wiewiorowski, *Pure & Appl. Chem.*, **1987**, 59, 407–414.

[4] T. Thomas, T. J. Thomas, *Cell. Mol. Life Sci.*, **2001**, *58*, 244–258.

[5] A. Gugliucci, *Clinica Chimica Acta*, **2004**, *344*, 23–35.

[6] L. A. E. Batista de Carvalho, M. P. M. Marques, J. Tomkinson, *J. Phys. Chem. A.*, **2006**, *110*, 12947–12954.

[7] M. P. M. Marques, L. A. E. Batista de Carvalho, *Biochem. Soc. Trans.*, **2007,** *35*, 374–380.

[8] E. Agostinelli, M. P. M. Marques, R. Calheiros, F. P. S. C. Gil, G. Tempera, N. Viceconte, V. Battaglia, S. Grancara, A. Toninello, *Amino Acids,*. **2010**, *38*, 393–403.

[9] J. L. Sudmeir, C. N. Reiley, *Anal. Chem.*, **1964**, 1698–1706.

[10] P. Paoletti, M. Ciampolini, A. Vacca, *J. Phys. Chem.*, **1963,** *67*, 1065–1067.

[11] A. Bencini, E. Bianchi, M. Garcia-Espana, J. A. Micheloni, *Coord. Chem. Rev.*, **1999**, *188*, 97-156.

[12] M. T. Albeda, J. C. Frias, E. Garcia-Espana, *Encyclopaedia of Supramolecular Chemistry*, **2007,** *1:1*, 1-37.

[13] M. Borkovec, D. Cakara, G. J. M. Koper, *J. Phys. Chem. B.*, **2012**, *116*, 4300–4309.

[14] M. M. Kimberly, J. H. Goldstein, *Anal. Chem.*, **1981**, *53*, 789–793.

[15] D. N. Hague, A. D. Moreton, *J. Chem. Soc. Perkin Trans.*, **1994**, *2*, 265–270.

[16] S. P. Dagnall, D. N. Hague, M. E. McAdam, *J. Chem. Soc. Perkin Trans.*, **1984**, *2*, 1111–1114.

[17] M. Delfini, A. L. Segre, F. Conti, R. Barbucci, V. Barone, P. Ferruti, *J. Chem. Soc. Perkin Trans.,*. **1980**, *2*, 900–903.

[18] G. R. Hedwig, H. K. J. Powell, *J.C.S. Dalton*, **1973**, 793–797.

[19] R. Barbucci, L. Fabbrizzi, P. Paoletti, *J.C.S. Dalton*, **1972**, 745–749.

[20] I. Cukrowski, C. F. Matta, *Comput. Theoret. Chem.*, **2011**, *966*, 213–219.

[21] A. Anichini, L. Fabrizzi, R. Barbucci, A. Mastorianni, *J.C.S Dalton*, **1977**, 2224–2228.

[22] D. Aikens, S. Bunce, F. Onasch, R. Parker III, C Hurwitz, S Clemans, *Biophys. Chem*., **1983**, *17*, 67–74.

[23] M. L. Reyzer, J. S. Brodbelt, *J. Am. Soc. Mass Spectrom.*, **1998**, *9*, 1043–1048.

[24] C. A. Ilioudis, K. S. B. Hancock, D. G. Georganopoulou, J. W. Steed, *New J. Chem.*, **2000**, *24*, 787–798.

[25] I. Cukrowski, A. S Adeyinka, D. C. Liles, *Acta Cryst. E*, **2012**, 68, o2387–o2387.

[26] R. Bhula, D. C. Weatherburn, *J. Chem. Soc. Perkin Trans*. II, **1988**, 1161–1164.

[27] A. S. Adeyinka, I. Cukrowski, *J. Mol. Model*, **2015**, *21*, 162–180.

[28] M. W. Lodewyk, M. R. Siebert, D. Tantillo, *Chem. Rev.*, **2012**, *112*, 1839–1862.

[29] M. Borkovec, G. J. M. Koper, *J. Phys. Chem*. **1994**, *98*, 6038–6045.

[30] M. Borkovec, D. Cakara, G. J. M. Koper, *J. Phys. Chem. B.*, **2012**, *116*, 4300–4309.

[31] S. Salezadeh, Y. Gholiee, M. Bayat, *Int. J. Quantum Chem.*, **2011**, *11*, 3608–3615.

[32] 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, U.S. Department of Commerce, National Institute of Standards and Technology.

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

[34] A. S. Adeyinka, B. W. Bulling, I. Cukrowski, *Theor. Chem. Acc*., **2016**, *135*, 139, 2–17.

Chapter 7

Conclusions

This work involved a series of studies to determine theoretically, the macroscopic protonation constants and to ascertain the protonation sequence of aliphatic linear polyamines in solution using *trien* as a case study. To accomplish this aim, a comprehensive structural, QTAIM and NCI analysis of numerous MM-generated conformers of **HL and H_2L** tautomers of *trien* was carried out in implicit aqueous environment. This culminated in the development of a combined (MM/DFT) conformational search protocol (CSP) which has been demonstrated to be of general purpose (with slight modifications depending on the size of aliphatic linear polyamine being investigated) for generating representative sets of low energy conformers (LECs) for any given aliphatic linear polyamines in reasonable time. Consequently, this facilitated determination of percentage distribution of each tautomer of mono and di– protonated forms of trien in line with previous reports[1–3] in both implicit and explicit aqueous environment. However, relative distribution of tautomers obtained from $\Delta G(aq)$ in the presence of explicit water molecules was in closer agreement with experimental observation than what was observed in implicit aqueous environment.

7.1. Protonation Sequence

With regards to protonation sequence in implicit aqueous environment and considering LECs, it was found that, although a mixture of HL_p and HL_s forms was predicted in accord with general literature reports[1–3], the HL_s was found to be predominant (60%) in contrast to recent literature reports from cluster expansion analysis developed by Borkovec and co-workers where the HL_p form was found to be the major species in solution as it constitutes about 86% of total solution composition at thermodynamic equilibrium. In fact only the HL_s form was predicted to form in solution when fully linear conformers were used. In contrast, implementation of a hybrid solvation model where *trien* was surrounded with four explicit water molecules and then immersed in a dielectric continuum to represent bulk solvent (PCM solvation model in this case), predicted a mixture of HL_p (84%) and HL_s (16%) in excellent agreement with results obtained from cluster analysis method. In contrast to the prediction of HL_s as the only tautomer when linear conformers of trien was used to determine protonation sequence, a mixture of HL_p and HL_s form in a 40 to 60 ratio was predicted in accord with experimental data reported recently when conformers obtained from the CSP developed in implicit aqueous environment were used to determine protonation sequence. Likewise for the diprotonated form, a mixture of the two possible forms H_2L_{ps} and H_2L_{pp} forms were predicted as opposed to only the H_2L_{ps} form predicted when linear conformers alone were assessed. This work demonstrates the importance of appropriate conformational search as a prerequisite for theoretical studies involving aliphatic

linear polyamines and similar flexible biomolecules. Furthermore our preliminary ^{13}C NMR–pH titration also revealed that due to the co-existence of all possible protonated species and the minute contribution of HL forms in solution, it might be difficult, if not impossible to establish with any degree of certainty the relative %-fraction composition of each possible tautomeric HL form. It is gratifying to note however that in agreement with result obtained using the G(aq) values of conformers, analysis of NMR–pH titration curves also suggest that a mixture of both monoprotonated forms would be formed in solution.

Hence, it is not surprising that although several authors use similar ^{13}C NMR–pH titration method they propose conflicting %-fractions of HL and H_2L tautomeric forms because they extracted information about tautomeric compositions from data obtained using various theoretical assumptions and statistical analysis methods. The implication of all these is that it might be more realistic to focus on identifying predominant tautomeric species of any given aliphatic linear polyamine rather than ascertaining their exact %-fraction in solution. This would at least facilitate a comparative structure-activity study of the various possible tautomeric forms as a means of verifying which of these tautomeric forms is mainly responsible for its biological activity. According to the results of this study, the best means of obtaining an accurate %-fraction of each HL and H_2L form will be a direct calculation of the change in Gibbs free energy for the protonation of each basic site or for the protonation of a combination of basic sites (i.e. microspecies). One can then obtain the so called microscopic protonation constants for such microspecies and consequently deduce directly, the protonation sequence. However as pointed out earlier, there is need to discover the most suitable theoretical model chemistry to obtain such information as accurately as possible and with affordable computational resources. Since only the use of a higher level of theory such as CCSD(T) might guarantee such accuracy, one can only look forward to the future with excitement, knowing that at the exponential rate at which computational power is improving over the past decades, it will only be a matter of few years before such calculations become a routine for interested computational chemists.

7.2. Protonation Constants.

Combination of the conformational search protocol (CSP) developed in this work with a competition reaction methodology made it possible to predict for the first time, the four stepwise macroscopic protonation constants of trien within 0.1(-0.8) log unit of experimental values. This accuracy is very satisfactory keeping in mind the tedious albeit slightly more accurate nature of glass electrode potentiometry (GEP) experimental evaluation of stepwise protonation constants. The level of accuracy achieved for protonation constants prediction in this study is the best

reported for aliphatic linear polyamines till date. More importantly, this results paves the way for reliable theoretical prediction of the stepwise protonation constants of yet to be synthesized aliphatic linear polyamines especially the unsymmetrical ones which are difficult to synthesize in the laboratory.[4] This is a significant milestone for medicinal chemistry purposes because their usefulness as new anticancer drug templates depends to a large extent on their proton binding ability as measured by their macroscopic stepwise protonation constants. Furthermore this is a crucial indicator of the ability of a molecule to pass across cell membranes. Therefore being to assess this important property theoretically, prior to embarking on a tedious synthetic endeavour in the laboratory will go a long way to save time and cost.

In addition, our success in predicting accurately the four stepwise protonation constants of *trien* gave us confidence that our conformational search protocol is generally applicable for identifying representative low energy conformers of any given aliphatic linear polyamine. It is also important to note that a pre-optimization and 5-step *EEBGB*- conformer selection protocol which affords up to 94% reduction of conformers submitted for frequency calculation in the course of protonation constants determination was outlined; this makes the task computationally feasible and practicable time-wise. With slight modifications to suit the unique structural properties of a given molecules, these protocols would likely find useful application in theoretical studies of other large and flexible molecules such as aliphatic linear polyamines and should be explored further in this regard.

7.3. Conformational Preference and nature of intramolecular interactions

An in-depth theoretical analysis of factors responsible for structural-topological and conformational preferences of polyamines using theoretical tools such as QTAIM, NCI and IQA revealed that in addition to NH•••N interactions which are chiefly responsible for the relative stability of confirmers, there are also a number of locally stabilizing (as revealed by IQA interaction energy) CH•••HC as well as CH•••N interactions which subtly influence conformer structure, topology and stability. A thorough investigation of the nature of these intramolecular interactions using standard procedures of QTAIM, IQA and NCI techniques to examine electron density topology in interatomic regions, as well as our in-house developed cross-sections of the electron and deformation densities showed that (i) not a single one of the topological indices tested in this work i.e. QTAIM-defined atomic interaction line (AIL), NCI-defined isosurfaces which distinguishes between local regions of accumulated ($\lambda_2 < 0$) or depleted ($\lambda_2 > 0$) density relative to immediate environment, IQA-defined interaction energy $E_{\text{int}}^{\text{A,B}}$, and deformation density for which $\Delta\rho(\boldsymbol{r}) > 0$ indicates an inflow or otherwise an outflow of density due to the

formation of an interaction; either individually or combined can be used to predict with certainty, the (de)stabilizing influence of an intramolecular interaction on a molecule. This implies that it might be difficult if not impossible to rationalize quantitatively why a conformer is preferred over another in polyatomic molecules such as aliphatic linear polyamines. However it is clear from this investigation that the classical chemist's classification of CH•••HC interactions as steric clashes is questionable. For instance, even though our initial set of conformers were generated with MMFF force field which is parameterized by default to recognise CH•••HC interactions as steric clashes, energetic analysis of conformers shows that almost all of the ten lowest in energy conformers of *trien* (in which CH•••HC interactions was observed) found after full optimizations with DFT and MP2 methods were also among the top 20 LECs generated from MMFFaq. These conformers should have ended up among medium and high energy conformers due to the destabilization caused by the CH•••HC interactions if truly they are steric clashes.

7.4. Solvation Method

Apart from being able to identify representative LECs, the use of appropriate solvation models to describe solvation environment appears to be very crucial for obtaining useful results in studies such as the one carried out in this work. For example, Predicted protonation sequence changed drastically when a hybrid solvation method involving four explicit water molecules and the continuum solvation model was utilized. Similarly, this solvation model facilitated a better prediction of protonation constants. However this was not without its own disadvantages as the protonation constants study showed that placement of water molecules around molecule of interest could lead to errors in thermodynamic quantities. Also our ^{13}C NMR–pH study showed that interaction of explicit solvent molecules with the molecule being investigated could also affect its geometry hence resulting in large deviations of computed NMR shifts from experimental values. Therefore the use of explicit solvent molecules in theoretical studies should be handled with extreme caution as it could introduce hidden errors which might lead to incorrect interpretation of results. The use of at least QM/MM or the more computationally expensive molecular dynamics method should be the first choice in such situations.

7.5 Level of Theory

An extensive comparison of various levels of theory (LoT) such as HF, B3LYP, B97D and MP2 was undertaken in this study. Results demonstrated that B3LYP was best reproducing MP2 data; hence it can be reliably used to carry out theoretical studies on aliphatic linear polyamines. There was only a slight difference in the performance of B3LYP and B97D LoT. There exists however a plethora of DFT methods which might perform better than B3LYP that were not

evaluated in this work due to time constraints. On the other hand HF method performed poorly in modelling structure of aliphatic linear polyamines and is therefore not recommended for future investigations.

7.6. Future Studies

There are a few interesting further research directions that could be pursued based on results of this study. Firstly, it would be important to find out the most cost-effective and adequate means of accounting for solvation effects when studies involving aliphatic linear polyamines are concerned. In this regard, it would be interesting to find out in particular the effect of implementing a stepwise increase in the number of explicit water molecules to dissipate charge throughout the macromolecular assembly of highly charged H_nL^{n+} species on computed protonation constants. Also, the use of QM/MM methods which is far less expensive than molecular dynamics simulations should be explored. It would also be interesting to see how this would influence accuracy of computed ^{13}C NMR chemical shifts. The attempt to understand factors responsible for conformational preference in aliphatic linear polyamines revealed the need for a thorough and in-depth study aimed at uncovering the overall (de)stabilizing effect of these CH•••HC interactions in conformers of aliphatic linear polyamines. Finally, it would be interesting to investigate the impact of various DFT functionals and available dispersion correction methods on the quality of computed protonation constants and ^{13}C NMR shifts.

7.7. References

[1] M. Borkovec, D. Cakara, G.J.M. Koper, *J Phys Chem B,*. **2012**, *116*, 4300–4309.

[2] D.N. Hague, A.D. Moreton, *J Chem Soc Perkin Trans,*. **1994**, *2*, 265–270.

[3] S.P. Dagnall, D.N. Hague, M.E. McAdam, *J Chem Soc Perkin Trans*, **1984**, *2*, 1111–1114

[4] T. Pirali, G. Callipari, E. Ercolano, A. A. Genazzani,G. B. Giovenzana, G. C. Tron, *Org. Lett.*, **2008**, *10*, 4199–4202

Appendix A

Supplementary Information for Chapter 3

Table A1. List of H_2L conformers with one imaginary frequency found at indicated level of theory

H_2L_{ps}			H_2L_{pp}		
HF	**B3LYP**	**B97D**	**HF**	**B3LYP**	**B97D**
$C_{ps}04$	$C_{ps}14$	$C_{ps}24$	$C_{pp}11$	$C_{pp}01$	$C_{pp}04$
			$C_{pp}16$	$C_{pp}04$	$C_{pp}11$
			$C_{pp}18$		$C_{pp}16$

Table A2. Relative electronic energies ($\Delta E = E_{Conf} - E_{LEC}$ in kcal/mol) of thirty lowest energy conformers found at indicated levels of theory and Boltzmann distribution, as a %-fraction, of the total population for: part (a) – HL_p, part (b) - HL_p, part (c) - H_2L_{ps} and part (d) - H_2L_{pp}

Part (a)[a]

HF			B3LYP			B97D			MP2		
HL_p	ΔE	%	HL_p	ΔE	%	HL_p	ΔE	%	HL_p	ΔE	%
C_p03	0.00	29.3	C_p02	0.00	35.7	C_p02	0.00	39.3	C_p02	0.00	57.3
C_p02	0.12	23.9	C_p03	0.09	30.8	C_p01	0.04	36.5	C_p01	0.70	17.4
C_p05	0.27	18.5	C_p01	0.45	16.8	C_p03	0.63	13.5	C_p03	0.71	17.3
C_p04	0.38	15.5	C_p04	0.82	8.9	C_p06	1.43	3.5	C_p04	1.52	4.4
C_p01	0.95	5.9	C_p05	1.31	3.9	C_p09	1.86	1.7	C_p05	2.09	1.7
C_p08	1.19	3.9	C_p09	1.79	1.7	C_p05	1.91	1.6	C_p06	2.47	0.9
C_p15	2.17	0.8	C_p06	2.19	0.9	C_p04	1.99	1.4	C_p08	2.57	0.7
C_p06	2.26	0.6	C_p08	2.47	0.6	C_p20	2.10	1.1	C_p09	3.63	0.1
C_p07	2.26	0.6	C_p20	2.71	0.4	C_p08	2.56	0.5	C_p20	3.75	0.1
C_p09	2.71	0.3	C_p15	3.69	0.1	C_p13	2.92	0.3	C_p23	4.61	0.0
C_p20	3.04	0.2	C_p29	4.05	0.0	C_p11	3.09	0.2	C_p29	4.95	0.0
C_p29	3.27	0.1	C_p13	4.40	0.0	C_p23	3.24	0.2	C_p26	5.15	0.0
C_p33	3.69	0.1	C_p11	4.43	0.0	C_p29	3.59	0.1	C_p15	5.16	0.0
C_p40	3.77	0.1	C_p23	4.54	0.0	C_p26	4.01	0.0	C_p24	5.17	0.0
C_p26	4.08	0.0	C_p24	4.75	0.0	C_p24	4.01	0.0	C_p13	5.54	0.0
C_p27	4.08	0.0	C_p26	4.82	0.0	C_p31	4.63	0.0	C_p11	5.60	0.0
C_p23	4.10	0.0	C_p27	5.24	0.0	C_p15	4.82	0.0	C_p27	5.72	0.0
C_p24	4.11	0.0	C_p31	5.45	0.0	C_p27	5.13	0.0	C_p31	6.36	0.0
C_p36	4.25	0.0	C_p32	5.58	0.0	C_p32	5.38	0.0	C_p32	6.85	0.0
C_p37	4.72	0.0	C_p33	6.66	0.0	C_p33	10.26	0.0	C_p33	8.96	0.0
C_p32	4.76	0.0	C_p36	7.76	0.0	C_p42	10.30	0.0	C_p42	10.12	0.0
C_p41	4.86	0.0	C_p40	7.98	0.0	C_p30	11.20	0.0	C_p30	10.21	0.0
C_p31	4.93	0.0	C_p42	8.06	0.0	C_p36	11.50	0.0	C_p36	10.31	0.0
C_p13	5.11	0.0	C_p37	8.32	0.0	C_p37	11.94	0.0	C_p37	10.96	0.0
C_p42	5.45	0.0	C_p41	8.69	0.0	C_p38	12.29	0.0	C_p40	11.53	0.0
C_p11	5.51	0.0	C_p30	8.81	0.0	C_p40	12.75	0.0	C_p38	11.57	0.0
C_p30	5.56	0.0	C_p38	9.41	0.0	C_p41	12.93	0.0	C_p41	12.02	0.0
C_p43	6.09	0.0	C_p43	9.95	0.0	C_p43	14.24	0.0	C_p43	12.93	0.0
C_p38	6.27	0.0	C_p54	12.45	0.0	C_p54	15.49	0.0	C_p54	15.04	0.0
C_p54	8.02	0.0									

[a] C_p06 and C_p07 MM-generated conformers optimized to the same structure at DFT and MP2 levels, hence they are showed only at HF level

Part (b)

HF			B3LYP			B97D			MP2		
HL_s	**ΔE**	**%**	**HL_s**	**ΔE**	**%**	**HL_s**	**ΔE**	**%**	**HL_s**	**ΔE**	**%**
C_s08	0.00	23.2	C_s04	0.00	41.2	C_s01	0.00	33.2	C_s04	0.00	40.7
C_s04	0.14	18.5	C_s01	0.28	25.5	C_s04	0.04	31.2	C_s01	0.14	32.0
C_s01	0.26	14.9	C_s03	0.43	19.9	C_s02	0.27	21.1	C_s03	0.64	13.9
C_s03	0.33	13.4	C_s02	0.83	10.1	C_s03	0.56	12.8	C_s02	0.71	12.3
C_s09	0.47	10.4	C_s06	2.36	0.8	C_s05	2.36	0.6	C_s05	2.72	0.4
C_s06	0.69	7.3	C_s22	2.44	0.7	C_s22	2.52	0.5	C_s22	2.90	0.3
C_s02	1.16	3.3	C_s07	2.75	0.4	C_s13	2.76	0.3	C_s06	3.58	0.1
C_s07	1.49	1.9	C_s08	2.81	0.4	C_s14	2.77	0.3	C_s13	3.59	0.1
C_s22	1.71	1.3	C_s05	2.91	0.3	C_s06	5.03	0.0	C_s14	3.60	0.1
C_s32	1.79	1.1	C_s09	3.05	0.2	C_s07	5.18	0.0	C_s07	3.98	0.0
C_s05	1.94	0.9	C_s13	3.27	0.2	C_s32	5.50	0.0	C_s09	5.05	0.0
C_s10	2.14	0.6	C_s14	3.27	0.2	C_s24	6.26	0.0	C_s08	5.12	0.0
C_s30	2.26	0.5	C_s32	3.44	0.1	C_s09	6.47	0.0	C_s32	5.19	0.0
C_s12	2.45	0.4	C_s10	3.98	0.0	C_s23	6.51	0.0	C_s10	5.61	0.0
C_s20	2.55	0.3	C_s24	4.06	0.0	C_s10	6.51	0.0	C_s24	5.73	0.0
C_s24	2.57	0.3	C_s12	4.32	0.0	C_s25	6.52	0.0	C_s12	6.02	0.0
C_s13	2.70	0.2	C_s20	4.66	0.0	C_s12	6.66	0.0	C_s23	6.19	0.0
C_s14	2.71	0.2	C_s23	4.81	0.0	C_s08	6.83	0.0	C_s20	6.32	0.0
C_s25	2.85	0.2	C_s11	5.20	0.0	C_s20	7.21	0.0	C_s25	6.33	0.0
C_s44	2.85	0.2	C_s25	5.24	0.0	C_s11	7.34	0.0	C_s11	6.61	0.0
C_s25	2.90	0.2	C_s30	5.36	0.0	C_s29	9.08	0.0	C_s30	8.05	0.0
C_s29	3.19	0.1	C_s29	6.15	0.0	C_s35	9.12	0.0	C_s29	8.13	0.0
C_s23	3.19	0.1	C_s35	6.29	0.0	C_s30	9.33	0.0	C_s35	8.36	0.0
C_s56	3.23	0.1	C_s27	6.41	0.0	C_s56	9.35	0.0	C_s27	8.61	0.0
C_s41	3.29	0.1	C_s39	6.45	0.0	C_s27	9.45	0.0	C_s56	8.91	0.0
C_s42	3.29	0.1	C_s56	6.46	0.0	C_s41	9.87	0.0	C_s41	9.13	0.0
C_s27	3.38	0.1	C_s44	6.60	0.0	C_s42	9.90	0.0	C_s42	9.19	0.0
C_s40	3.54	0.1	C_s41	6.73	0.0	C_s44	10.43	0.0	C_s44	9.40	0.0
C_s11	3.68	0.0	C_s42	6.74	0.0	C_s39	10.61	0.0	C_s39	9.56	0.0
C_s35	3.84	0.0	C_s40	6.94	0.0	C_s40	10.61	0.0	C_s40	9.79	0.0

Part (c)[a]

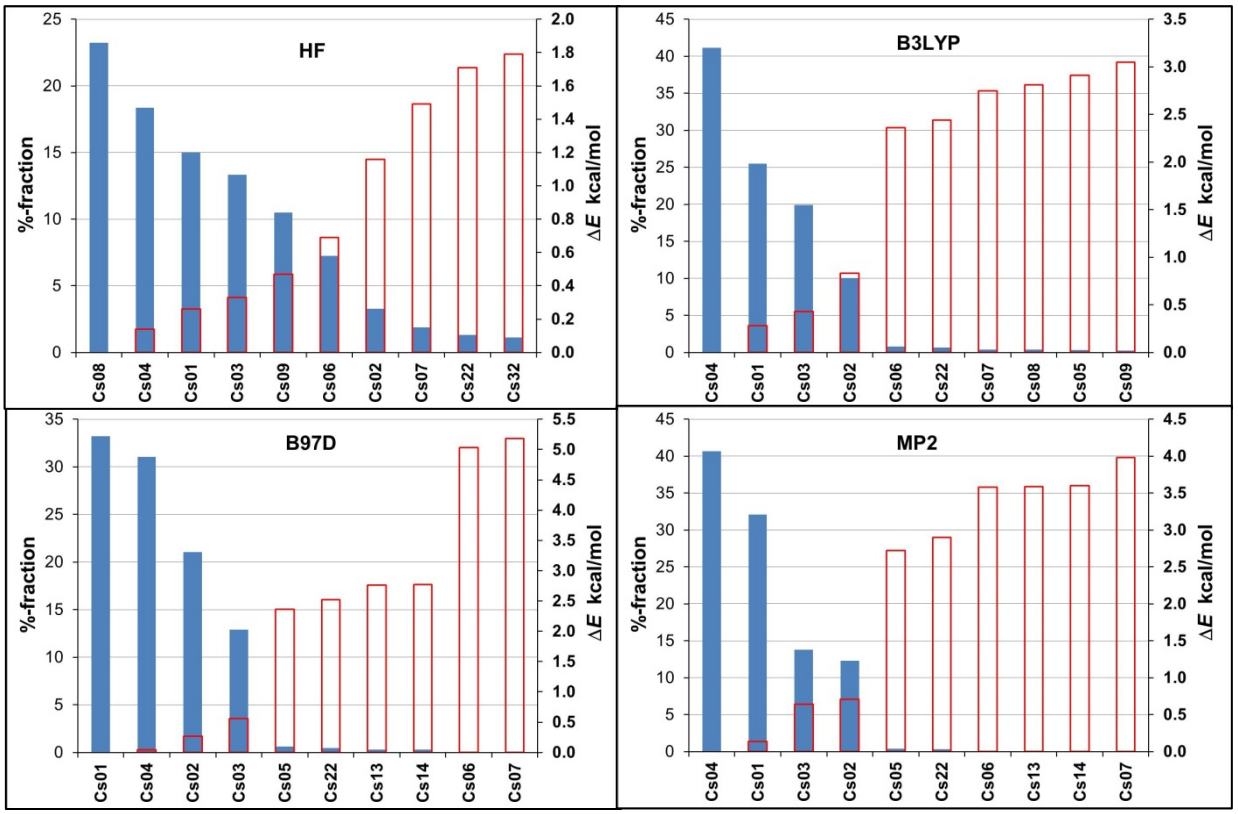
HF			B3LYP			B97D			MP2		
H_2L_{ps}	ΔE	%	H_2L_{ps}	ΔE	%	H_2L_{ps}	ΔE	%	H_2L_{ps}	ΔE	%
$C_{ps}01$	0.00	58.0	$C_{ps}01$	0.00	82.0	$C_{ps}01$	0.00	55.6	$C_{ps}01$	0.00	71.8
$C_{ps}03$	0.79	15.3	$C_{ps}41$	1.30	9.2	$C_{ps}41$	0.29	34.0	$C_{ps}41$	0.99	13.6
$C_{ps}41$	0.90	12.7	$C_{ps}10$	1.87	3.5	$C_{ps}10$	1.00	10.2	$C_{ps}10$	0.99	13.4
$C_{ps}02$	1.89	2.4	$C_{ps}03$	1.89	3.4	$C_{ps}03$	3.84	0.1	$C_{ps}03$	2.60	0.9
$C_{ps}05$	1.91	2.3	$C_{ps}02$	3.24	0.3	$C_{ps}14$	4.38	0.0	$C_{ps}14$	3.69	0.1
$C_{ps}10$	2.10	1.7	$C_{ps}14$	3.31	0.3	$C_{ps}15$	6.12	0.0	$C_{ps}05$	4.98	0.0
$C_{ps}06$	2.19	1.4	$C_{ps}05$	3.51	0.2	$C_{ps}21$	6.24	0.0	$C_{ps}17$	5.07	0.0
$C_{ps}17$	2.48	0.9	$C_{ps}12$	3.62	0.2	$C_{ps}12$	6.29	0.0	$C_{ps}02$	5.07	0.0
$C_{ps}13$	2.64	0.7	$C_{ps}04$	3.81	0.1	$C_{ps}18$	6.46	0.0	$C_{ps}34$	5.23	0.0
$C_{ps}20$	2.64	0.7	$C_{ps}17$	3.88	0.1	$C_{ps}04$	6.50	0.0	$C_{ps}43$	5.24	0.0
$C_{ps}12$	2.68	0.6	$C_{ps}13$	3.94	0.1	$C_{ps}43$	6.53	0.0	$C_{ps}04$	5.25	0.0
$C_{ps}14$	2.72	0.6	$C_{ps}21$	4.03	0.1	$C_{ps}17$	6.55	0.0	$C_{ps}12$	5.31	0.0
$C_{ps}04$	2.80	0.5	$C_{ps}15$	4.11	0.1	$C_{ps}02$	6.56	0.0	$C_{ps}21$	5.33	0.0
$C_{ps}09$	2.84	0.5	$C_{ps}09$	4.12	0.1	$C_{ps}13$	6.56	0.0	$C_{ps}09$	5.33	0.0
$C_{ps}34$	2.88	0.4	$C_{ps}18$	4.22	0.1	$C_{ps}24$	6.70	0.0	$C_{ps}15$	5.34	0.0
$C_{ps}21$	3.05	0.3	$C_{ps}34$	4.22	0.1	$C_{ps}29$	6.74	0.0	$C_{ps}13$	5.41	0.0
$C_{ps}07$	3.32	0.2	$C_{ps}06$	4.35	0.1	$C_{ps}34$	6.74	0.0	$C_{ps}18$	5.58	0.0
$C_{ps}43$	3.35	0.2	$C_{ps}43$	4.45	0.0	$C_{ps}05$	6.75	0.0	$C_{ps}06$	5.66	0.0
$C_{ps}15$	3.56	0.1	$C_{ps}07$	4.50	0.0	$C_{ps}11$	6.80	0.0	$C_{ps}11$	5.91	0.0
$C_{ps}24$	3.60	0.1	$C_{ps}24$	4.57	0.0	$C_{ps}09$	6.81	0.0	$C_{ps}24$	5.95	0.0
$C_{ps}18$	3.68	0.1	$C_{ps}11$	4.62	0.0	$C_{ps}07$	7.12	0.0	$C_{ps}07$	5.96	0.0
$C_{ps}11$	3.92	0.1	$C_{ps}29$	5.13	0.0	$C_{ps}06$	7.15	0.0	$C_{ps}29$	6.18	0.0
$C_{ps}29$	4.48	0.0	$C_{ps}19$	6.70	0.0	$C_{ps}22$	9.13	0.0	$C_{ps}22$	8.33	0.0
$C_{ps}27$	4.50	0.0	$C_{ps}22$	6.88	0.0	$C_{ps}19$	10.15	0.0	$C_{ps}19$	8.93	0.0
$C_{ps}19$	4.52	0.0	$C_{ps}32$	7.30	0.0	$C_{ps}32$	10.29	0.0	$C_{ps}48$	9.78	0.0
$C_{ps}49$	4.89	0.0	$C_{ps}48$	7.48	0.0	$C_{ps}48$	10.59	0.0	$C_{ps}49$	9.80	0.0
$C_{ps}48$	4.89	0.0	$C_{ps}49$	7.50	0.0	$C_{ps}49$	10.60	0.0			
$C_{ps}08$	4.92	0.0									
$C_{ps}32$	5.33	0.0									
$C_{ps}22$	5.41	0.0									

[a]Some higher HF-energy C_{ps} conformers failed to optimize at higher level of theory

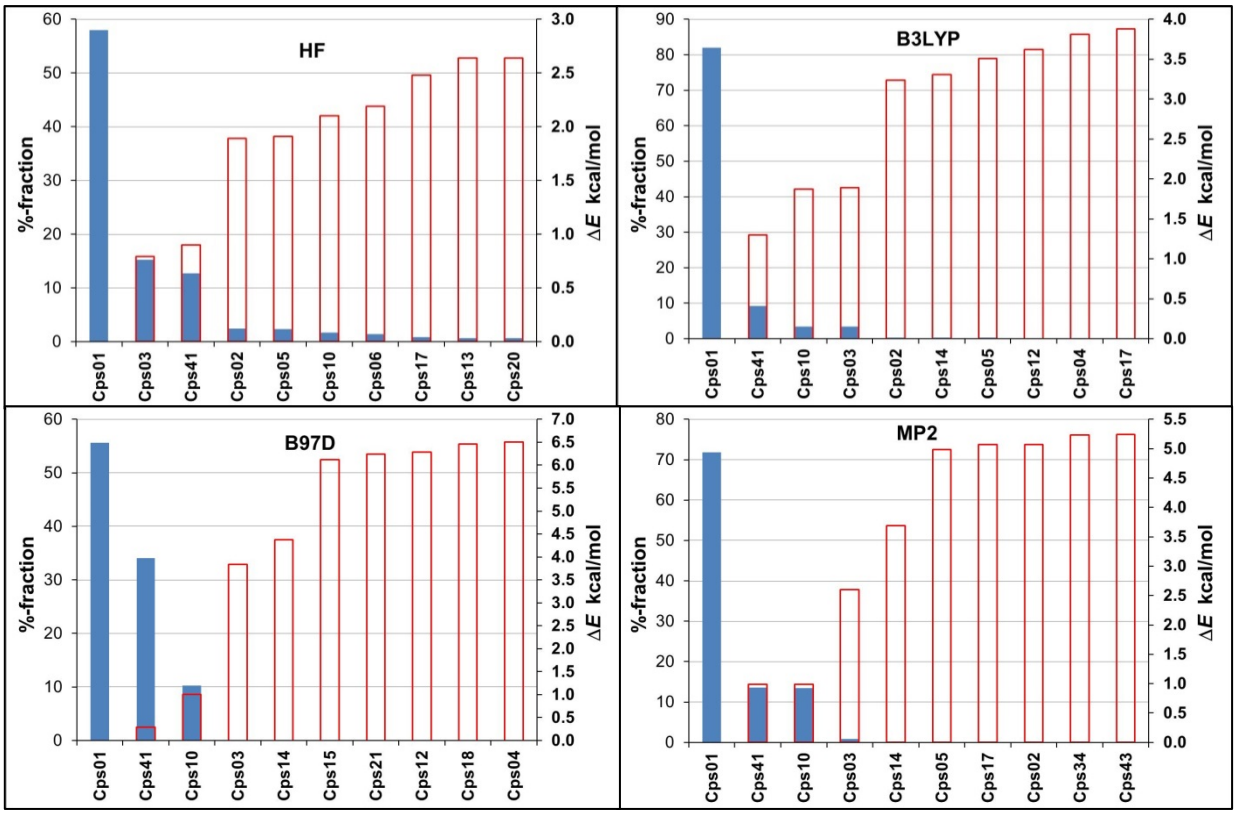
Part (d)

HF			B3LYP			B97D			MP2		
H_2L_{pp}	ΔE	%	H_2L_{pp}	ΔE	%	H_2L_{pp}	ΔE	%	H_2L_{pp}	ΔE	%
$C_{pp}02$	0.00	19.9	$C_{pp}12$	0.00	16.4	$C_{pp}12$	0.00	26.7	$C_{pp}12$	0.00	32.7
$C_{pp}01$	0.09	17.0	$C_{pp}02$	0.20	11.7	$C_{pp}10$	0.37	14.4	$C_{pp}10$	0.73	9.5
$C_{pp}05$	0.46	9.2	$C_{pp}01$	0.31	9.6	$C_{pp}14$	0.78	7.1	$C_{pp}07$	0.86	7.6
$C_{pp}17$	0.51	8.4	$C_{pp}05$	0.36	8.9	$C_{pp}21$	0.87	6.1	$C_{pp}16$	1.09	5.2
$C_{pp}03$	0.52	8.2	$C_{pp}10$	0.50	7.1	$C_{pp}19$	0.87	6.1	$C_{pp}11$	1.09	5.2
$C_{pp}19$	0.54	7.9	$C_{pp}09$	0.59	6.1	$C_{pp}11$	0.89	5.9	$C_{pp}09$	1.13	4.9
$C_{pp}08$	0.76	5.5	$C_{pp}04$	0.60	5.9	$C_{pp}16$	0.89	5.9	$C_{pp}05$	1.26	3.9
$C_{pp}23$	0.91	4.3	$C_{pp}07$	0.66	5.4	$C_{pp}27$	1.05	4.5	$C_{pp}02$	1.38	3.2
$C_{pp}04$	0.97	3.9	$C_{pp}03$	0.75	4.6	$C_{pp}13$	1.19	3.6	$C_{pp}23$	1.40	3.1
$C_{pp}06$	1.08	3.2	$C_{pp}23$	0.81	4.2	$C_{pp}18$	1.19	3.6	$C_{pp}04$	1.41	3.0
$C_{pp}15$	1.12	3.0	$C_{pp}11$	0.98	3.1	$C_{pp}07$	1.27	3.2	$C_{pp}01$	1.46	2.8
$C_{pp}10$	1.37	2.0	$C_{pp}16$	0.98	3.1	$C_{pp}05$	1.52	2.0	$C_{pp}27$	1.50	2.6
$C_{pp}24$	1.69	1.1	$C_{pp}06$	1.09	2.6	$C_{pp}09$	1.69	1.5	$C_{pp}14$	1.51	2.5
$C_{pp}16$	1.74	1.1	$C_{pp}08$	1.24	2.0	$C_{pp}04$	1.72	1.5	$C_{pp}13$	1.61	2.1
$C_{pp}09$	1.84	0.9	$C_{pp}18$	1.31	1.8	$C_{pp}08$	1.78	1.3	$C_{pp}18$	1.61	2.1
$C_{pp}07$	1.84	0.9	$C_{pp}13$	1.31	1.8	$C_{pp}03$	1.86	1.1	$C_{pp}03$	1.63	2.1
$C_{pp}18$	2.14	0.5	$C_{pp}14$	1.38	1.6	$C_{pp}02$	1.92	1.0	$C_{pp}19$	1.67	2.0
$C_{pp}12$	2.16	0.5	$C_{pp}21$	1.59	1.1	$C_{pp}23$	1.93	1.0	$C_{pp}21$	1.67	1.9
$C_{pp}11$	2.18	0.5	$C_{pp}19$	1.59	1.1	$C_{pp}01$	2.06	0.8	$C_{pp}06$	1.82	1.5
$C_{pp}26$	2.19	0.5	$C_{pp}27$	1.85	0.7	$C_{pp}06$	2.10	0.8	$C_{pp}08$	1.89	1.3
$C_{pp}14$	2.41	0.3	$C_{pp}15$	2.57	0.2	$C_{pp}15$	2.30	0.6	$C_{pp}15$	2.89	0.2
$C_{pp}13$	2.71	0.2	$C_{pp}20$	2.63	0.2	$C_{pp}17$	2.37	0.5	$C_{pp}20$	2.98	0.2
$C_{pp}27$	2.73	0.2	$C_{pp}17$	2.63	0.2	$C_{pp}20$	2.37	0.5	$C_{pp}17$	2.98	0.2
$C_{pp}21$	2.80	0.2	$C_{pp}24$	2.91	0.1	$C_{pp}30$	4.25	0.0	$C_{pp}24$	3.86	0.0
$C_{pp}31$	2.89	0.2	$C_{pp}26$	2.91	0.1	$C_{pp}31$	4.34	0.0	$C_{pp}26$	3.86	0.0
$C_{pp}22$	2.92	0.1	$C_{pp}31$	3.42	0.1	$C_{pp}24$	4.37	0.0	$C_{pp}31$	4.08	0.0
$C_{pp}40$	3.04	0.1	$C_{pp}30$	3.62	0.0	$C_{pp}26$	4.37	0.0	$C_{pp}30$	4.13	0.0
$C_{pp}30$	3.19	0.1	$C_{pp}40$	3.75	0.0	$C_{pp}40$	4.50	0.0	$C_{pp}40$	4.53	0.0
$C_{pp}25$	3.21	0.1	$C_{pp}22$	4.07	0.0	$C_{pp}25$	5.56	0.0	$C_{pp}22$	5.59	0.0
$C_{pp}20$	3.53	0.1	$C_{pp}25$	4.37	0.0	$C_{pp}22$	5.88	0.0	$C_{pp}25$	5.62	0.0

Part (a)



Part (b)



Part (c)

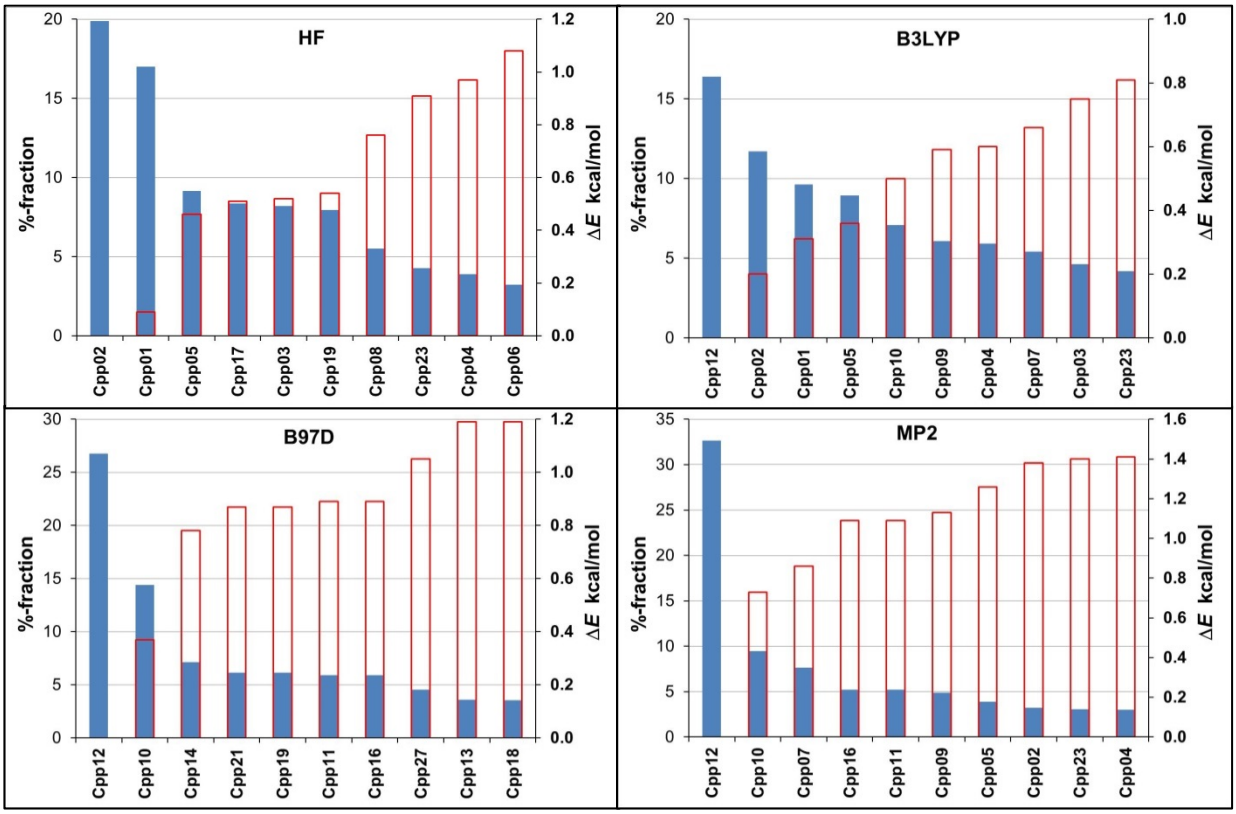
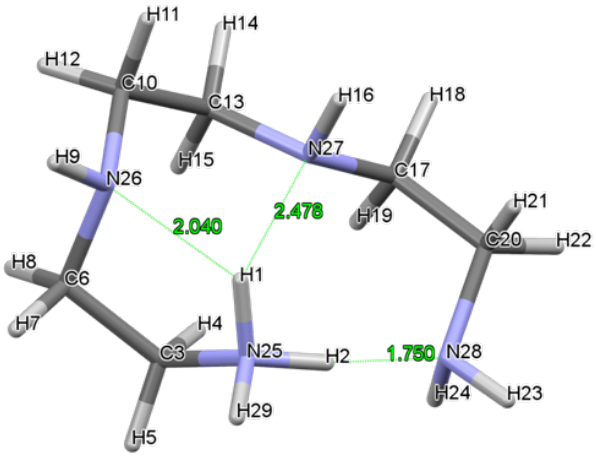
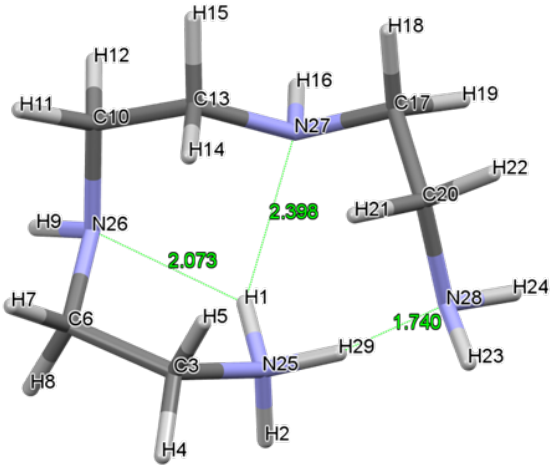


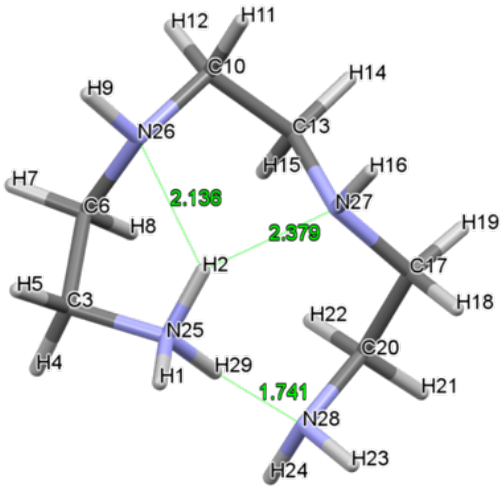
Figure. A1. Graphical presentation of ten lowest energy conformers of *trien*, in terms of %-fraction computed from Boltzmann distribution (solid bars) and relative energies (in kcal/mol) obtained at the indicated level of theory for: part (a) - HL_s, part (b) - H_2L_{ps} and part (c) H_2L_{pp} tautomers



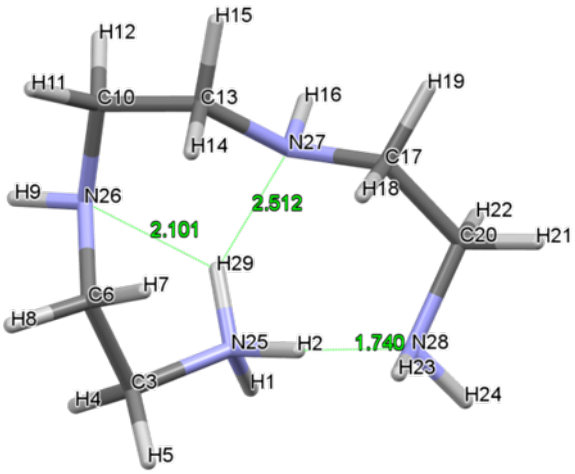
C_p02



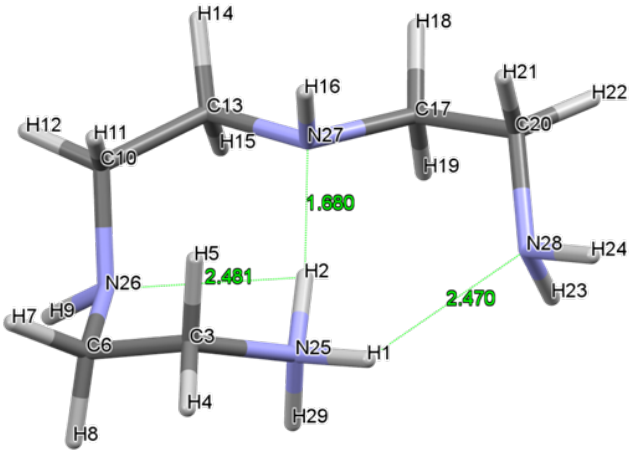
C_p01



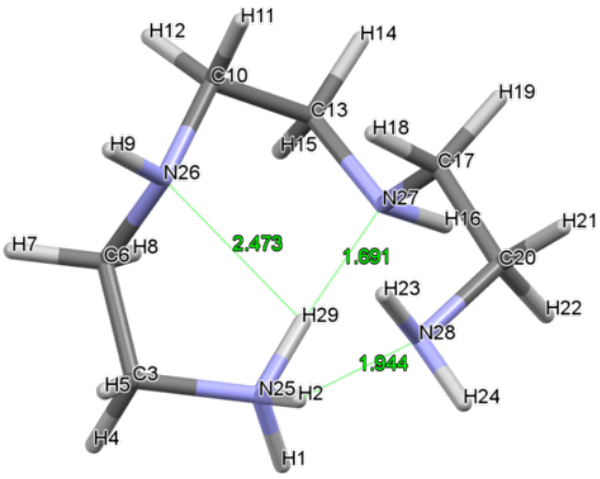
C_p03



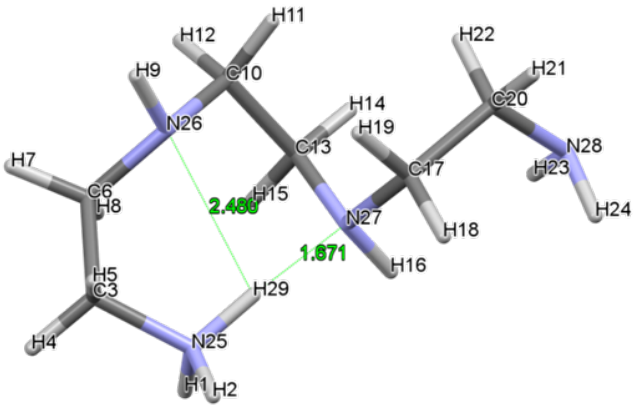
C_p04



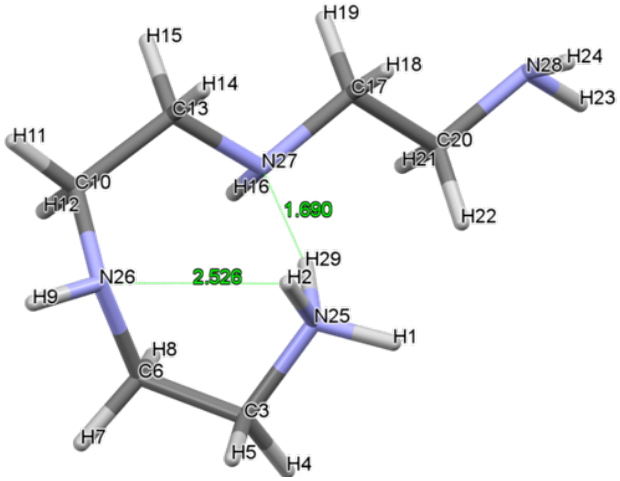
C_p05



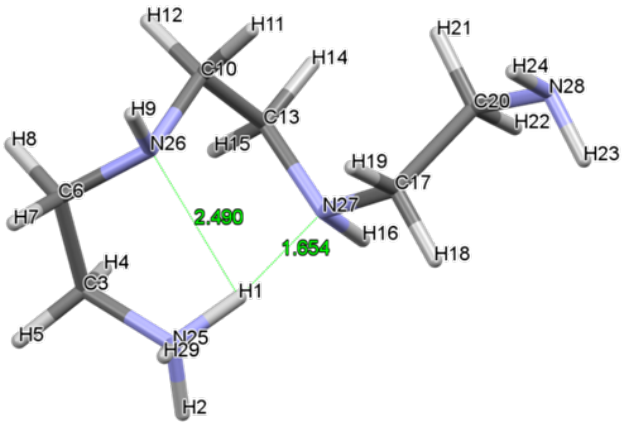
C_p06



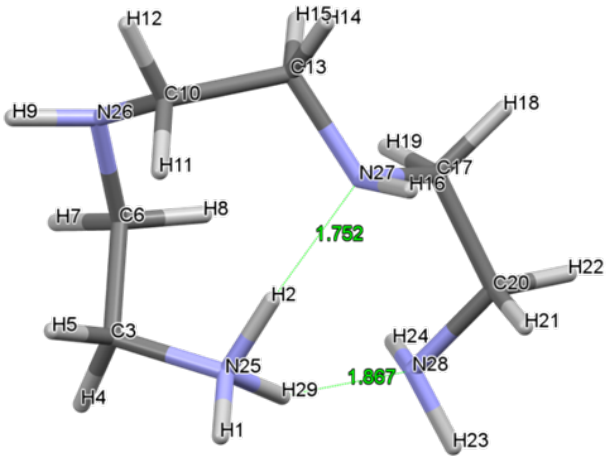
C_p26



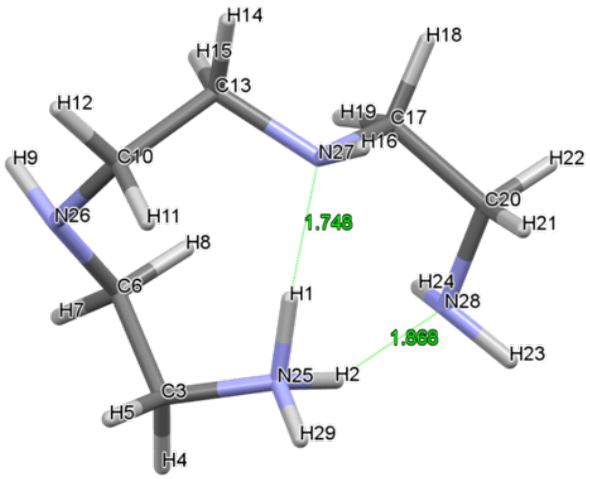
C_p15



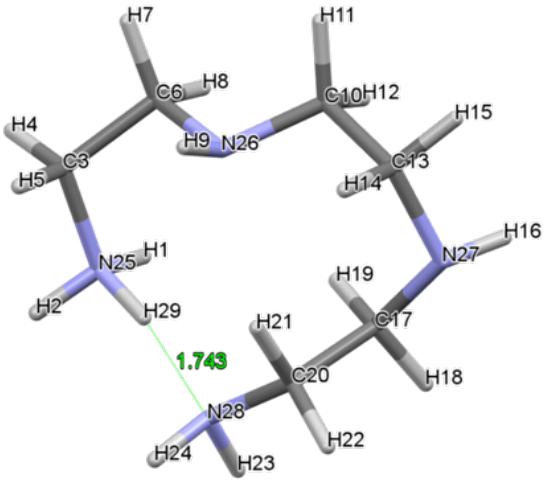
C_p24



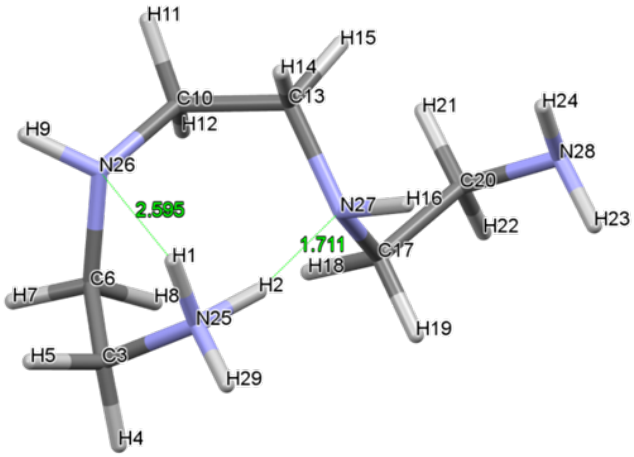
C_p13



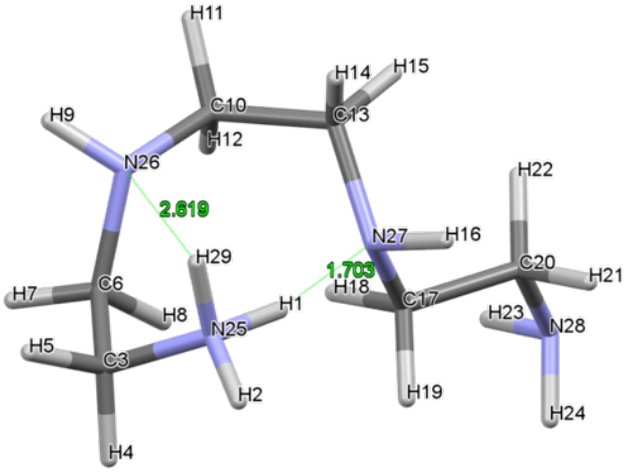
C_p11



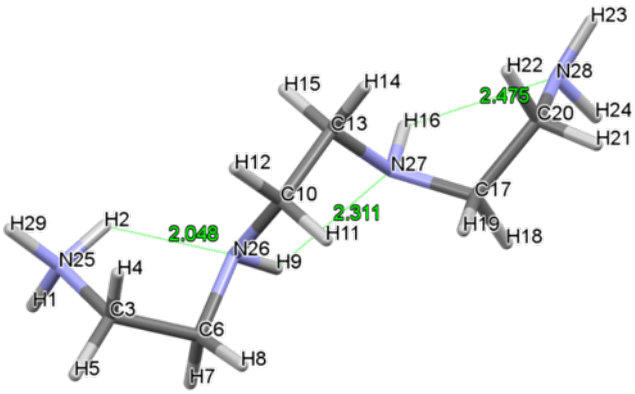
C_p27



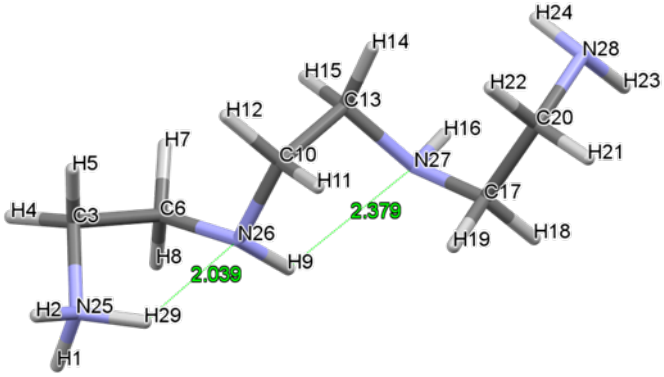
C_p31



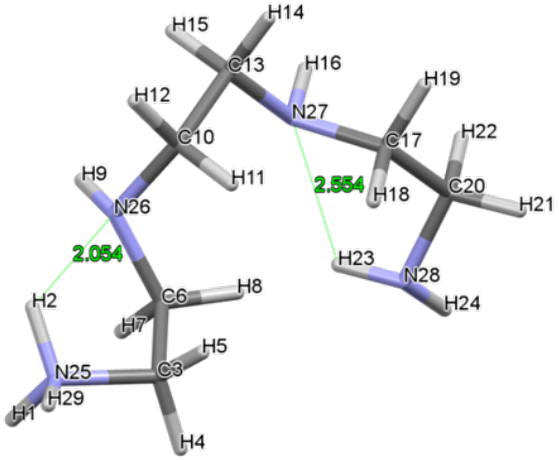
C_p32



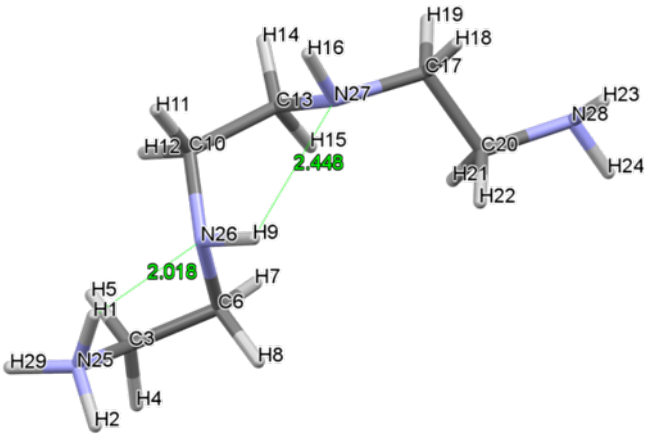
C_p33



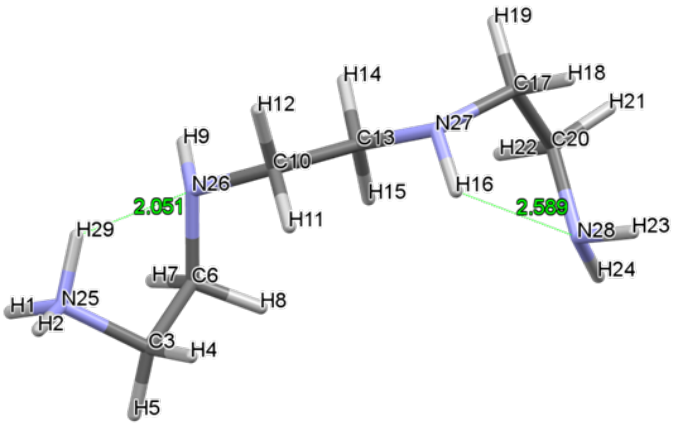
C_p42



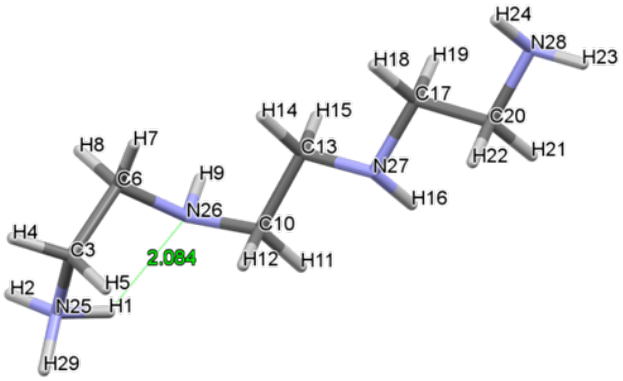
C_p30



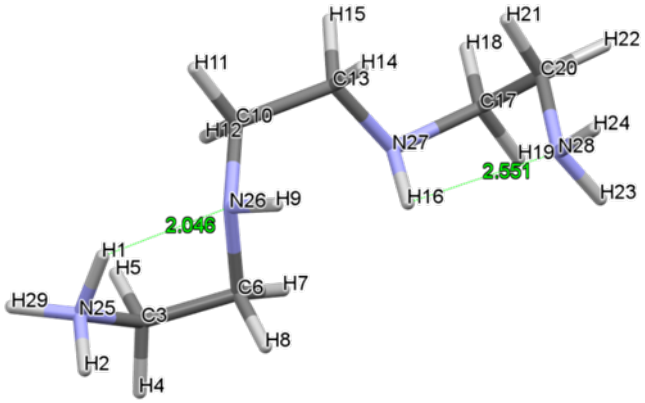
C_p36



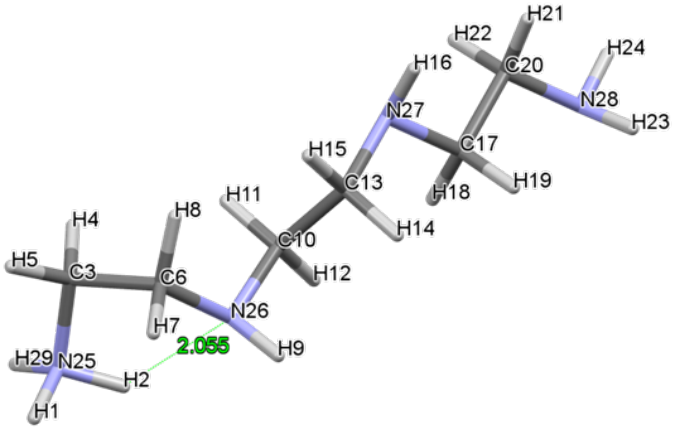
C_p37



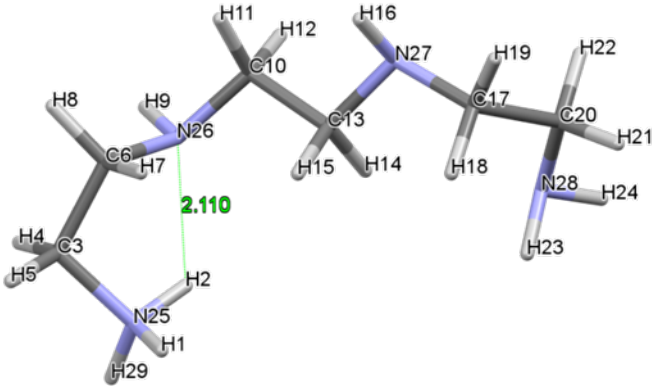
C_p40



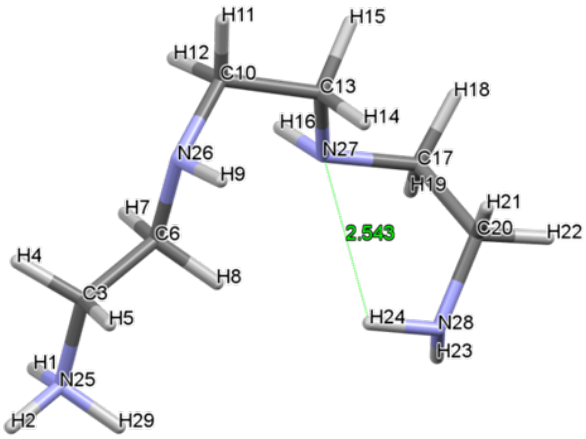
C_p38



C_p41

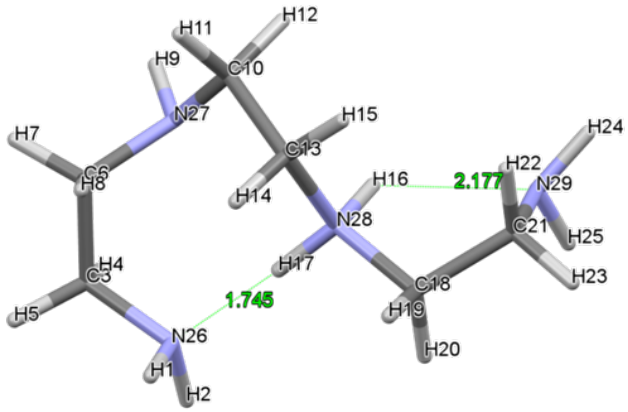


C_p43

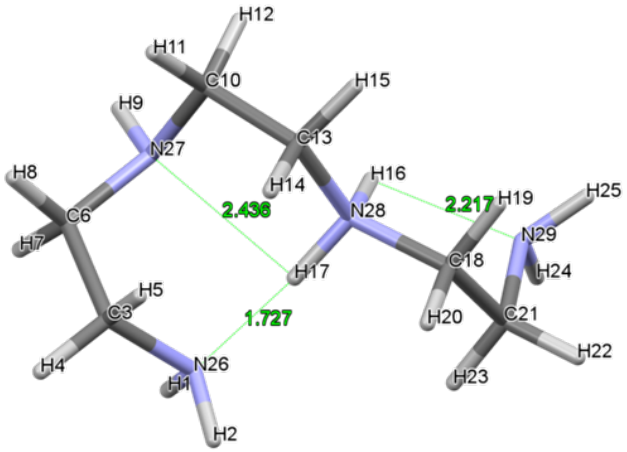


C_p54

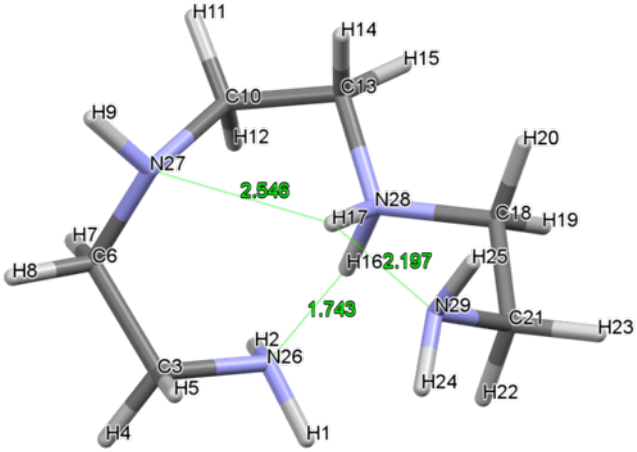
Figure A2. Capped-stick representation of HL_p structures of 30 LECs generated at MP2 during the fourth and final stage of the conformational protocol developed in this work, also showing atoms' numbering as well as interatomic distances in Å of short NH--N contacts



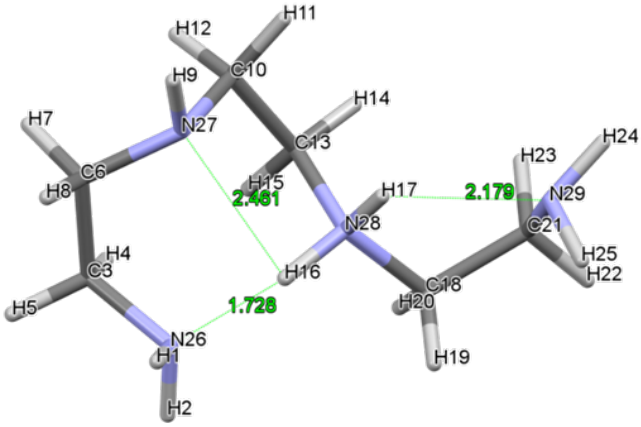
C_s04



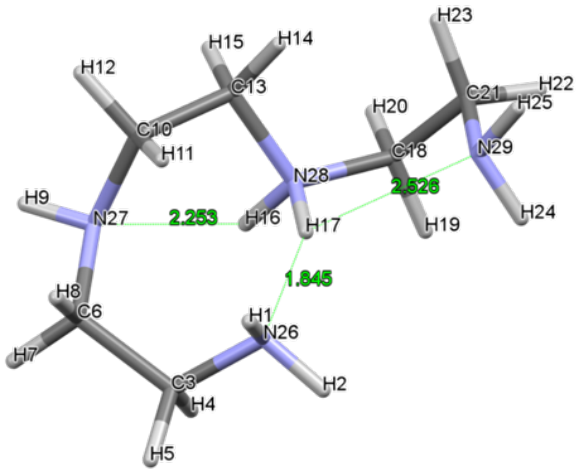
C_s01



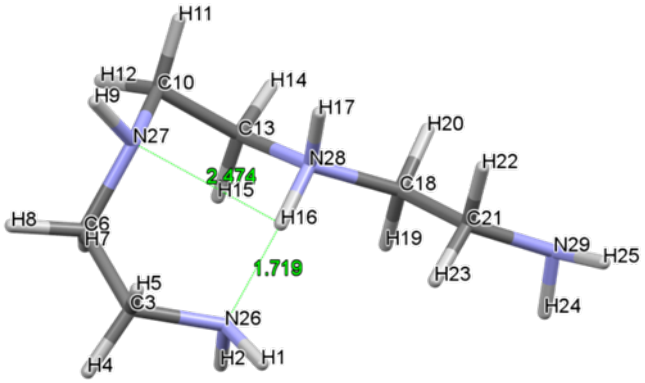
C_s03



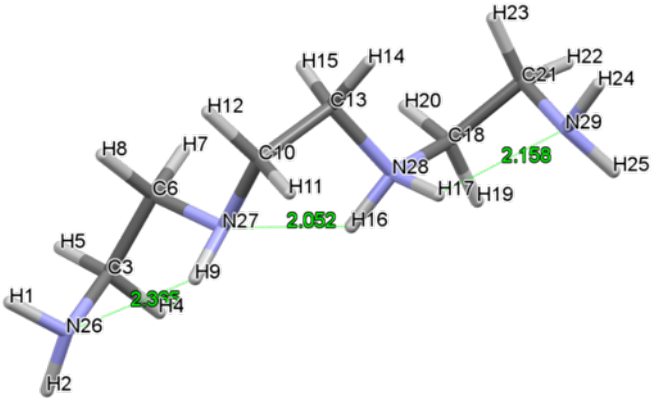
C_s02



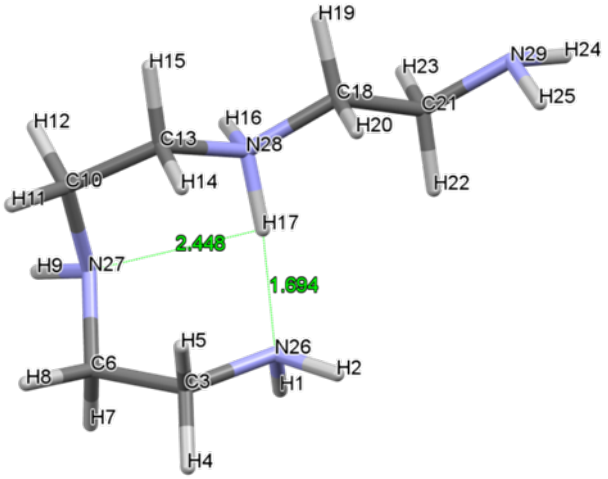
C_s05



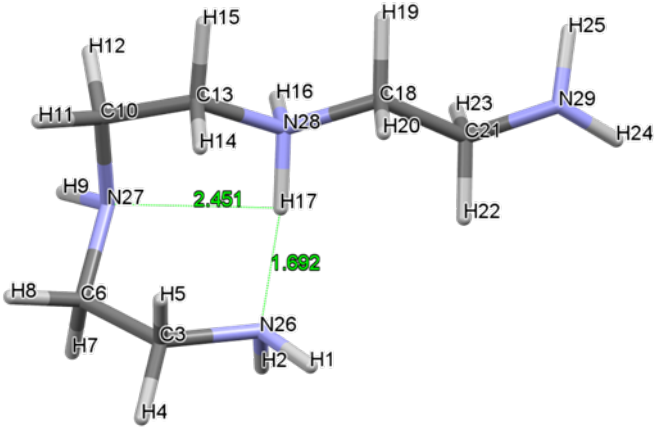
C_s22



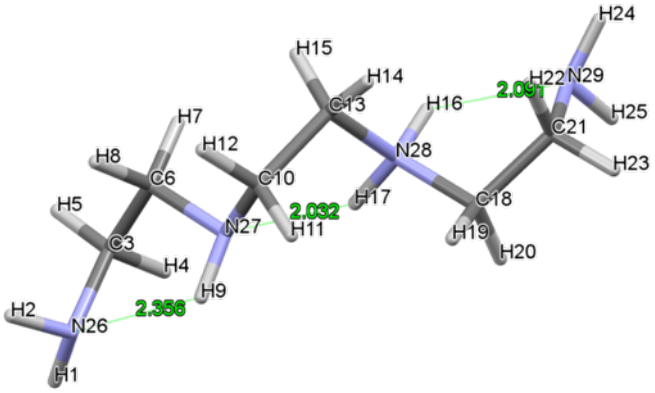
C_s06



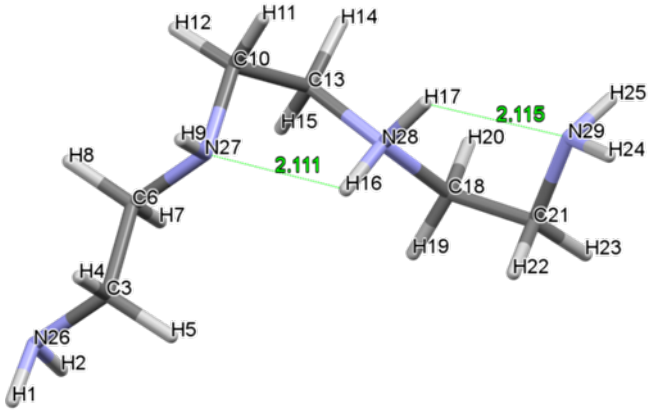
C_s13



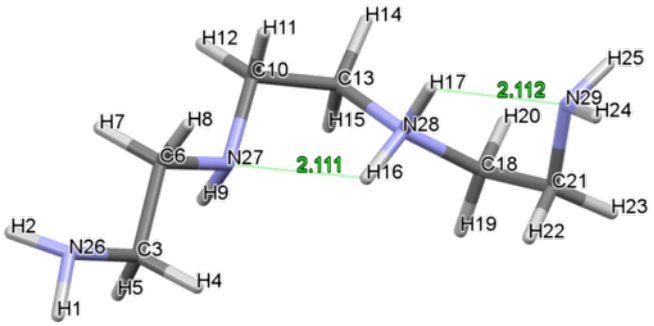
C_s14



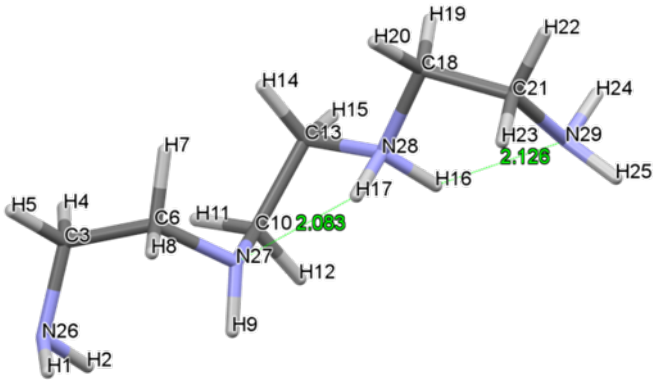
C_s07



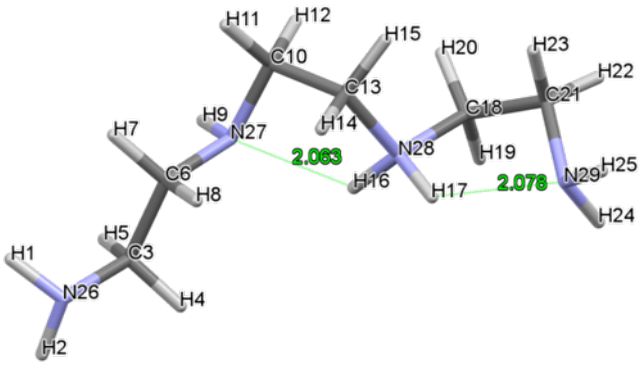
C_s09



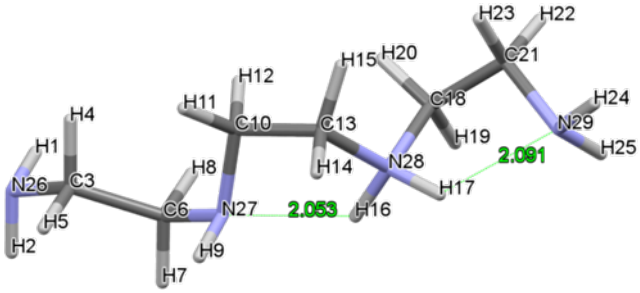
C_s08



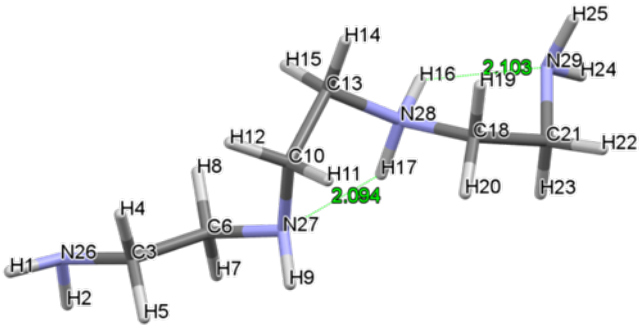
C_s32



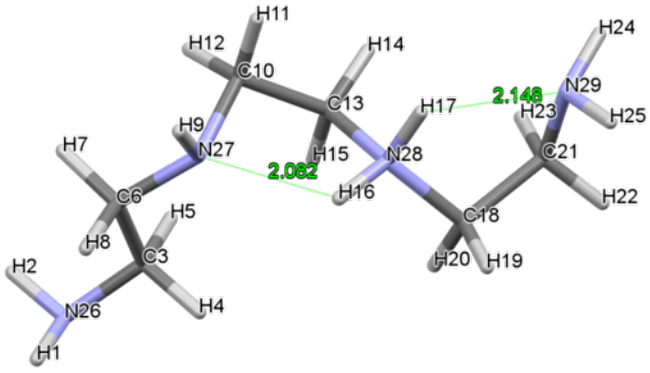
C_s10



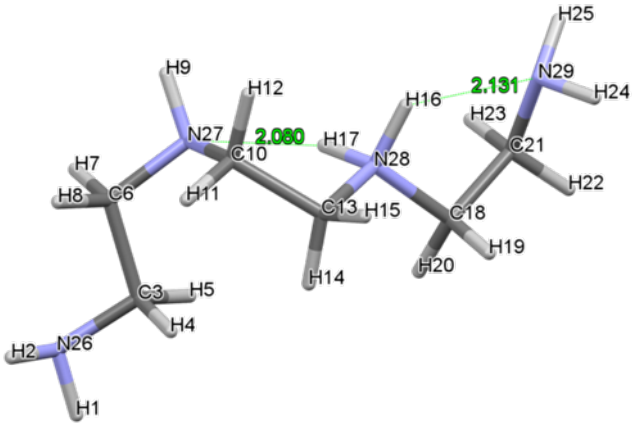
C_s24



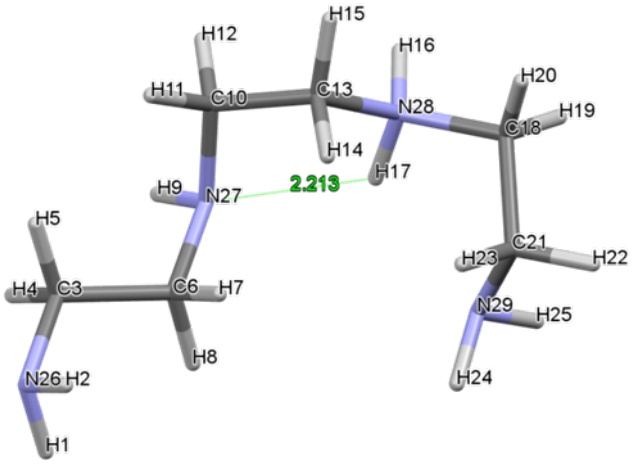
C_s12



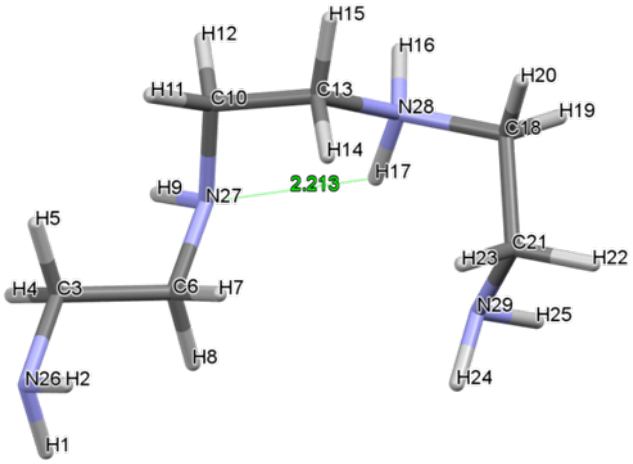
C_s23



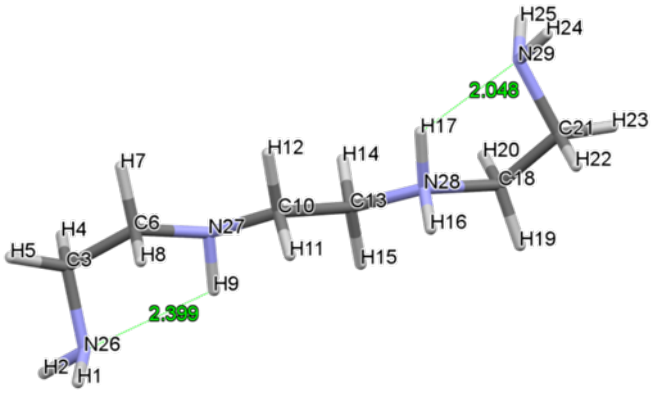
C_s20



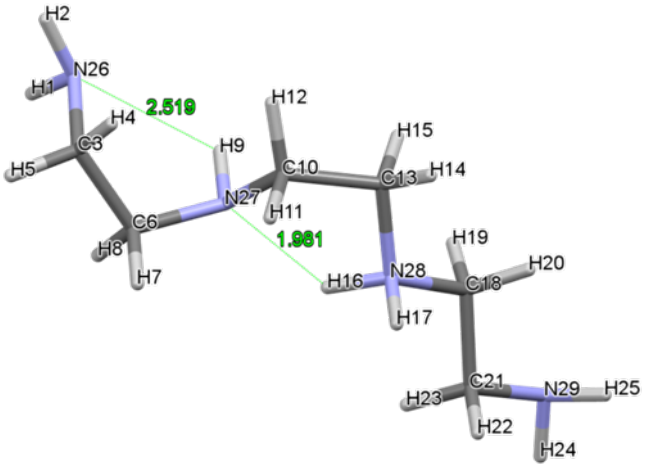
C_s25



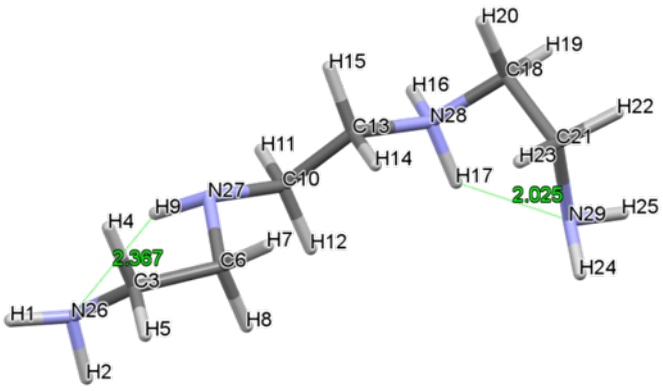
C_s11



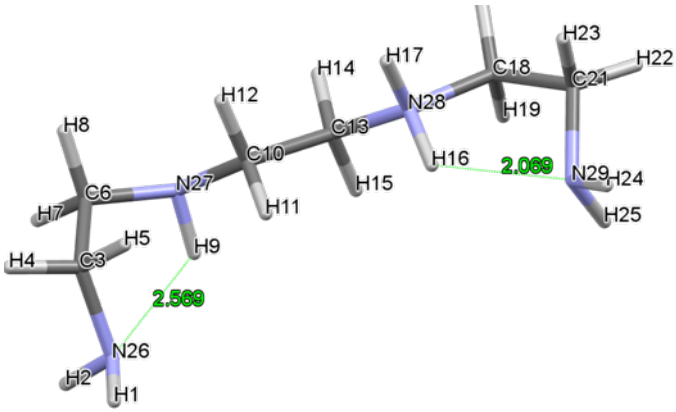
C_s30



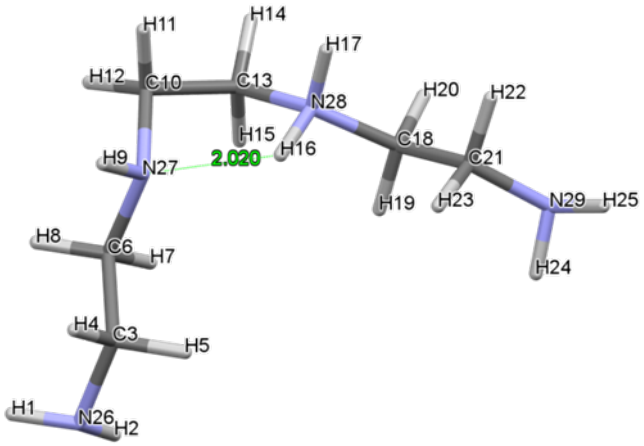
C_s29



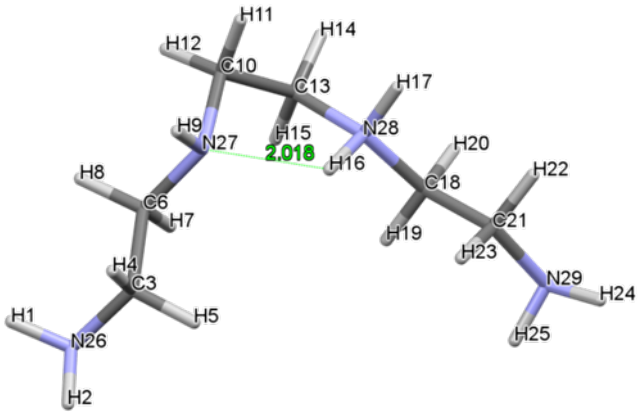
C_s35



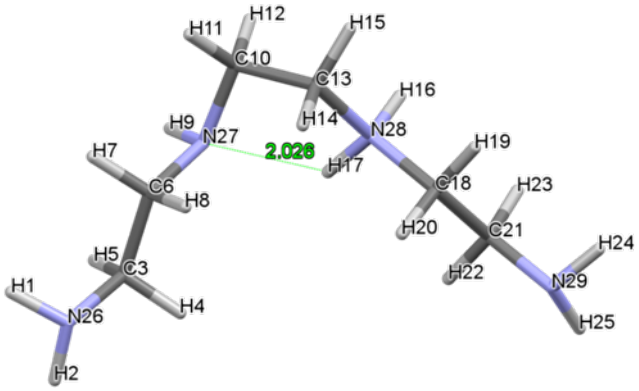
C_s27



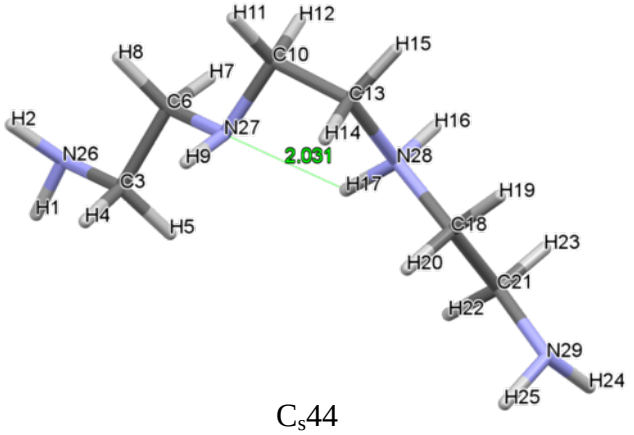
C_s56



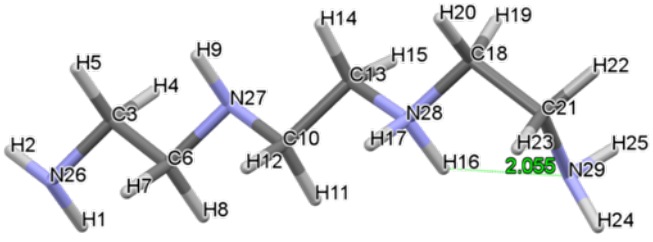
C_s41



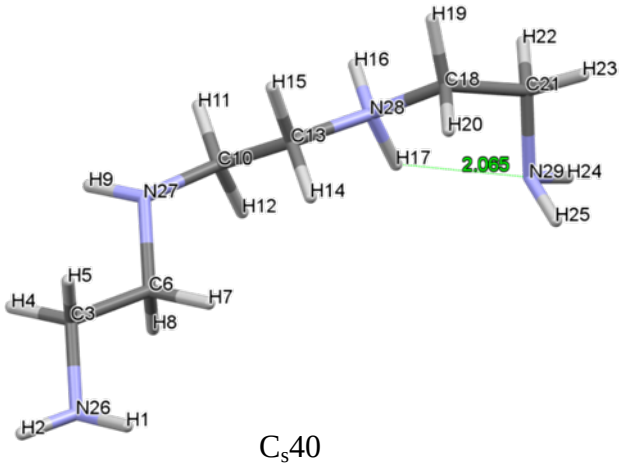
C_s42



C_s44

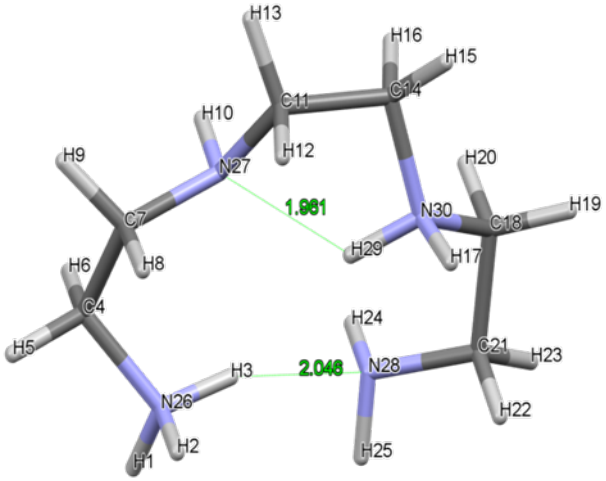


C_s39

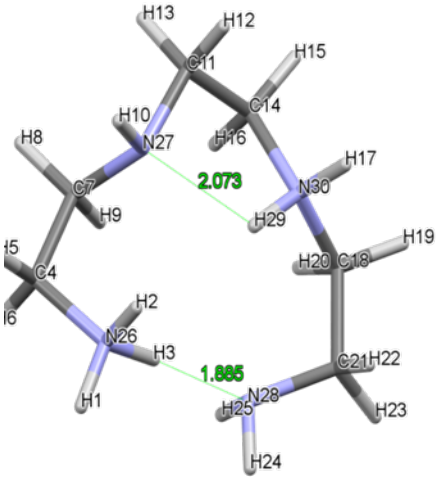


C_s40

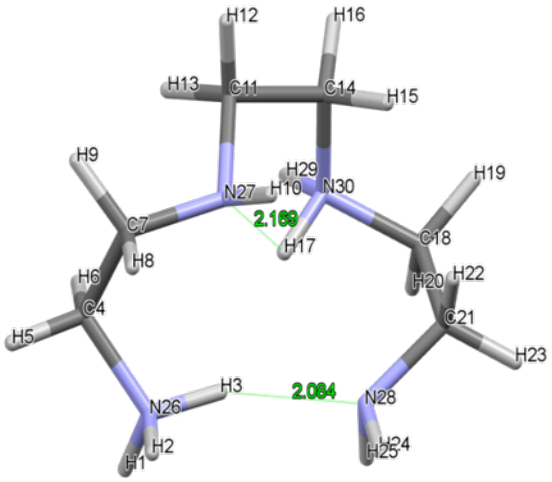
Figure. A3 Capped-stick representation of HL_s structures of 30 LECs generated at MP2 during the fourth and final stage of the conformational protocol developed in this work, also showing atoms' numbering as well as interatomic distances in Å of short NH--N contacts



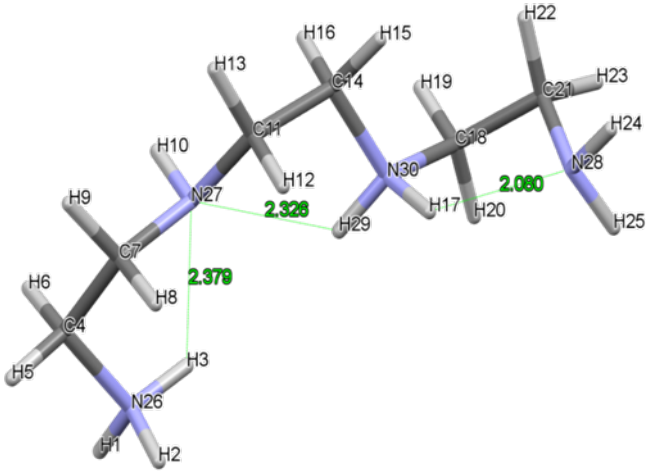
$C_{ps}01$



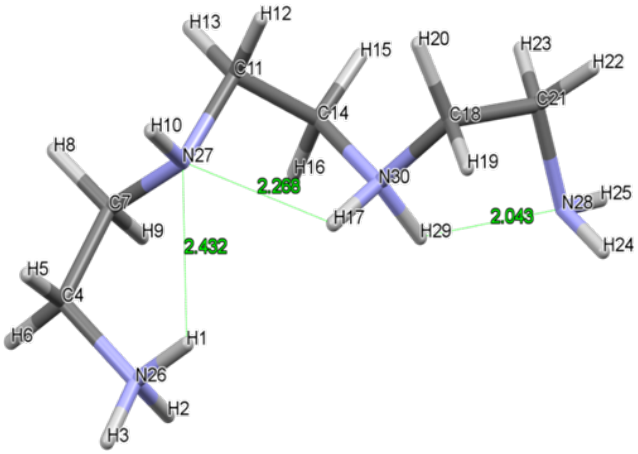
$C_{ps}41$



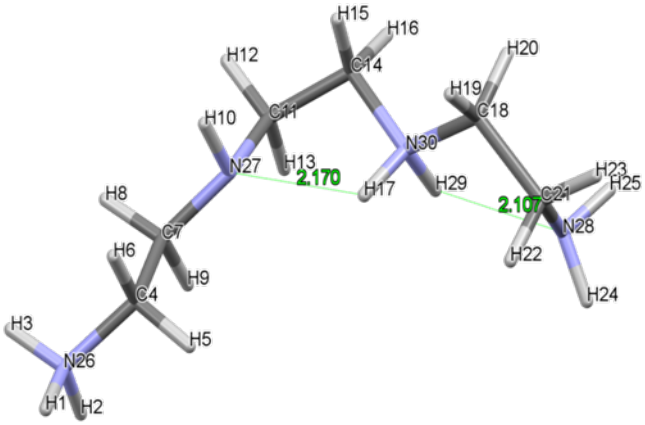
$C_{ps}10$



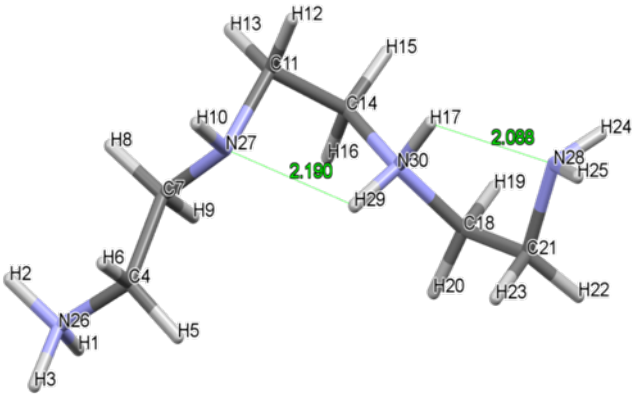
$C_{ps}03$



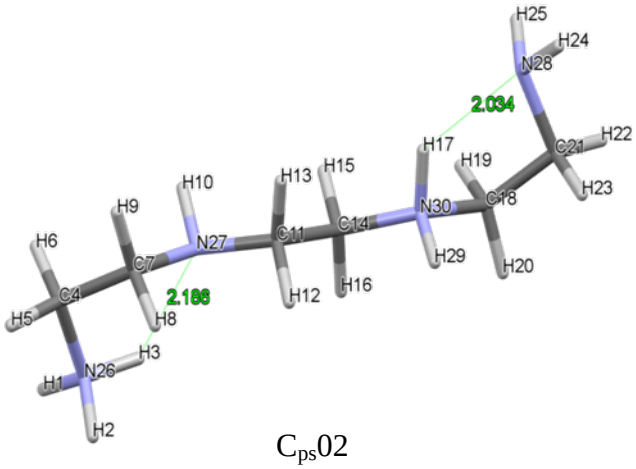
$C_{ps}14$



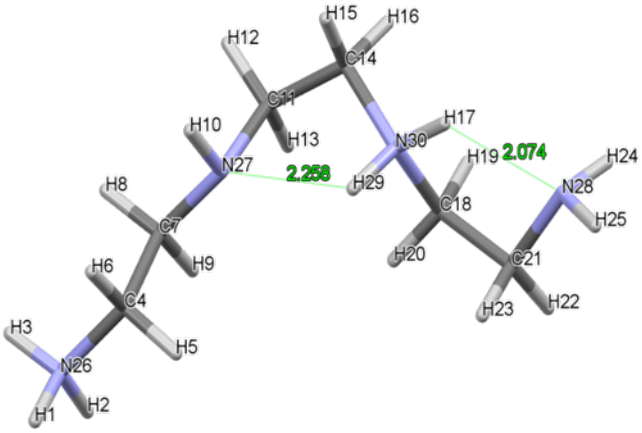
$C_{ps}05$



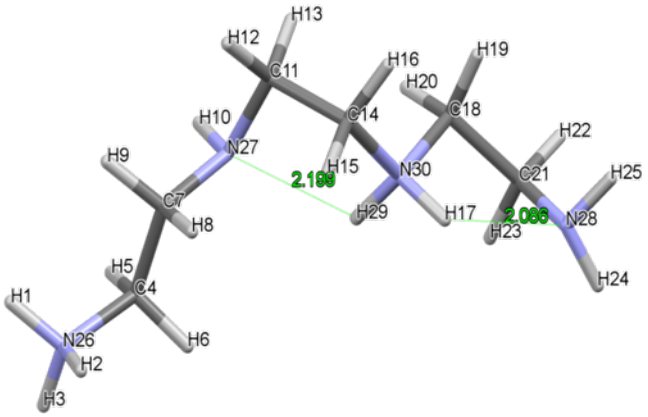
$C_{ps}17$



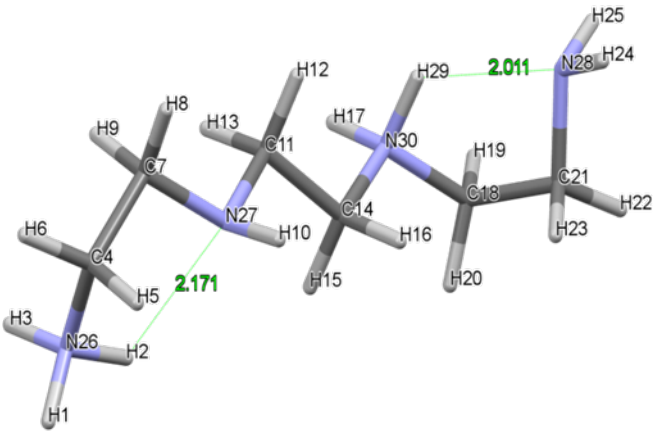
$C_{ps}02$



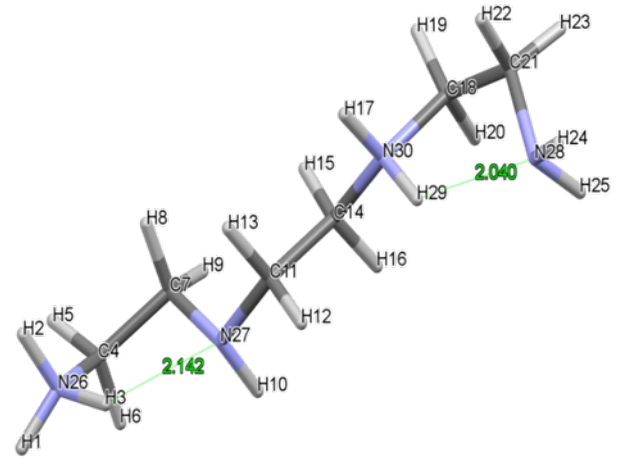
$C_{ps}34$



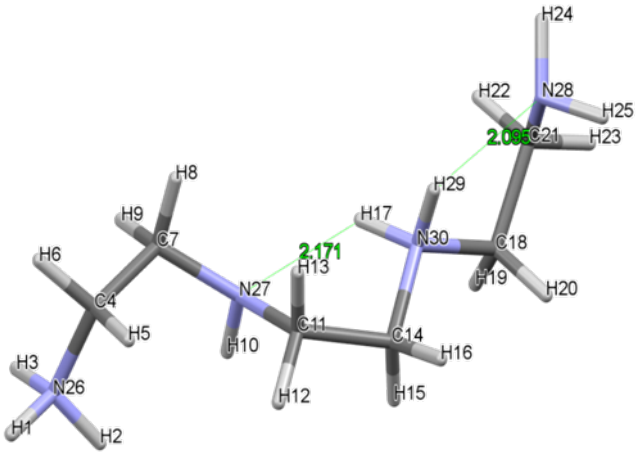
$C_{ps}43$



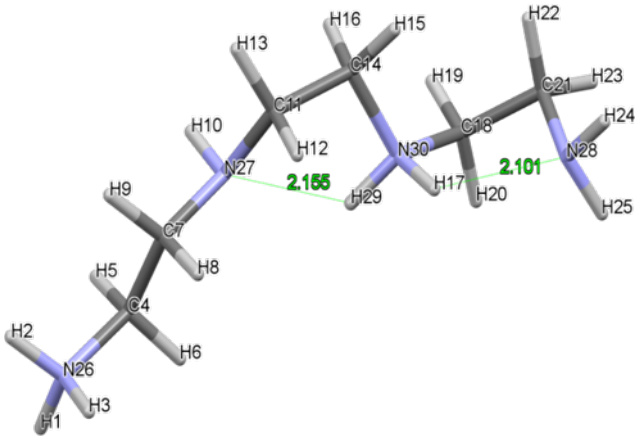
$C_{ps}04$



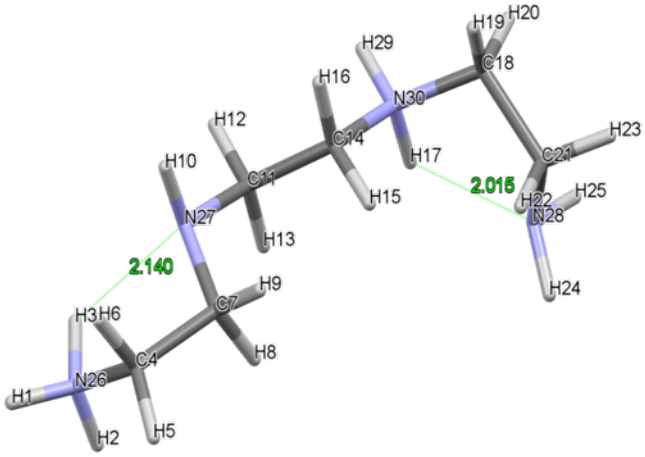
$C_{ps}12$



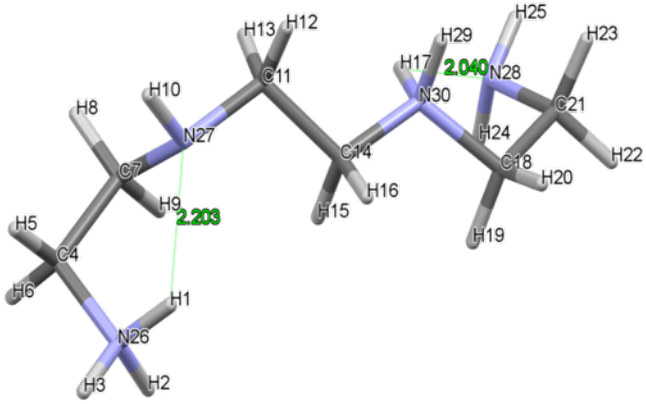
$C_{ps}21$



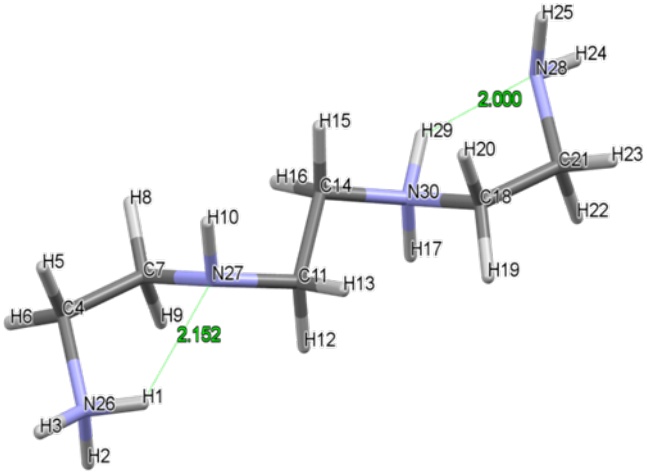
$C_{ps}09$



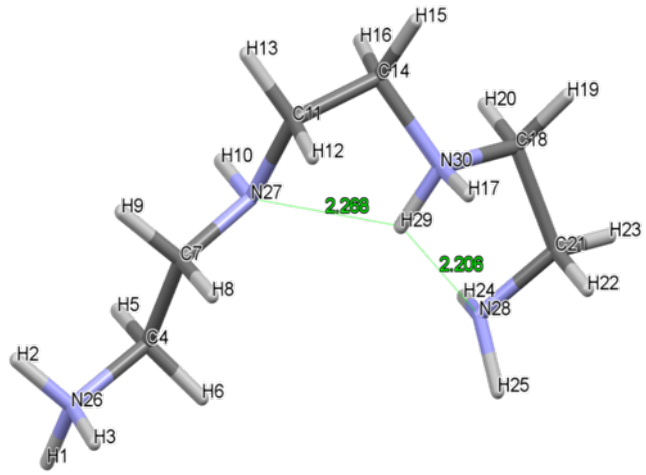
$C_{ps}15$



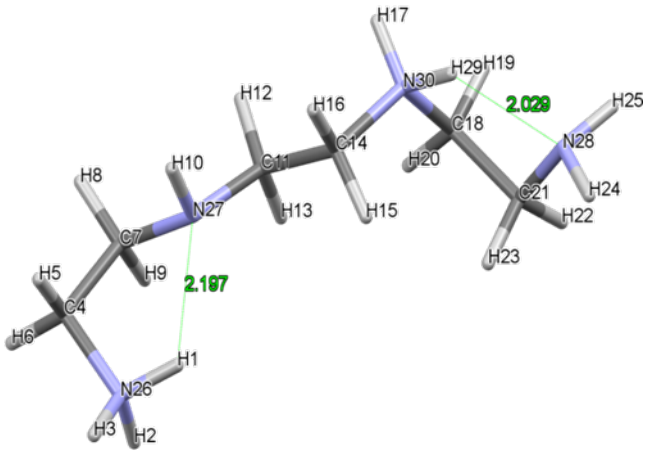
$C_{ps}13$



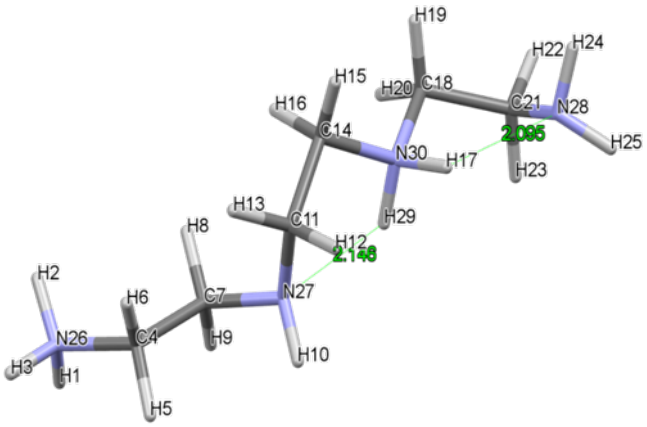
$C_{ps}18$



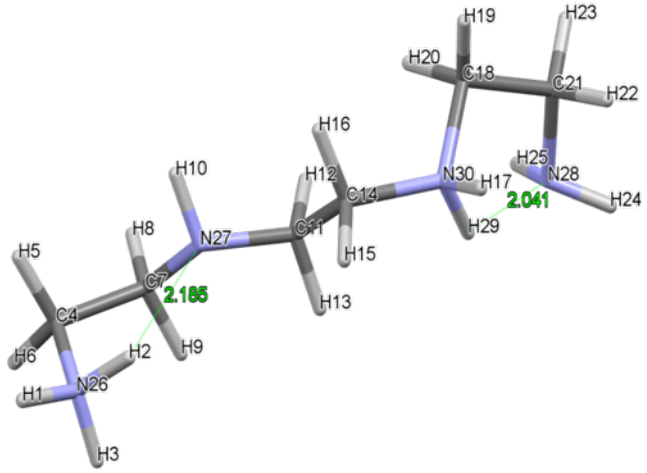
$C_{ps}06$



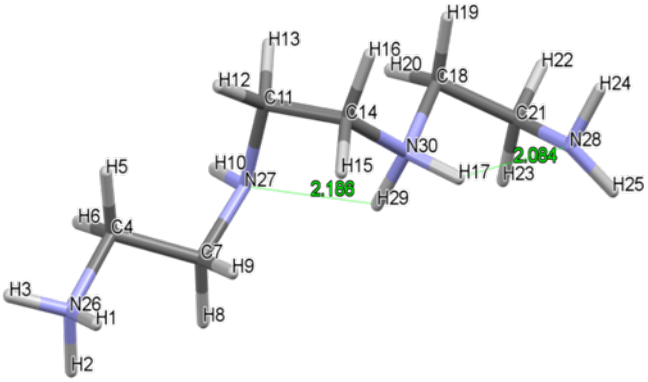
$C_{ps}11$



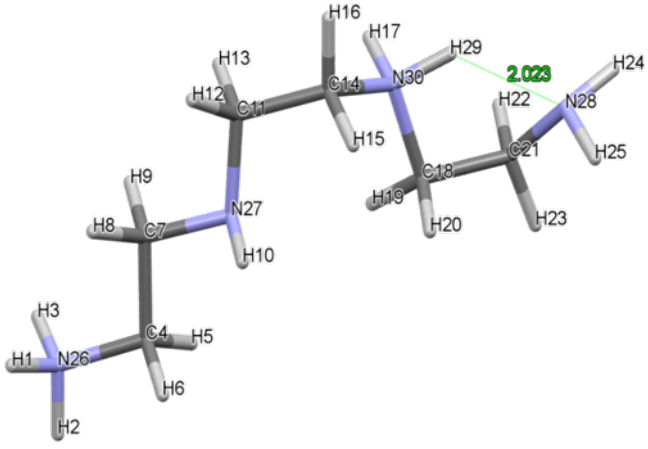
$C_{ps}24$



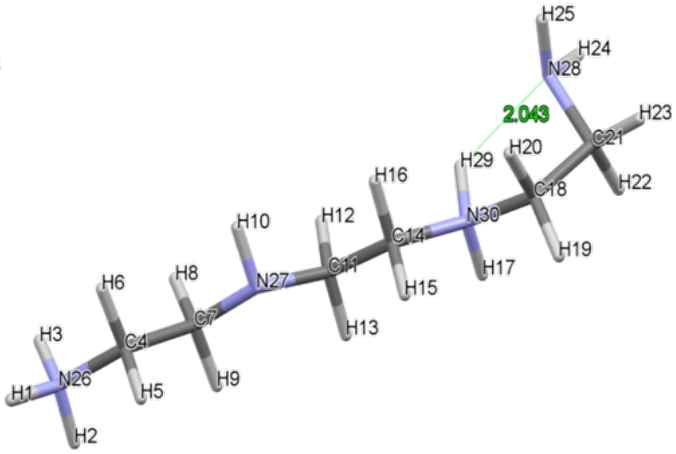
$C_{ps}07$



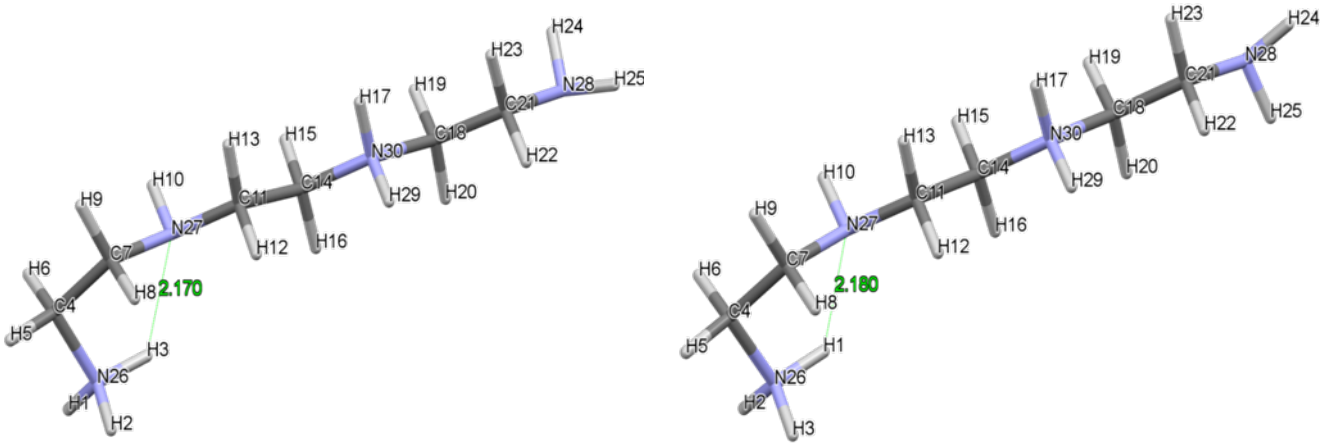
$C_{ps}29$



$C_{ps}22$

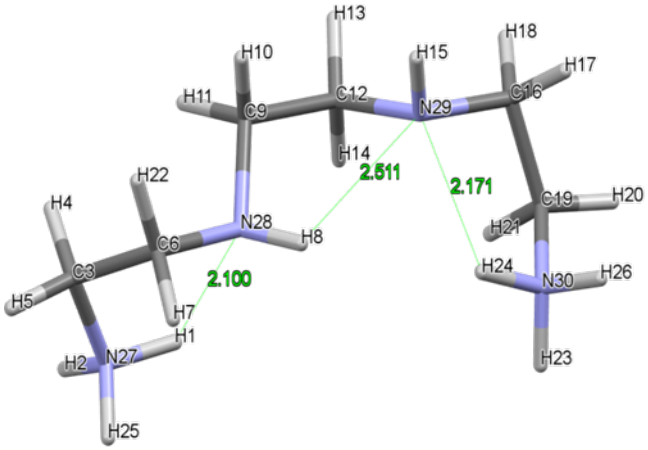


$C_{ps}19$

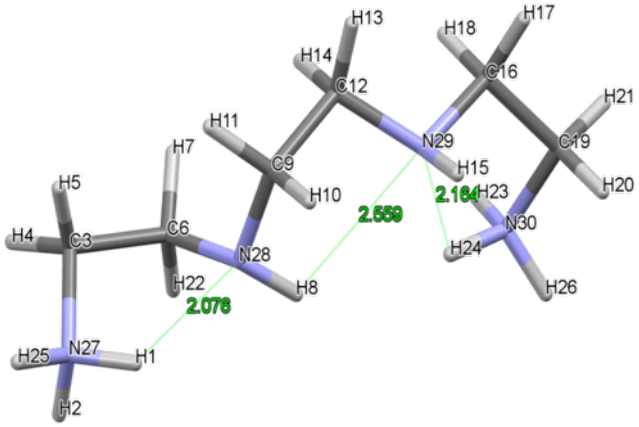


$C_{ps}48$ $C_{ps}49$

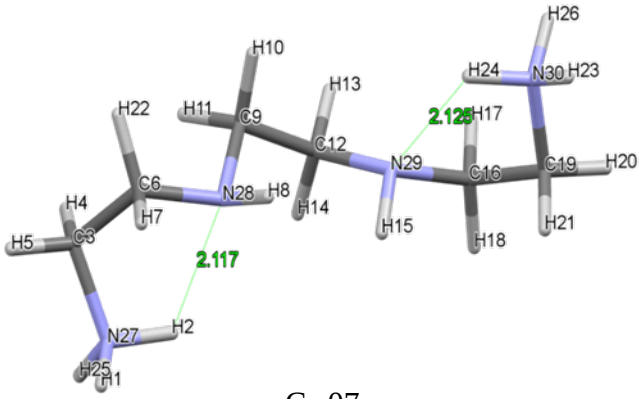
Figure A4 Capped-stick representation of H_2L_{ps} structures of 30 LECs generated at MP2 during the fourth and final stage of the conformational protocol developed in this work, also showing atoms' numbering as well as interatomic distances in Å of short NH--N contacts.



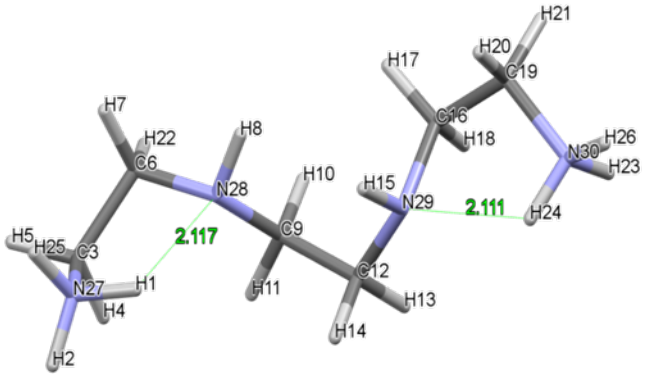
$C_{pp}12$



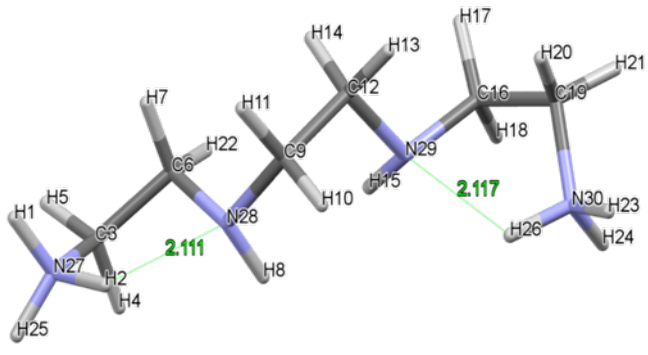
$C_{pp}10$



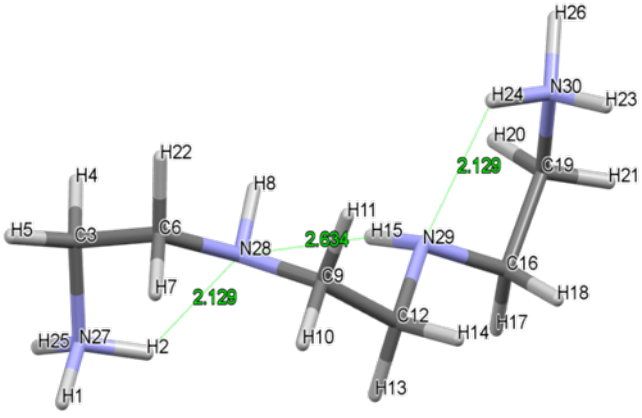
$C_{pp}07$



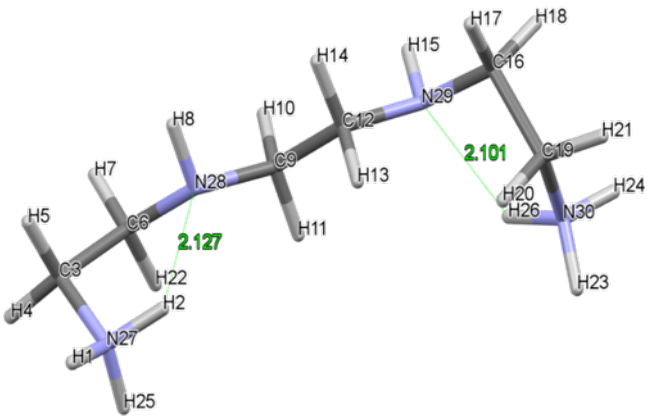
$C_{pp}16$



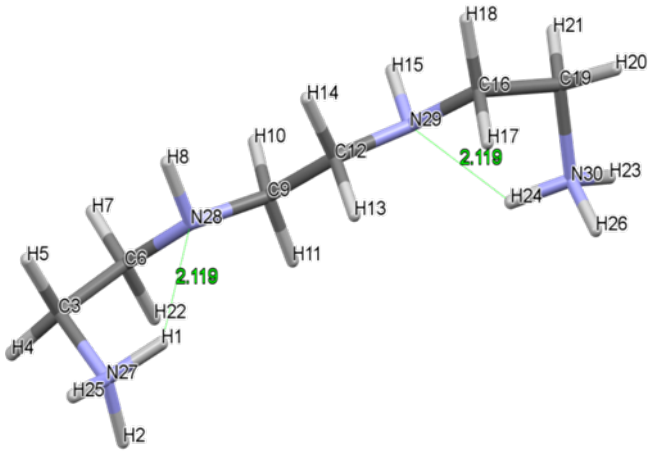
$C_{pp}11$



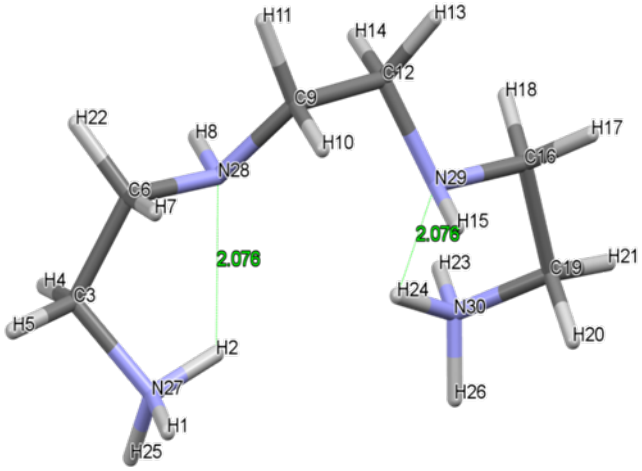
$C_{pp}09$



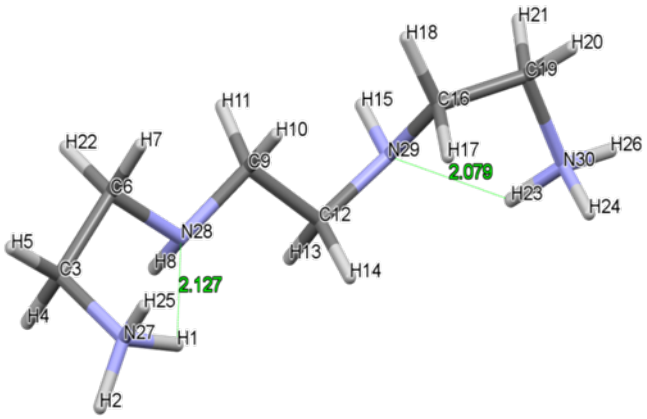
$C_{pp}05$



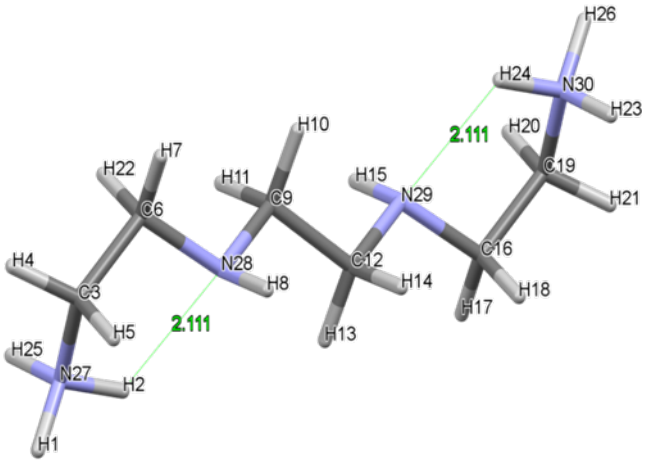
$C_{pp}02$



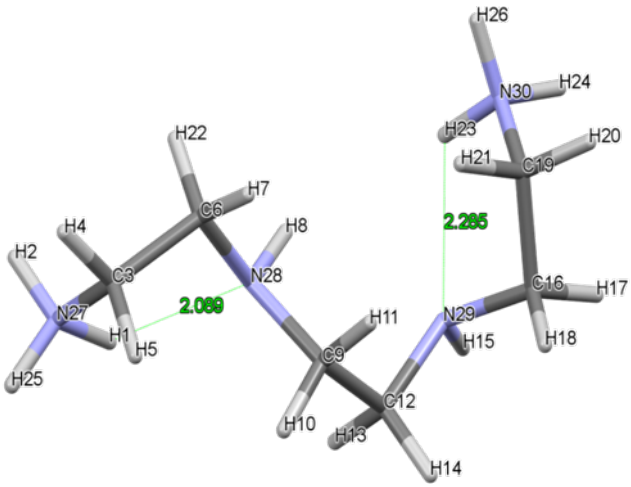
$C_{pp}23$



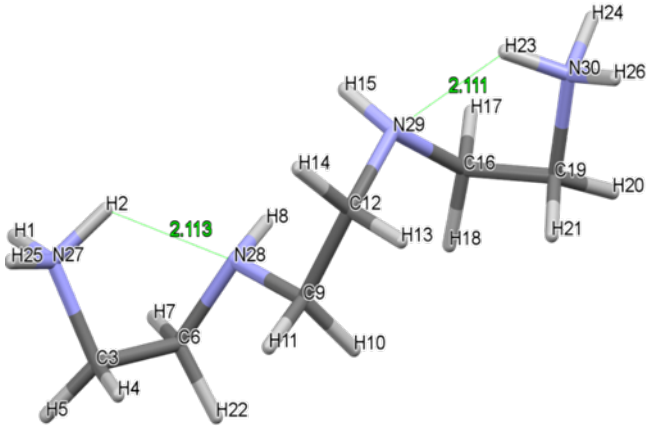
$C_{pp}04$



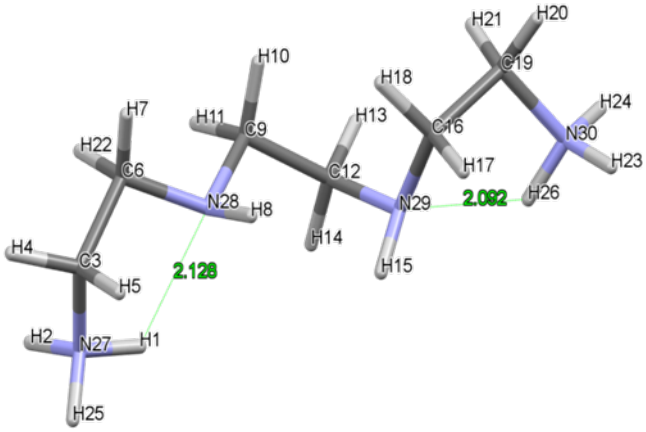
$C_{pp}01$



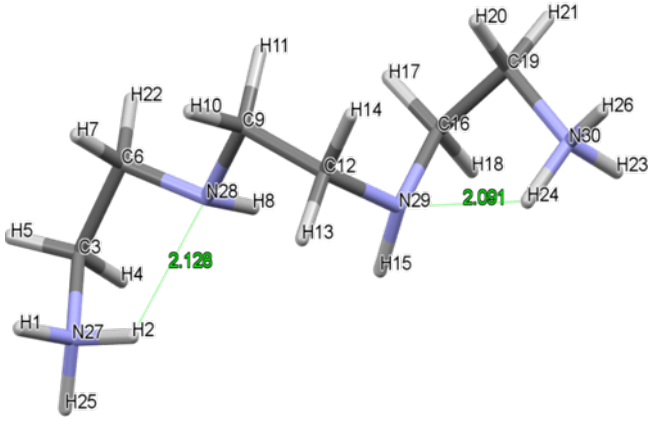
$C_{pp}27$



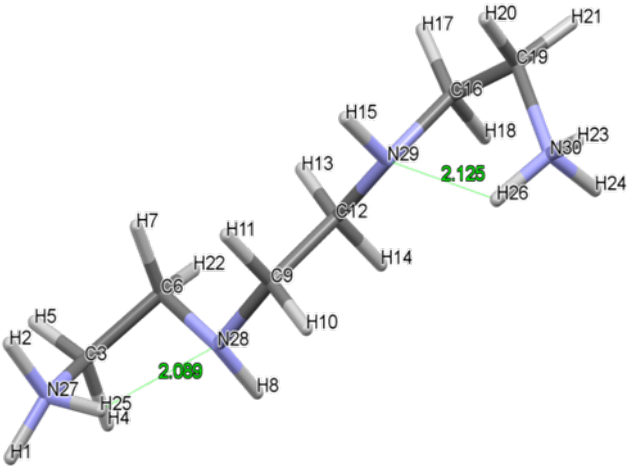
$C_{pp}14$



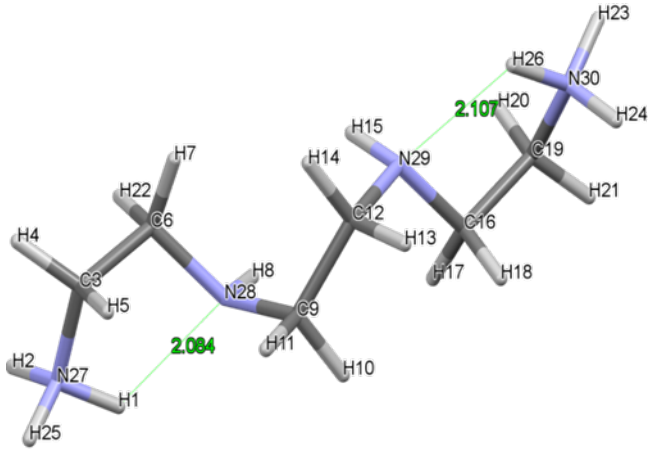
$C_{pp}13$



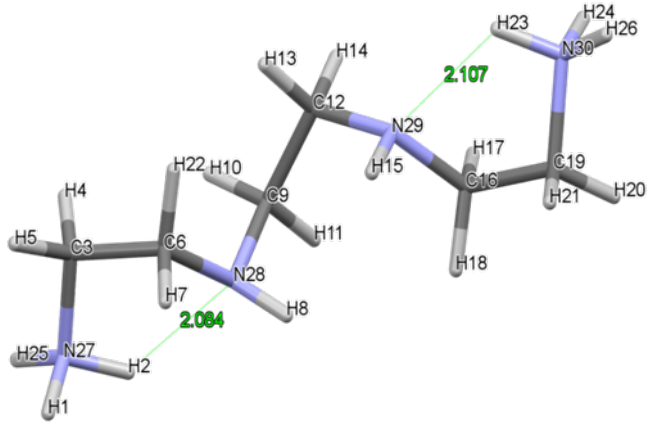
$C_{pp}18$



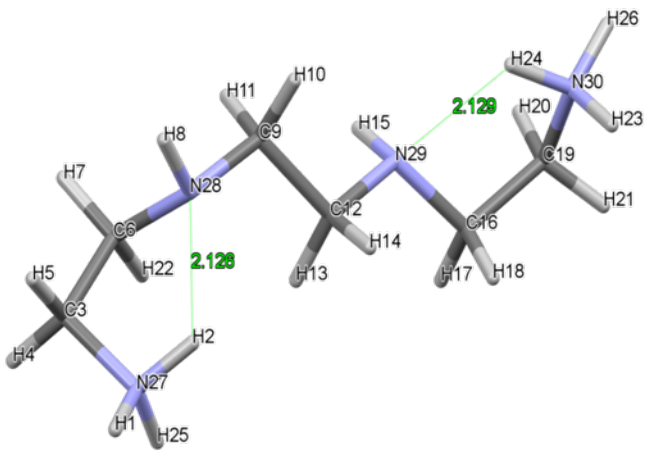
$C_{pp}03$



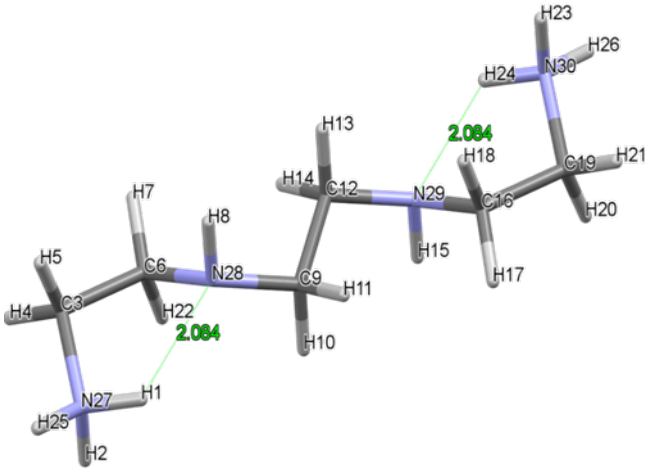
$C_{pp}19$



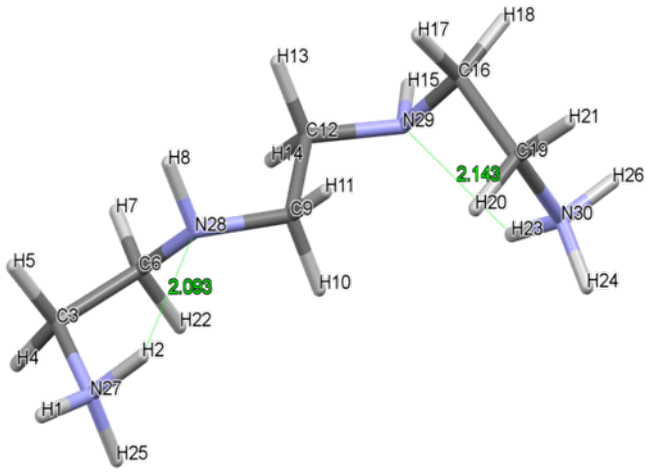
$C_{pp}21$



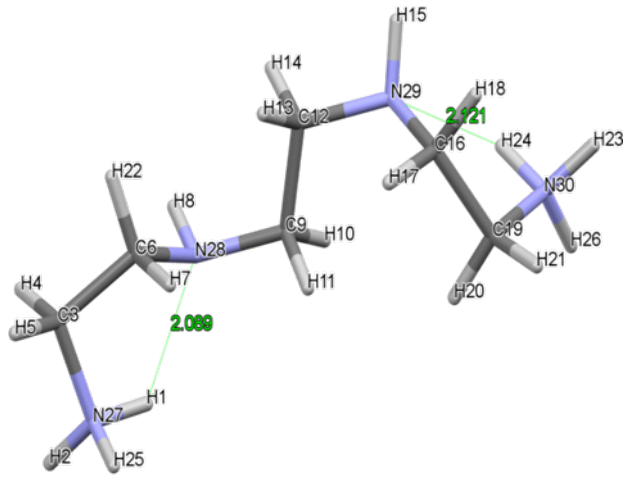
$C_{pp}06$



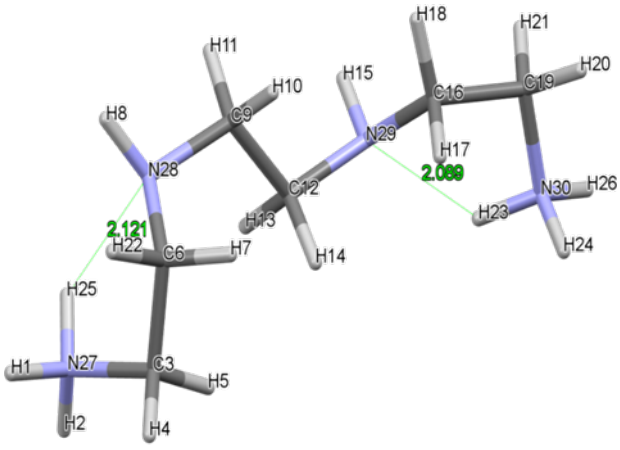
$C_{pp}08$



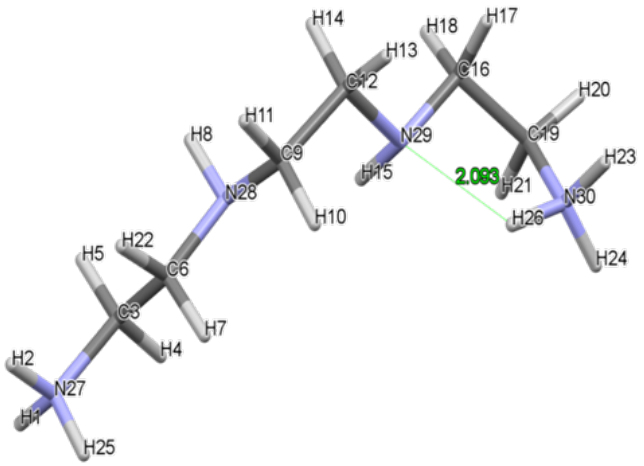
$C_{pp}15$



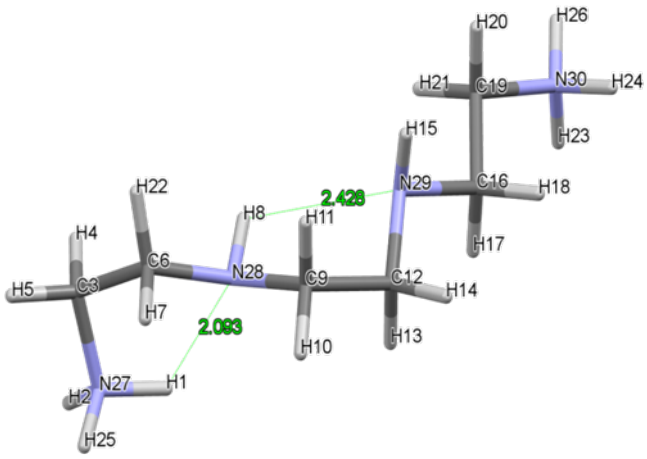
$C_{pp}20$



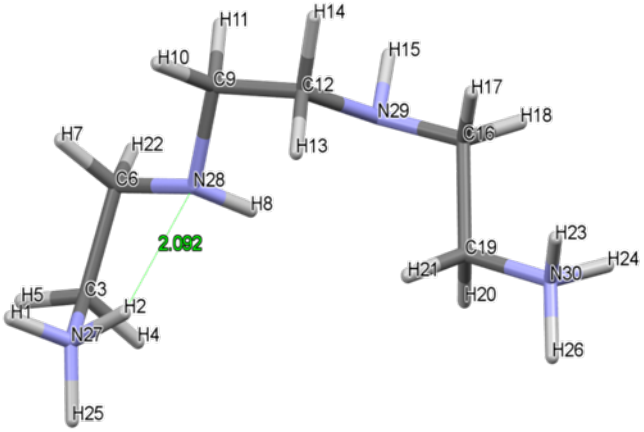
$C_{pp}17$



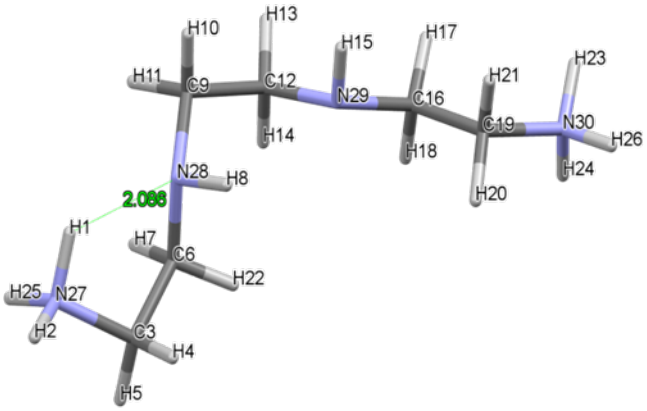
$C_{pp}24$



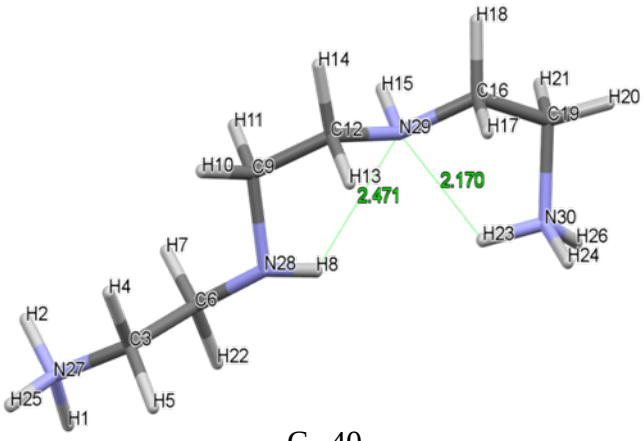
$C_{pp}26$



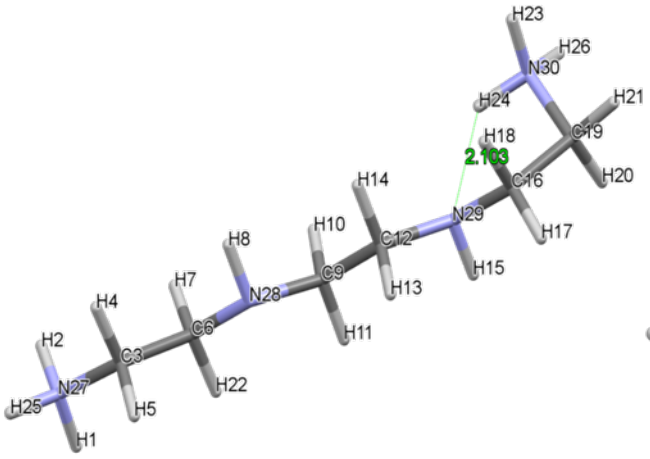
$C_{pp}31$



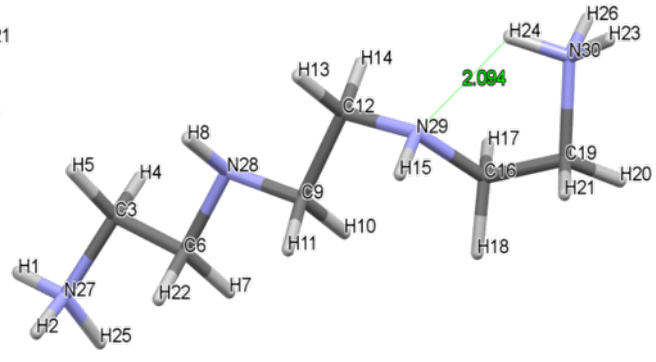
$C_{pp}30$



$C_{pp}40$



$C_{pp}22$



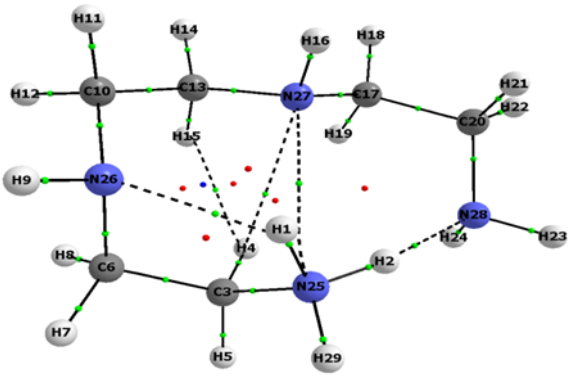
$C_{pp}25$

Figure A5. Capped-stick representation of H_2L_{pp} structures of 30 LECs generated at MP2 during the fourth and final stage of the conformational protocol developed in this work, also showing atoms' numbering as well as interatomic distances in Å of short NH--N

Table A3. Interatomic distance and electron density at a CP of interactions found at MP2 in 15 LECs of HL_p also showing short contacts without AILs on a relevant molecular graph

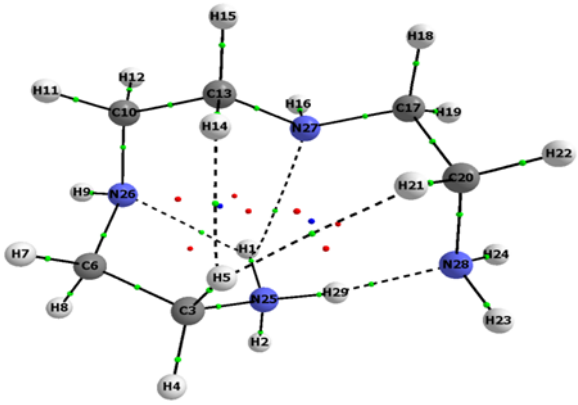
Conformer C_p02

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N28	H2	1.750	0.0507
	N26	H1	2.040	0.0287
CH•••N	N27	H4	2.601	0.0115
CH•••HC	H4	H15	2.124	0.0105
N•••N	N25	N27	2.897	0.0147
Contact without AIL				
NH--N	N27	H1	2.478	–



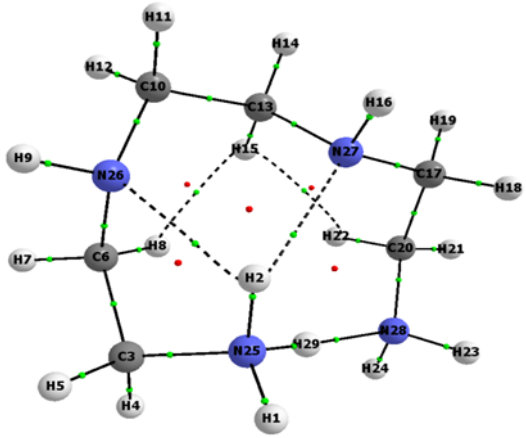
Conformer C_p01

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N28	H29	1.740	0.0521
	N26	H1	2.073	0.0269
	N27	H1	2.398	0.0154
CH•••HC	H5	H14	2.048	0.0113
	H5	H21	2.472	0.0046
Contact without AIL				
CH--HC	H14	H21	2.299	–



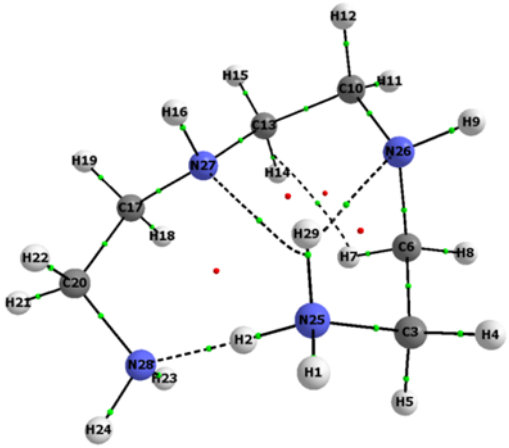
Conformer C_p03

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N28	H29	1.741	0.0517
	N26	H2	2.136	0.0241
	N27	H2	2.379	0.0153
CH•••HC	H8	H15	2.034	0.0134
	H15	H22	2.133	0.0105
Contact without AIL				
None				



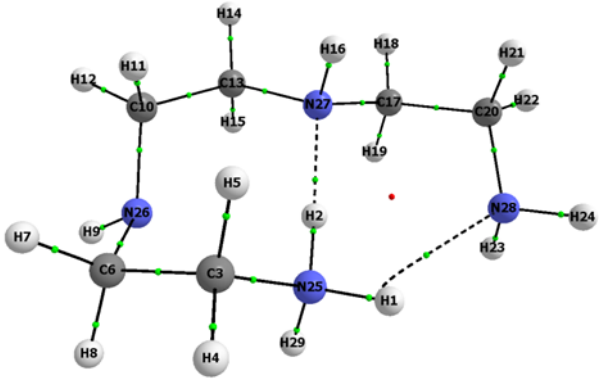
Conformer C_p04

Interaction	Atoms A	Atoms B	d(A,B) Å	ρ_{CP} a.u.
NH•••N	N28	H2	1.740	0.0518
	N26	H29	2.101	0.0258
	N27	H29	2.512	0.0126
CH•••C	H7	C13	2.466	0.0140
Contact without AIL				
CH--HC	H7	H14	2.047	–



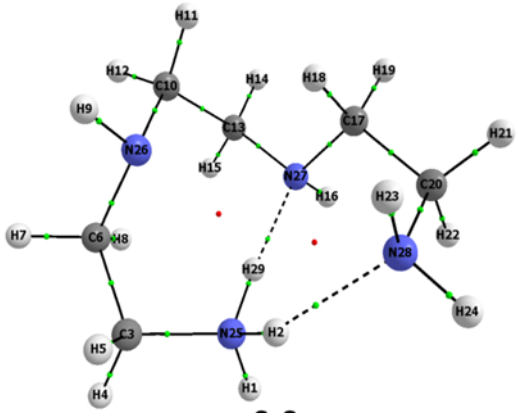
Conformer C_p05

Interaction	Atoms A	Atoms B	d(A,B) Å	ρ_{CP} a.u.
NH•••N	N27	H2	1.680	0.0614
	N28	H1	2.470	0.0136
Contact without AIL				
NH--N	N26	H2	2.482	–
	N28	H2	2.646	–



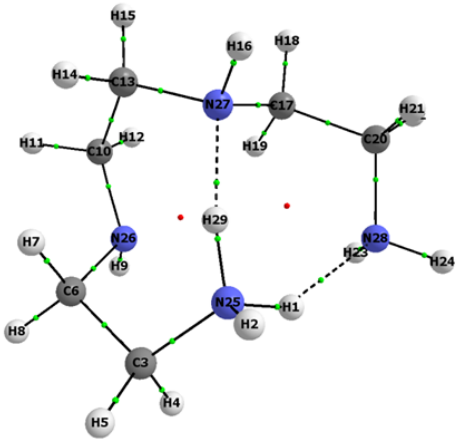
Conformer C_p06

Interaction	Atoms A	Atoms B	d(A,B) Å	ρ_{CP} a.u.
NH•••N	N27	H29	1.691	0.0599
	N28	H2	1.944	0.0336
Contact without AIL				
NH--N	N26	H29	2.473	–
CH--HC	H8	H15	2.225	–
	H11	H18	2.225	–



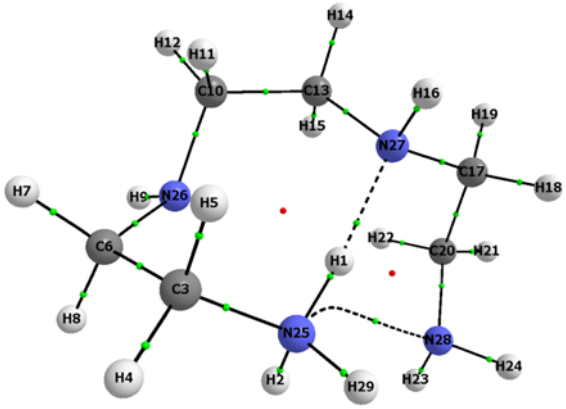
Conformer C_p07

Interaction	Atoms A	Atoms B	d(A,B) Å	ρ_{CP} a.u.
NH•••N	N27	H29	1.691	0.0599
	N28	H2	1.944	0.0336
Contact without AIL				
NH--N	N26	H29	2.473	–
CH--HC	H8	H15	2.225	–
	H11	H18	2.225	–



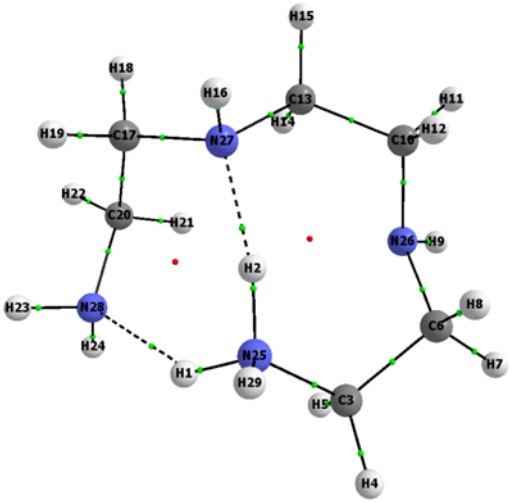
Conformer C_p08

Interaction	Atoms A	Atoms B	d(A,B) Å	ρ_{CP} a.u.
NH•••N	N27	H1	1.717	0.0559
N•••N	N25	N28	2.978	0.0125
Contact without AIL				
NH--N	N26	H1	2.487	–
CH--HC	H15	H22	2.189	–
	H5	H11	2.282	–



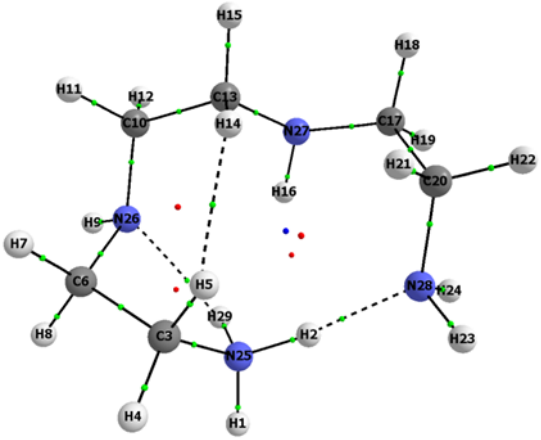
Conformer C_p09

Interaction	Atoms A	Atoms B	d(A,B) Å	ρ_{CP} a.u.
NH•••N	N27	H2	1.689	0.0601
	N28	H1	1.989	0.0308
Contact without AIL				
NH--N	N26	H1	2.533	–
CH--HC	H15	H22	2.268	–



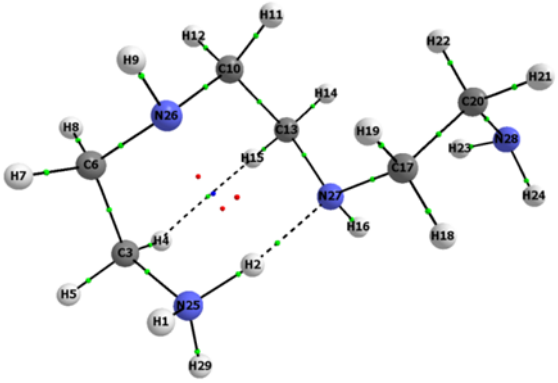
Conformer C_p20

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N28	H2	1.708	0.0564
	N26	H29	2.112	0.0255
CH•••HC	H5	H14	2.048	0.0110
Contact without AIL				
NH--N	N26	H16	2.513	–
CH--HC	H14	H21	2.186	–



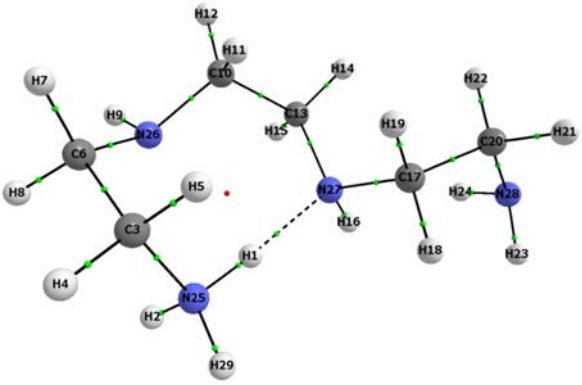
Conformer C_p23

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N27	H2	1.661	0.0642
CH•••HC	H4	H15	2.179	0.0089
Contact without AIL				
NH--N	N26	H2	2.478	–
CH--HC	H14	H22	2.274	–
	H11	H19	2.363	–



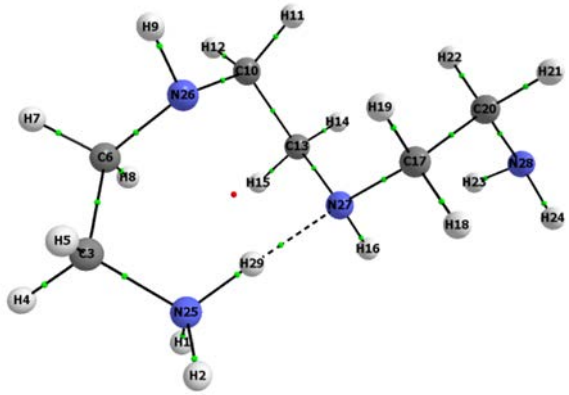
Conformer C_p29

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N27	H1	1.694	0.0595
Contact without AIL				
NH--N	N26	H1	2.481	–
CH--HC	H14	H22	2.155	–
	H11	H19	2.349	–



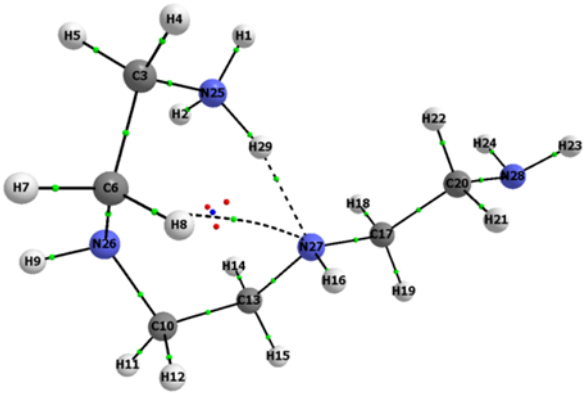
Conformer C_p26

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N27	H29	1.671	0.0627
Contact without AIL				
NH--N	N26	H29	2.480	–
	N26	H19	2.688	–
CH--HC	H8	H15	2.255	–
	H14	H22	2.304	–
	H11	H19	2.350	–



Conformer C_p15

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N27	H29	1.690	0.0603
CH•••N	H8	N27	2.708	0.0092
Contact without AIL				
NH--N	N26	H2	2.526	–
	N26	H29	2.658	–



Conformer C_p24

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N27	H1	1.654	0.0653
CH•••HC	H4	H15	2.157	0.0093
Contact without AIL				
NH--N	N26	H1	2.490	–
CH--HC	H14	H21	2.273	–
	H11	H19	2.344	–

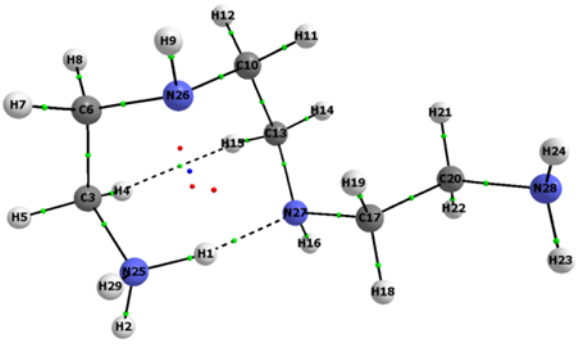
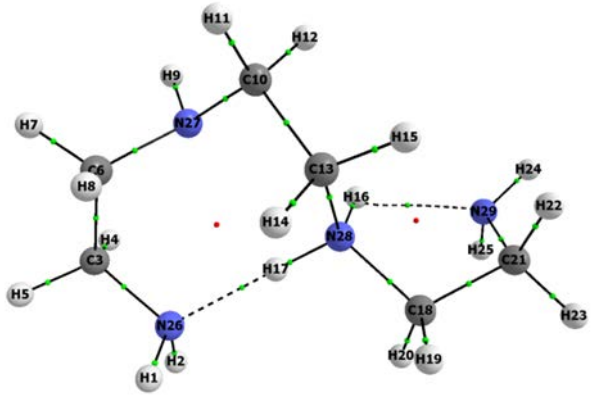


Table A4. Interatomic distance and electron density at a CP of interactions found at MP2 in 15 LECs of HL_s also showing short contacts without AILs on a relevant molecular graph.

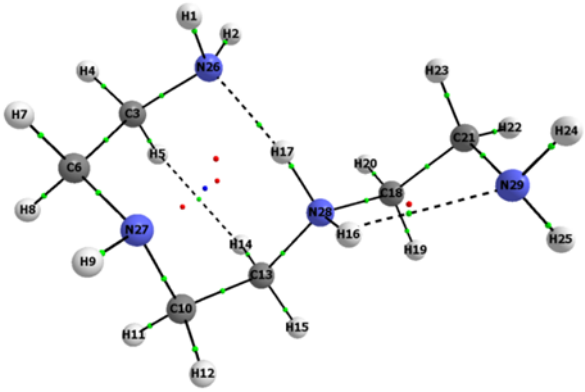
Conformer C_s04

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N26	H17	1.745	0.0521
	N29	H16	2.177	0.0221
Contact without AIL				
NH--N	N27	H17	2.477	–
CH--HC	H15	H22	2.277	–
	H8	H14	2.319	–



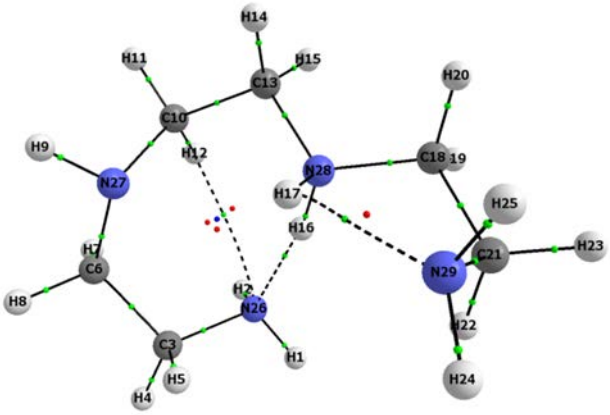
Conformer C_s01

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N26	H17	1.728	0.0542
	N29	H16	2.217	0.0205
CH•••HC	H5	H14	2.136	0.0096
Contact without AIL				
NH--N	N27	H17	2.436	–
CH--HC	H15	H19	2.373	–



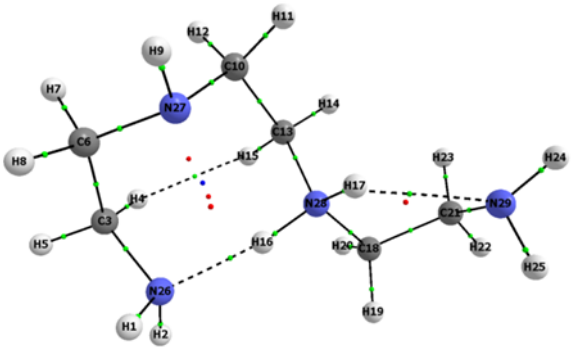
Conformer C_s03

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N26	H16	1.743	0.0526
	N29	H17	2.197	0.0213
CH•••N	N26	H12	2.691	0.0092
Contact without AIL				
NH--N	N27	H17	2.546	–
	N27	H16	2.654	–
CH--HC	H7	H12	2.251	–



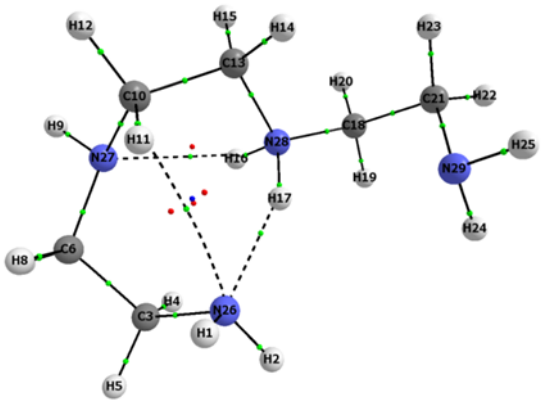
Conformer C_s02

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N26	H16	1.728	0.0540
	N29	H17	2.179	0.0220
CH•••HC	H4	H15	2.136	0.0096
Contact without AIL				
NH--N	N27	H16	2.461	–
CH--HC	H14	H23	2.278	–



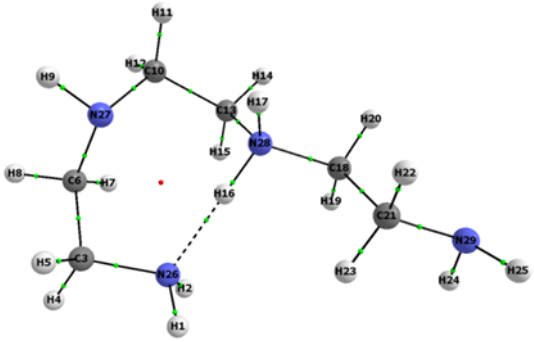
Conformer C_s05

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N26	H17	1.845	0.0418
	N27	H16	2.253	0.0210
CH•••N	N26	H11	2.679	0.0093
Contact without AIL				
NH--N	N29	H17	2.526	–
CH--HC	H14	H23	2.208	–
	H8	H11	2.251	–



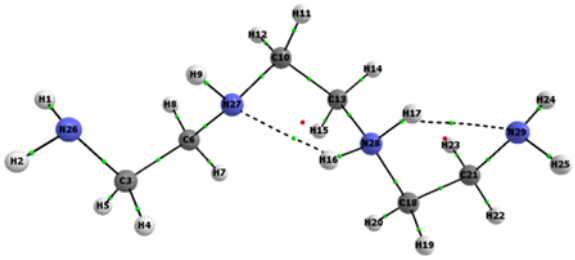
Conformer C_s22

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N26	H16	1.719	0.0556
Contact without AIL				
NH--N	N27	H16	2.474	–
CH--HC	H7	H15	2.336	–



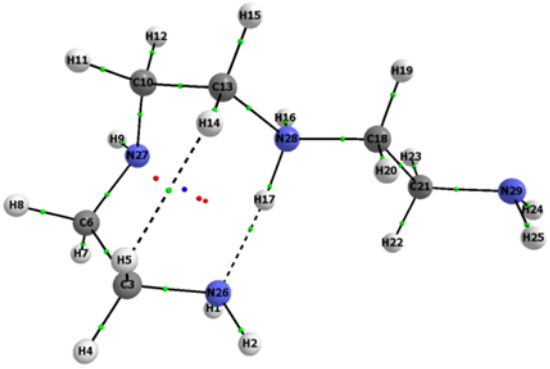
Conformer C_s06

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N27	H16	2.052	0.0285
	N29	H17	2.158	0.0232
Contact without AIL				
NH--N	N26	H9	2.365	–
CH--HC	H7	H15	2.392	–



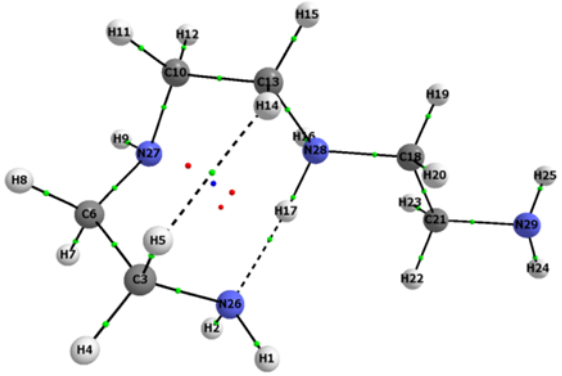
Conformer C_s13

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N26	H17	1.694	0.0588
CH•••HC	H5	H14	2.135	0.0097
Contact without AIL				
NH--N	N27	H17	2.448	–



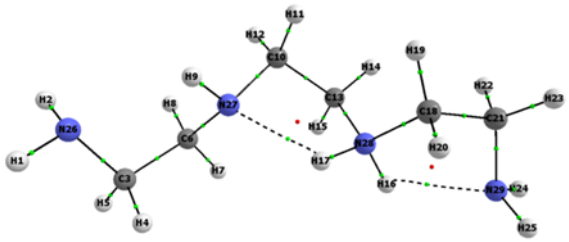
Conformer C_s14

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N26	H17	1.692	0.0590
CH•••HC	H5	H14	2.137	0.0097
Contact without AIL				
NH--N	N27	H17	2.451	–



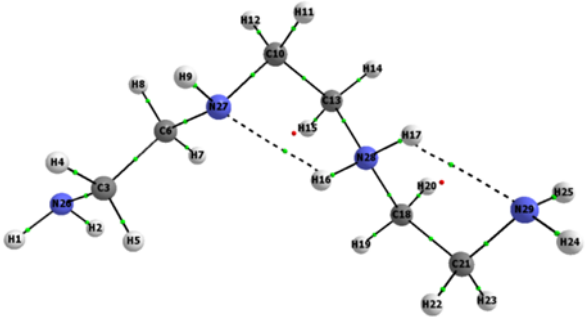
Conformer C_s07

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N27	H17	2.032	0.0298
	N29	H16	2.091	0.0259
Contact without AIL				
NH--N	N26	H9	2.356	–
CH--HC	H14	H22	2.338	–
	H11	H19	2.361	–



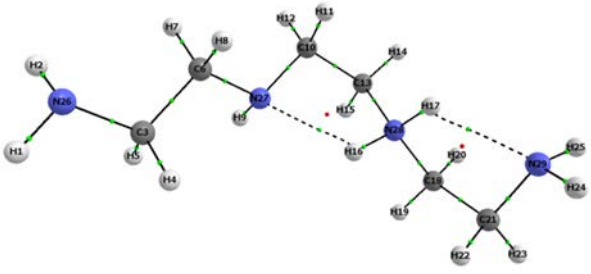
Conformer C_s09

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N27	H16	2.111	0.0255
	N29	H17	2.115	0.0246
Contact without AIL				
CH--HC	H7	H15	2.321	–



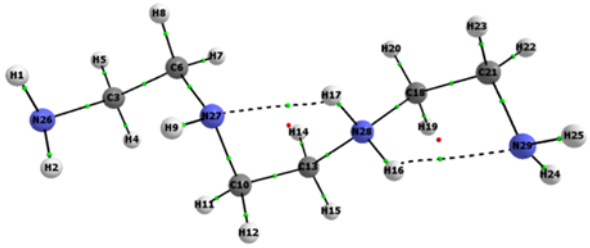
Conformer C_s08

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N27	H16	2.111	0.0253
	N29	H17	2.112	0.0248
Contact without AIL				
None				



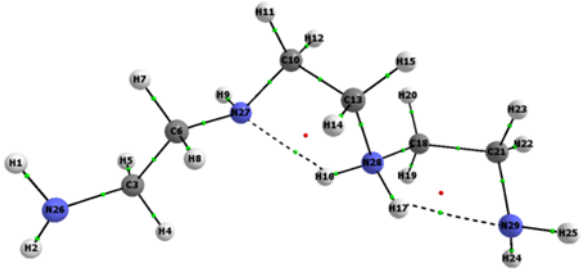
Conformer C_s32

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N27	H17	2.083	0.0270
	N29	H16	2.126	0.0242
Contact without AIL				
CH--HC	H4	H11	2.178	–
	H7	H14	2.325	–



Conformer C_s10

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N27	H16	2.063	0.0279
	N29	H17	2.078	0.0265
Contact without AIL				
CH--HC	H8	H14	2.319	–
	H15	H23	2.344	–
	H12	H20	2.360	–



Conformer C_s24

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N27	H16	2.053	0.0286
	N29	H17	2.091	0.0259
Contact without AIL				
CH--HC	H8	H14	2.319	–
	H12	H20	2.360	–

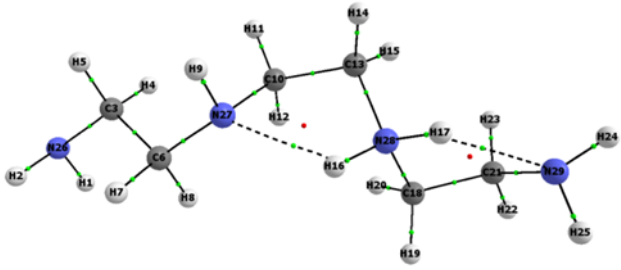
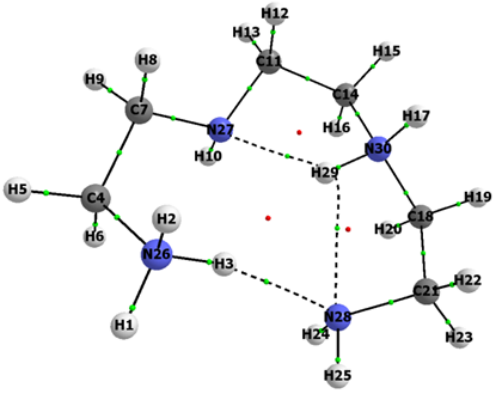


Table A5. Interatomic distance and electron density at a CP of interactions found at MP2 in 15 LECs of HL_{ps} also showing short contacts without AILs on a relevant molecular graph

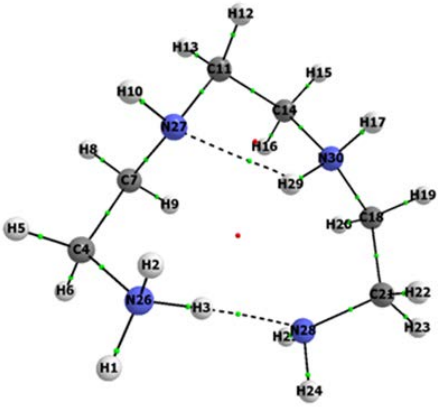
Conformer $C_{ps}01$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N28	H3	2.046	0.0344
	N27	H29	1.961	0.0253
	N28	H29	2.289	0.0175
Contact without AIL				
None				



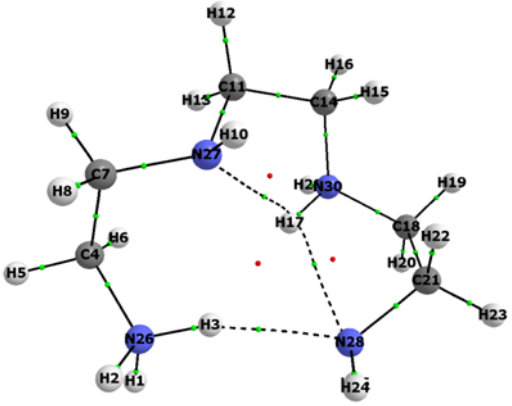
Conformer $C_{ps}41$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N27	H29	2.073	0.0272
	N28	H3	1.885	0.0363
Contact without AIL				
NH--N	N27	H2	2.566	–
CH--HC	H9	H16	2.252	–



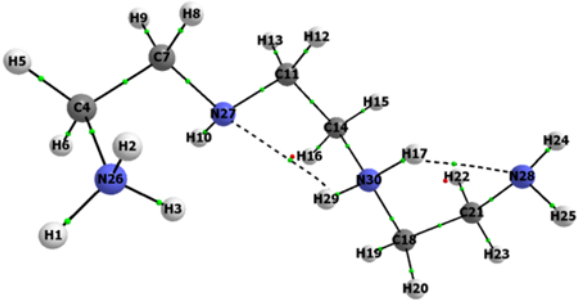
Conformer $C_{ps}10$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N27	H17	2.169	0.0231
	N28	H17	2.201	0.0203
	N28	H3	2.084	0.0233
Contact without AIL				
NH--N	N27	H3	2.494	–
CH--HC	H3	H17	2.194	–
	H15	H22	2.344	–



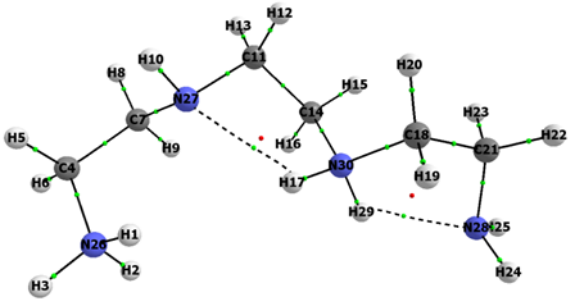
Conformer $C_{ps}03$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N27	H29	2.326	0.0166
	N28	H17	2.080	0.0268
Contact without AIL				
NH--N	N27	H3	2.379	–
CH--HC	H15	H22	2.375	–



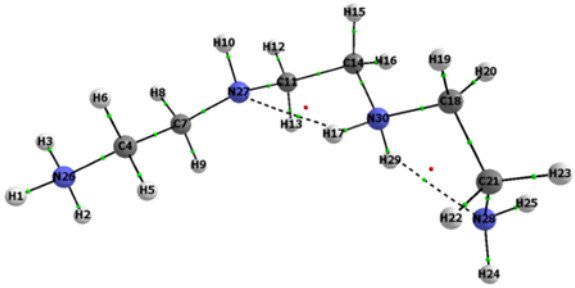
Conformer $C_{ps}14$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N27	H17	2.268	0.0188
	N28	H29	2.043	0.0285
Contact without AIL				
NH--N	N27	H1	2.432	–
CH--HC	H9	H16	2.337	–
	H12	H20	2.298	–
	H15	H23	2.313	–



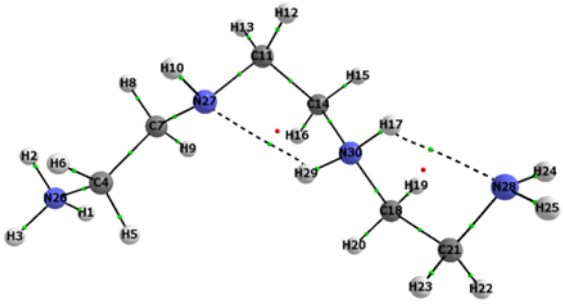
Conformer $C_{ps}05$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N27	H17	2.170	0.0224
	N28	H29	2.107	0.0250
Contact without AIL				
None				



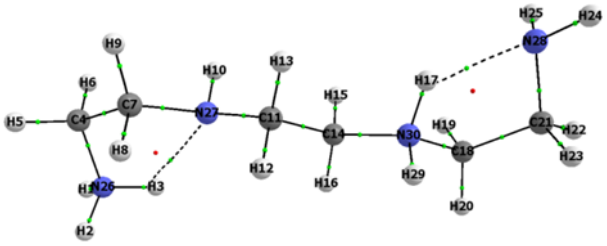
Conformer $C_{ps}17$

Interaction	Atoms A	Atoms B	d(A,B) Å	ρ_{CP} a.u.
NH•••N	N27	H29	2.190	0.0218
	N28	H17	2.088	0.0260
Contact without AIL				
CH--HC	H9	H16	2.321	–



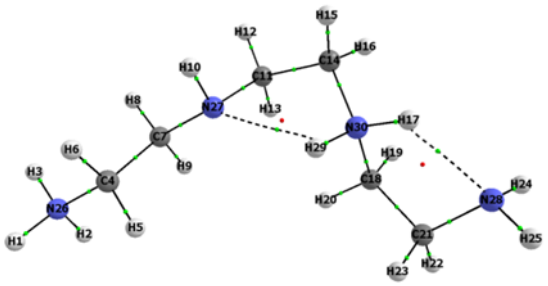
Conformer $C_{ps}02$

Interaction	Atoms A	Atoms B	d(A,B) Å	ρ_{CP} a.u.
NH•••N	N27	H3	2.186	0.0218
	N28	H17	2.034	0.0289
Contact without AIL				
None				



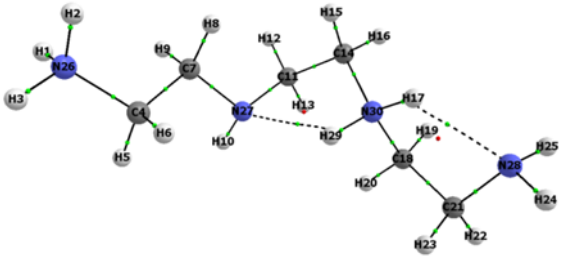
Conformer $C_{ps}34$

Interaction	Atoms A	Atoms B	d(A,B) Å	ρ_{CP} a.u.
NH•••N	N27	H29	2.258	0.0197
	N28	H17	2.074	0.0266
Contact without AIL				
None				



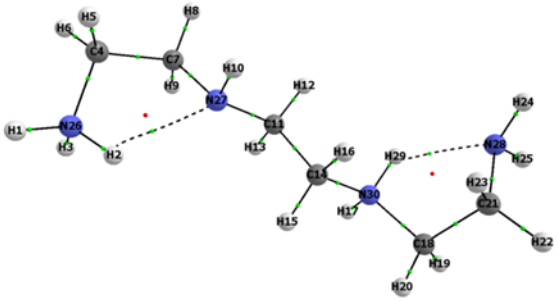
Conformer $C_{ps}43$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N27	H29	2.199	0.0220
	N28	H17	2.086	0.0260
Contact without AIL				
CH--HC	H8	H15	2.332	–



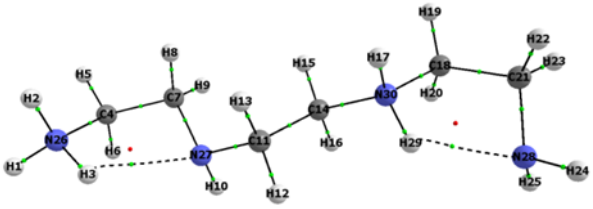
Conformer $C_{ps}04$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N27	H2	2.171	0.0225
	N28	H29	2.011	0.0305
Contact without AIL				
CH--HC	H16	H23	2.291	–



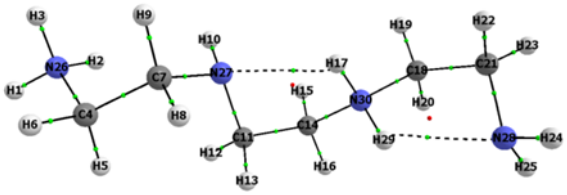
Conformer $C_{ps}12$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N27	H3	2.142	0.0240
	N28	H29	2.040	0.0286
Contact without AIL				
None				



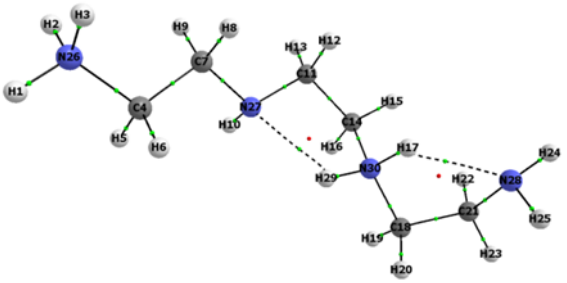
Conformer $C_{ps}21$

Interaction	Atoms A	Atoms B	d(A,B) Å	ρ_{CP} a.u.
NH•••N	N27	H17	2.171	0.0227
	N28	H29	2.095	0.0256
Contact without AIL				
None				



Conformer $C_{ps}09$

Interaction	Atoms A	Atoms B	d(A,B) Å	ρ_{CP} a.u.
NH•••N	N27	H29	2.155	0.0231
	N28	H17	2.101	0.0257
Contact without AIL				
None				



Conformer $C_{ps}15$

Interaction	Atoms A	Atoms B	d(A,B) Å	ρ_{CP} a.u.
NH•••N	N27	H3	2.140	0.0241
	N28	H17	2.015	0.0302
Contact without AIL				
CH--HC	H15	H22	2.280	–

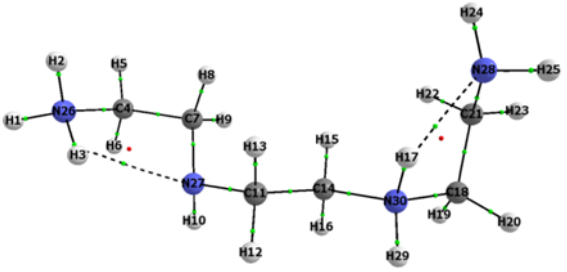
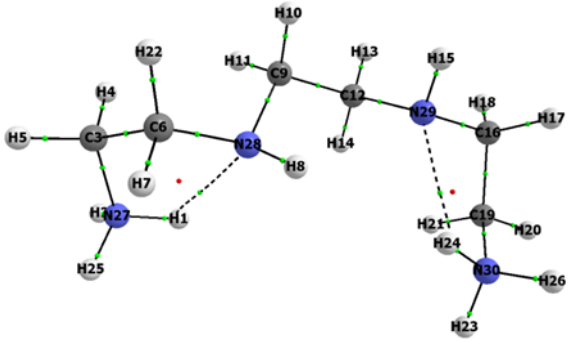


Table A6. Interatomic distance and electron density at a CP of interactions found at MP2 in 15 LECs of HL_{pp} also showing short contacts without AILs on a relevant molecular graph.

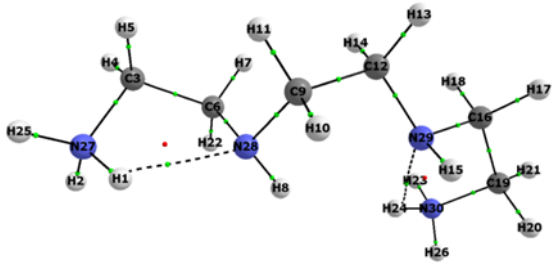
Conformer $C_{pp}12$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N28	H1	2.135	0.0260
	N29	H24	2.566	0.0227
Contact without AIL				
NH--N	N29	H8	2.511	–
CH--HC	H4	H11	2.332	–
	H14	H21	2.345	–



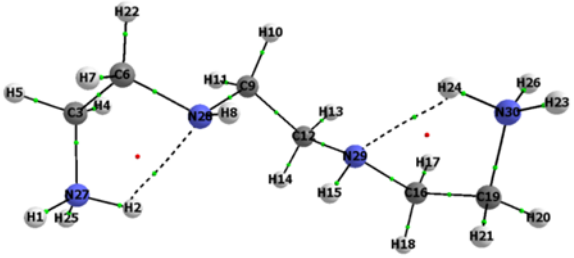
Conformer $C_{pp}10$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N28	H1	2.076	0.0275
	N29	H24	2.559	0.0228
Contact without AIL				
NH--N	N29	H8	2.559	–
CH--HC	H7	H14	2.210	–



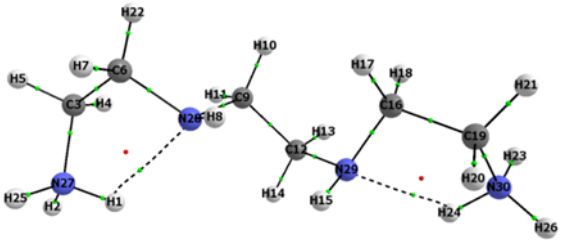
Conformer $C_{pp}07$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N28	H2	2.117	0.0252
	N29	H24	2.125	0.0246
Contact without AIL				
CH--HC	H4	H11	2.311	–
	H8	H15	2.256	–



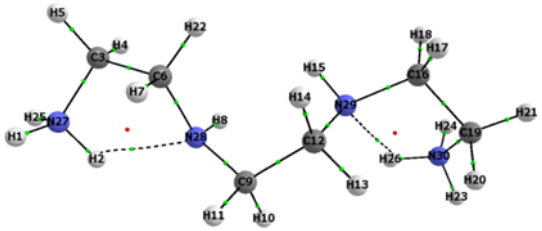
Conformer $C_{pp}16$

Interaction	Atoms A	Atoms B	d(A,B) Å	ρ_{CP} a.u.
NH•••N	N28	H1	2.117	0.0252
	N29	H24	2.111	0.0255
Contact without AIL				
CH--HC	H4	H11	2.334	–
	H8	H15	2.309	–



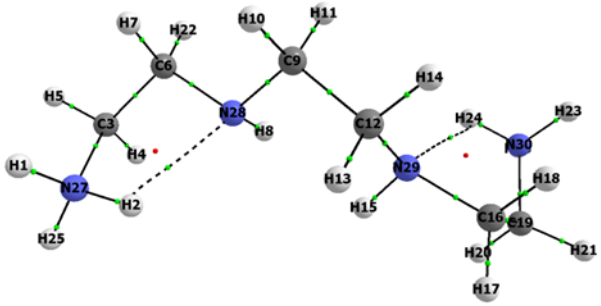
Conformer $C_{pp}11$

Interaction	Atoms A	Atoms B	d(A,B) Å	ρ_{CP} a.u.
NH•••N	N28	H2	2.111	0.0255
	N29	H26	2.117	0.0252
Contact without AIL				
CH--HC	H8	H15	2.309	–
	H20	H13	2.334	–



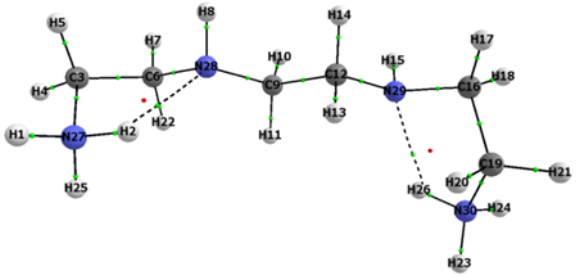
Conformer $C_{pp}09$

Interaction	Atoms A	Atoms B	d(A,B) Å	ρ_{CP} a.u.
NH•••N	N28	H2	2.129	0.0245
	N29	H24	2.129	0.0245
Contact without AIL				
CH--HC	H8	H15	2.211	–



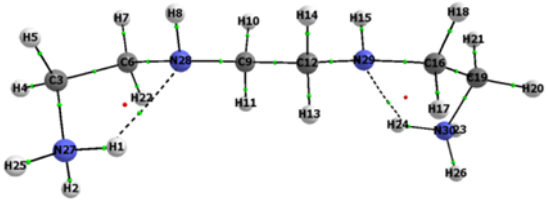
Conformer $C_{pp}05$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N28	H2	2.127	0.0246
	N29	H26	2.101	0.0260
Contact without AIL				
CH--HC	H13	H20	2.328	–



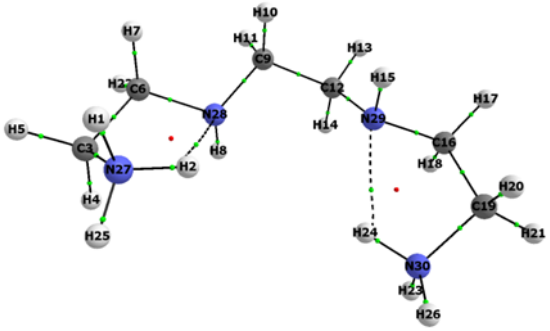
Conformer $C_{pp}02$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N28	H1	2.119	0.0248
	N29	H24	2.119	0.0248
Contact without AIL				
None				



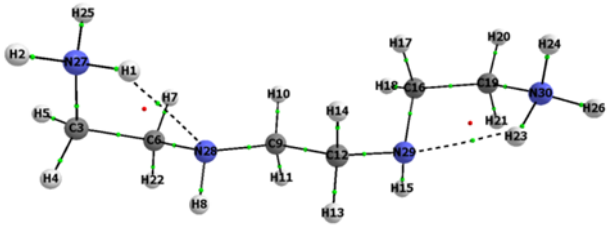
Conformer $C_{pp}23$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N28	H2	2.076	0.0272
	N29	H24	2.076	0.0272
Contact without AIL				
None				



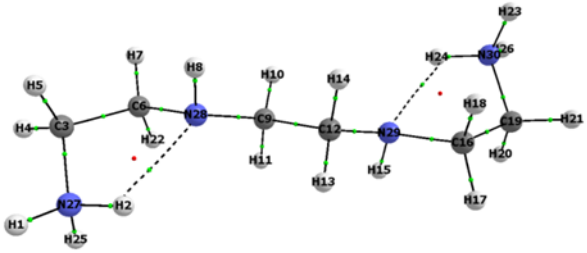
Conformer $C_{pp}04$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N28	H1	2.127	0.0245
	N29	H23	2.079	0.0272
Contact without AIL				
None				



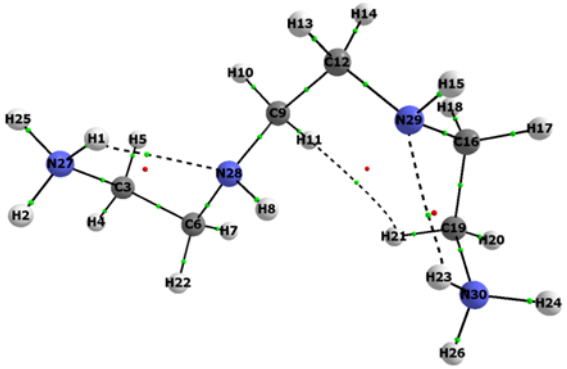
Conformer $C_{pp}01$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N28	H2	2.111	0.0253
	N29	H24	2.111	0.0253
Contact without AIL				
None				



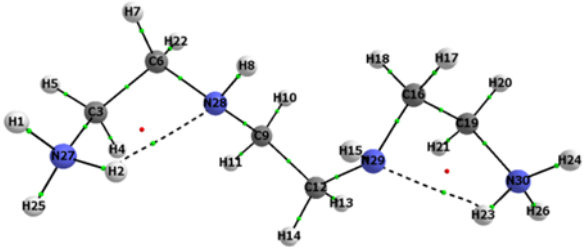
Conformer $C_{pp}27$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N28	H1	2.089	0.0266
	N29	H23	2.475	0.0061
CH•••HC	H11	H21	2.365	0.0061
Contact without AIL				
None				



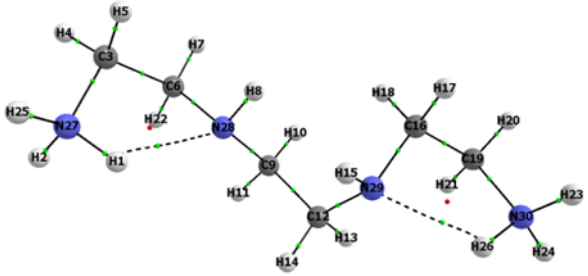
Conformer $C_{pp}14$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N28	H1	2.089	0.0266
CH•••HC	H11	H21	2.365	0.0061
Contact without AIL				
CH--HC	H4	H11	2.322	–
	H8	H15	2.316	–
	H10	H18	2.178	–



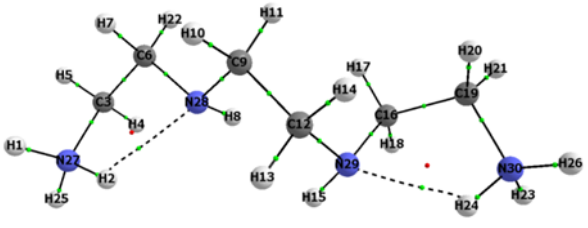
Conformer $C_{pp}13$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N28	H1	2.128	0.0244
	N29	H26	2.092	0.0265
Contact without AIL				
CH--HC	H8	H15	2.311	–
	H10	H18	2.187	–



Conformer $C_{pp}18$

	Atoms		d(A,B)	ρ_{CP}
Interaction	**A**	**B**	**Å**	**a.u.**
NH•••N	N28	H2	2.128	0.0244
	N29	H24	2.091	0.0266
Contact without AIL				
CH--HC	H11	H17	2.188	–



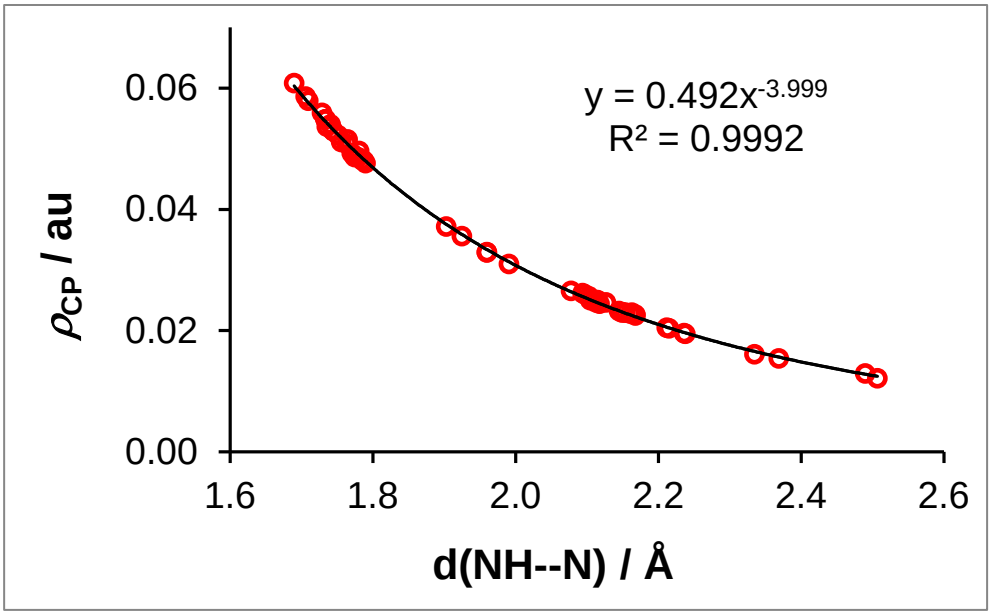
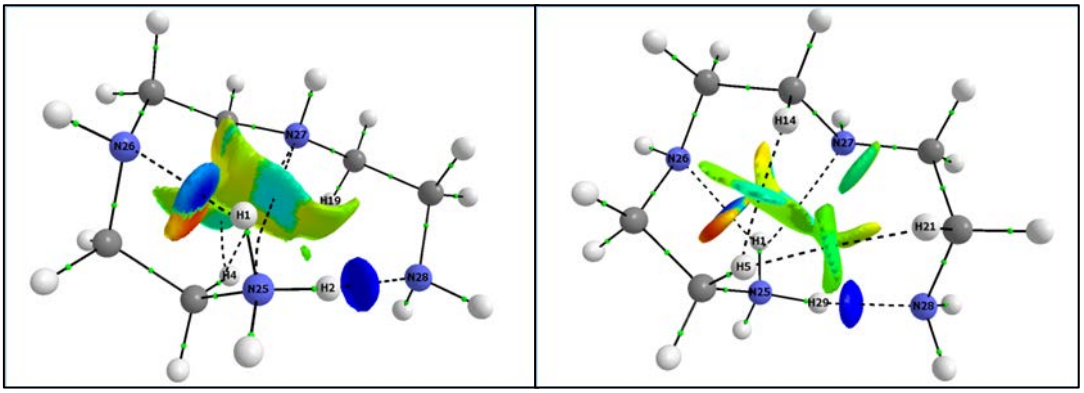
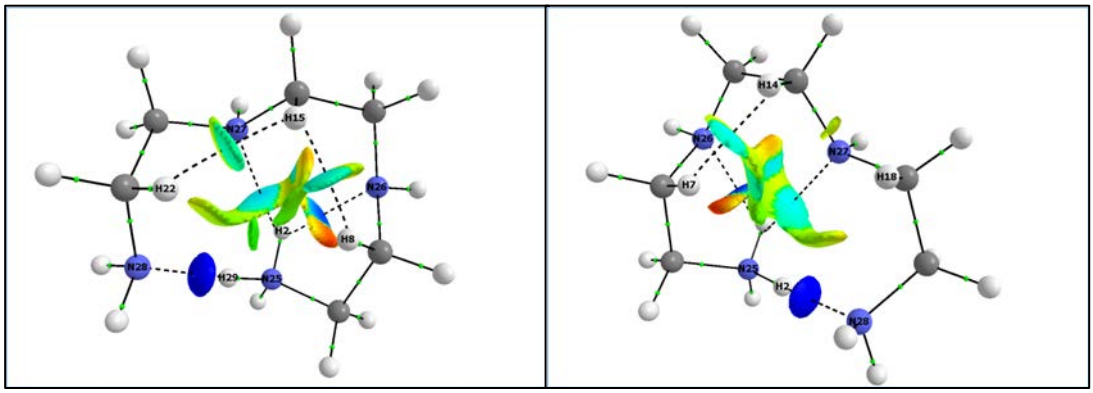


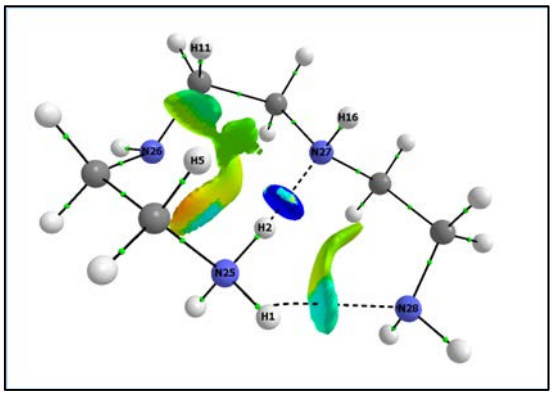
Figure A6. Exponential decrease in ρ_{CP} with interatomic distance d(N,H) for all NH•••N interactions in 15 lowest energy HL_p and HL_s conformers obtained at B3LYP



C_p02 C_p01

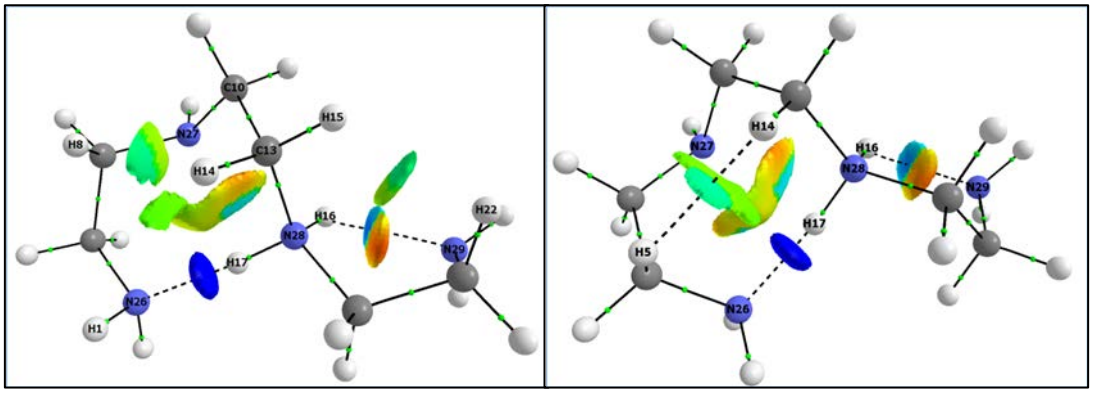


C_p03 C_p04

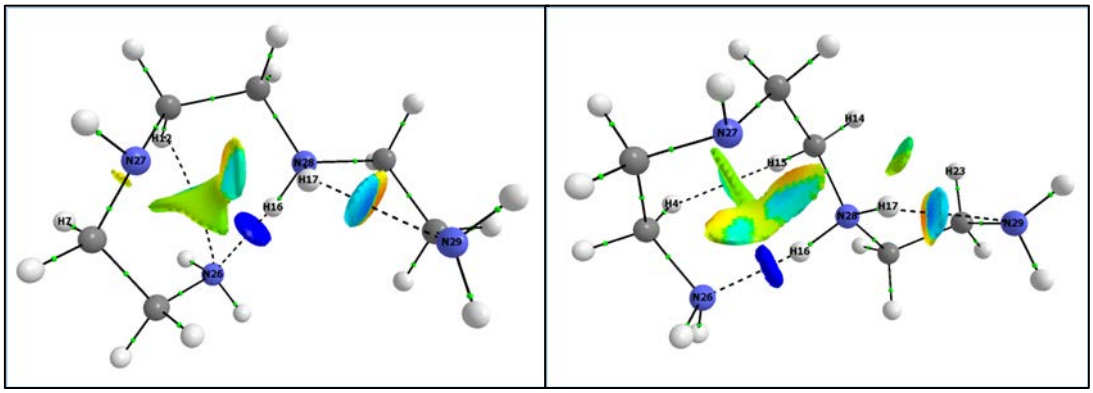


C_p05

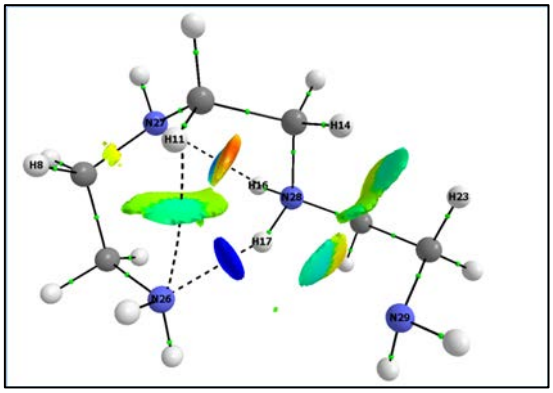
Figure. A7. NCI isosurfaces (RDG isovalue = 0.5 a.u.) for top LECs of HL_p. Isosurfaces are coloured from blue to red using a $-0.03 \leq \rho(r) \times \text{sign}(\lambda_2) \leq +0.03$ range



C_s04 C_s01

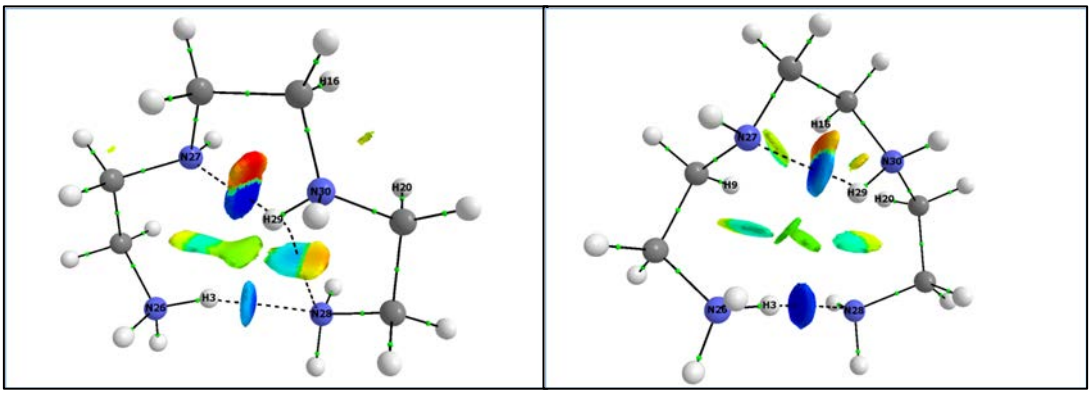


C_s03 C_s02

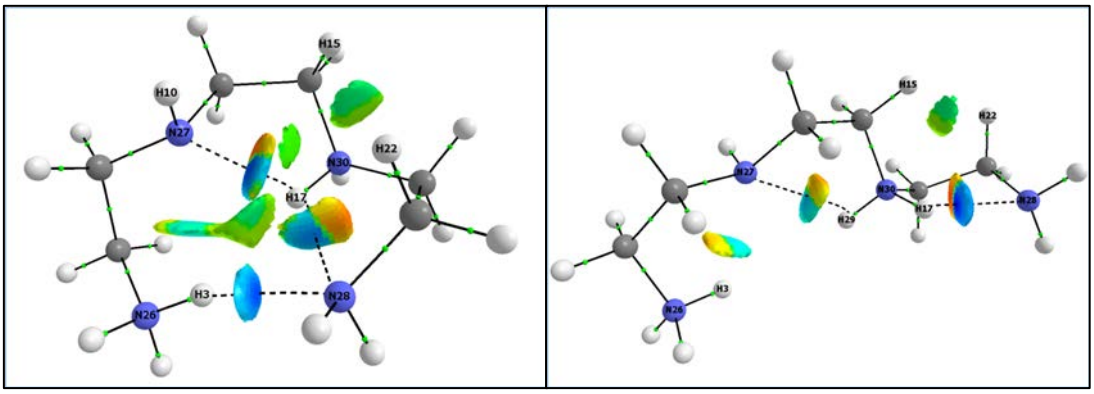


C_s05

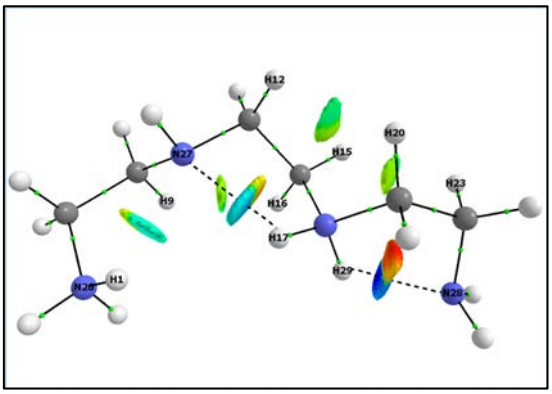
Figure. A8. NCI isosurfaces (RDG isovalue = 0.5 a.u.) for top LECs of HLs. Isosurfaces are coloured from blue to red using a $-0.03 \leq \rho(r) \times \text{sign}(\lambda 2) \leq +0.03$ range



$C_{ps}01$ $C_{ps}41$

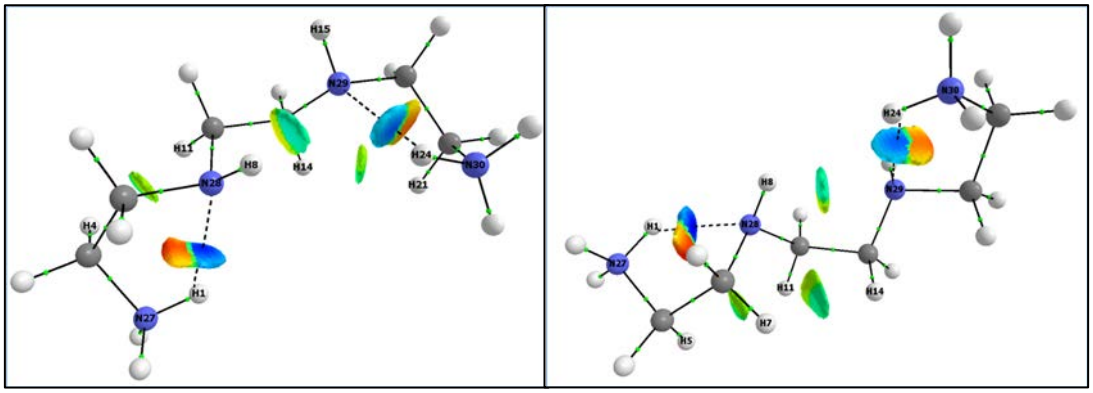


$C_{ps}10$ $C_{ps}03$

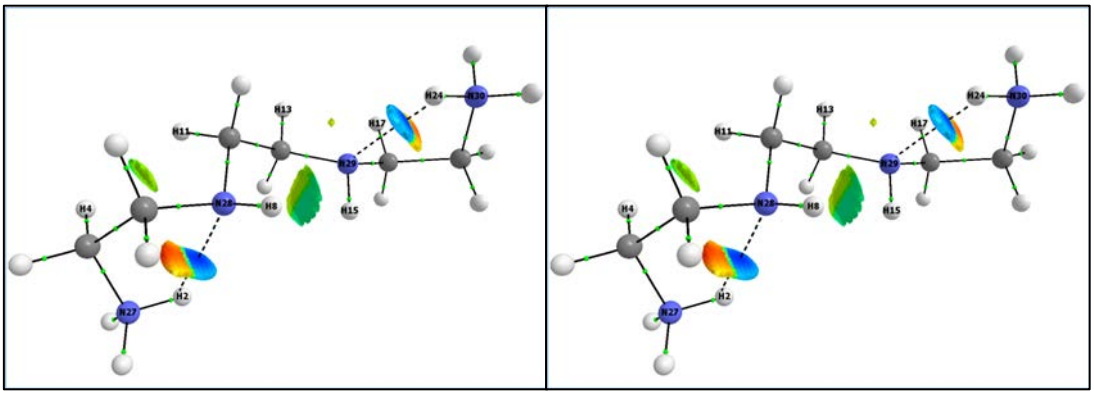


$C_{ps}14$

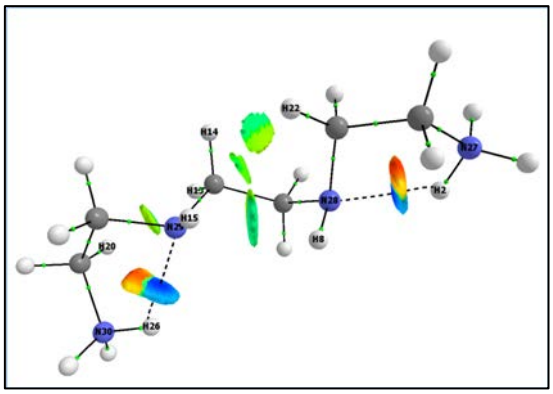
Figure A9. NCI isosurfaces (RDG isovalue = 0.5 a.u.) for top LECs of H_2L_{ps}. Isosurfaces are coloured from blue to red using a $-0.03 \leq \rho(r) \times \text{sign}(\lambda_2) \leq +0.03$ range



$C_{pp}12$ $C_{pp}10$



$C_{pp}07$ $C_{pp}16$



$C_{pp}11$

Figure A10. NCI isosurfaces (RDG isovalue = 0.5 a.u.) for top LECs of H_2L_{pp}. Isosurfaces are coloured from blue to red using a $-0.03 \leq \rho(r) \times \text{sign}(\lambda_2) \leq +0.03$ range

Table A7 Relative to MP2 values of d(N,H) value, performance of HF, B3LYP and B97D in terms of Δd(N,H) in top LECs of HL_p (all values in Å)

Conformer	Contact	d(N,H)	Δd_{HF}	Δd_{B3LYP}	Δd_{B97D}
C_p03	N28--H29	1.741	0.214	0.029	–0.060
	N26--H2	2.136	0.136	0.032	0.107
	N27--H2	2.379	0.138	–0.044	–0.131
	Average for |Δd|:		0.163	0.035	0.099
	StDev for |Δd|:		0.044	0.008	0.036
C_p02	N28--H2	1.750	0.219	0.025	–0.072
	N26--H1	2.040	0.127	0.038	0.081
	N27--H1	2.478	0.066	0.025	–0.001
	Average for |Δd|:		0.137	0.029	0.051
	StDev for |Δd|:		0.077	0.008	0.044
C_p01	N28--H29	1.740	0.227	0.016	–0.079
	N26--H1	2.073	0.125	0.036	0.086
	N27--H1	2.398	0.052	–0.029	–0.043
	Average for |Δd|:		0.135	0.027	0.069
	StDev for |Δd|:		0.088	0.010	0.023
C_p05	N27--H2	1.680	0.276	0.031	–0.061
	N28--H1	2.470	0.303	0.020	–0.050
	N26--H2	2.482	0.053	0.050	0.078
	N28--H2	2.646	0.008	0.038	0.033
	Average for |Δd|:		0.160	0.035	0.056
	StDev for |Δd|:		0.151	0.013	0.019
C_p04	N28--H2	1.740	0.242	0.032	–0.072
	N26--H29	2.101	0.162	0.012	0.077
	N27--H29	2.512	0.157	–0.005	–0.031
	Average for |Δd|:		0.187	0.016	0.060
	StDev for |Δd|:		0.048	0.014	0.025
C_p09	N27--H2	1.689	0.245	0.001	–0.061
	N28--H1	1.989	0.563	0.003	–0.140
	Average for |Δd|:		0.404	0.002	0.101
	StDev for |Δd|:		0.225	0.001	0.056
C_p08	N27--H1	1.717	0.315	0.036	–0.083
C_p06	N27--H29	1.691	0.286	0.043	–0.018
	N28--H2	1.944	0.914	0.016	–0.124
	Average for |Δd|:		0.600	0.029	0.071
	StDev for |Δd|:		0.444	0.020	0.075
	Overall Average for |Δd|:		**0.220**	**0.023**	**0.065**
	Overall StDev for |Δd|:		**0.183**	**0.014**	**0.032**
	Overall Average for Δd:		***0.220***	***0.019***	***0.006***
	Overall StDev for Δd:		***0.183***	***0.019***	***0.073***

Table A8 Relative to MP2 values of d(N,H), performance of HF, B3LYP and B97D in terms of Δd(N,H) in top LECs of HL_s (all values in Å)

Conformer	Contact	d(N,H)	Δd_{HF}	Δd_{B3LYP}	Δd_{B97D}
C_s04	N26--H17	1.745	0.272	0.045	–0.064
	N29--H16	2.177	0.158	0.035	0.160
	Average for |Δd|:		0.215	0.040	0.112
	StDev for |Δd|:		0.081	0.007	0.068
C_s03	N26--H16	1.743	0.305	0.046	–0.058
	N29--H17	2.197	0.173	0.054	0.139
	Average for |Δd|:		0.239	0.050	0.099
	StDev for |Δd|:		0.093	0.006	0.057
C_s01	N26--H17	1.728	0.287	0.058	–0.054
	N29--H16	2.217	0.155	0.022	0.141
	Average for |Δd|:		0.221	0.040	0.098
	StDev for |Δd|:		0.093	0.025	0.062
C_s02	N26--H16	1.728	0.279	0.052	–0.056
	N29--H17	2.461	0.021	0.056	0.071
	Average for |Δd|:		0.150	0.054	0.064
	StDev for |Δd|:		0.182	0.003	0.011
C_s06	N27--H16	2.052	0.248	0.052	0.070
	N29--H17	2.158	0.173	–0.003	0.101
	Average for |Δd|:		0.211	0.028	0.086
	StDev for |Δd|:		0.053	0.035	0.022
C_s09	N27--H16	2.111	0.208	0.057	0.075
	N29--H17	2.115	0.202	0.035	0.075
	Average for |Δd|:		0.205	0.046	0.102
	StDev for |Δd|:		0.004	0.016	0.037
C_s07	N27--H17	2.032	0.231	0.062	0.085
	N29--H16	2.091	0.181	0.027	0.123
	Average for |Δd|:		0.206	0.044	0.104
	StDev for |Δd|:		0.035	0.025	0.027
C_s05	N26--H17	1.845	0.425	0.089	–0.135
	N27--H16	2.253	0.129	0.056	0.176
	N26--H11	2.679	–0.081	0.067	0.055
	Average for |Δd|:		0.212	0.071	0.122
	StDev for |Δd|:		0.186	0.017	0.062
	Overall Average for |Δd|:		**0.179**	**0.040**	**0.084**
	Overall StDev for |Δd|:		**0.090**	**0.021**	**0.039**
	Overall Average for Δd:		***0.174***	***0.040***	***0.062***
	Overall StDev for Δd:		***0.099***	***0.021***	***0.070***

Table A9 Relative to MP2 values of d(N,H), performance of HF, B3LYP and B97D in terms of Δd(N,H) in top LECs of H_2L_{ps} (all values in Å)

Conformer	Contact	d(N,H)	Δd_{HF}	Δd_{B3LYP}	Δd_{B97D}
$C_{ps}01$	N28--H3	2.046	–	0.003	–0.219
	N27--H29	1.961	0.557	0.034	–0.106
	Average for |Δd|:			0.019	0.163
	StDev for |Δd|:			0.022	0.080
$C_{ps}02$	N27--H3	2.186	0.201	0.005	0.079
	N28--H17	2.034	0.246	0.036	0.085
	Average for |Δd|:		0.224	0.021	0.082
	StDev for |Δd|:		0.032	0.022	0.004
$C_{ps}05$	N27--H17	2.170	0.196	0.061	0.113
	N28--H29	2.107	0.212	0.024	0.098
	Average for |Δd|:		0.204	0.042	0.106
	StDev for |Δd|:		0.011	0.026	0.011
$C_{ps}17$	N27--H29	2.190	0.170	0.055	0.077
	N28--H17	2.088	0.220	0.035	0.130
	Average for |Δd|:		0.195	0.045	0.104
	StDev for |Δd|:		0.035	0.014	0.037
$C_{ps}41$	N27--H2	2.566	–0.038	0.081	0.032
	N27--H29	2.073	0.467	0.033	–0.022
	N28--H3	1.885	–	0.006	–0.108
	Average for |Δd|:		0.215	0.040	0.054
	StDev for |Δd|:		0.357	0.038	0.047
$C_{ps}10$	N27--H3	2.494	0.017	0.057	0.125
	N27--H17	2.169	0.208	–0.180	0.025
	N28--H3	2.084	0.476	–0.472	–0.216
	N28--H17	2.201	0.041	0.030	0.201
	Average for |Δd|:		0.186	0.185	0.142
	StDev for |Δd|:		0.211	0.202	0.087
$C_{ps}09$	N27--H29	2.155	0.209	0.060	0.033
	N28--H17	2.101	0.199	0.010	0.179
	Average for |Δd|:		0.204	0.035	0.106
	StDev for |Δd|:		0.007	0.035	0.103
$C_{ps}34$	N27--H29	2.258	0.124	–0.013	0.046
	N28--H17	2.074	0.213	0.030	0.132
	Average for |Δd|:		0.169	0.022	0.089
	StDev for |Δd|:		0.063	0.012	0.061
	Overall Average for |Δd|:		**0.223**	**0.064**	**0.107**
	Overall StDev for |Δd|:		**0.150**	**0.106**	**0.063**
	Overall Average for Δd:		***0.219***	***–0.006***	***0.036***
	Overall StDev for Δd:		***0.157***	***0.125***	***0.121***

Table A10 Relative to MP2 values of d(N,H), performance of HF, B3LYP and B97D in terms of Δd(N,H) in top LECs of H_2L_{pp} (all values in Å)

Conformer	Contact	d(N,H)	Δd_{HF}	Δd_{B3LYP}	Δd_{B97D}
$C_{pp}07$	N28--H2	2.117	–	0.016	0.035
	N29--H24	2.125	0.222	0.022	0.071
	Average for |Δd|:			0.019	0.053
	StDev for |Δd|:			0.004	0.025
$C_{pp}12$	N28--H1	2.100	0.273	0.011	0.035
	N29--H8	2.511	–	0.092	0.055
	N29--H24	2.171	–	0.032	0.098
	Average for |Δd|:			0.045	0.063
	StDev for |Δd|:			0.042	0.032
$C_{pp}09$	N28--H2	2.129	–	0.014	0.057
	N29--H24	2.129	–	0.014	0.057
	Average for |Δd|:			0.014	0.057
	StDev for |Δd|:			0.000	0.000
$C_{pp}02$	N28--H1	2.119	–	0.018	0.054
	N29--H24	2.119	0.248	0.018	0.054
	Average for |Δd|:			0.018	0.054
	StDev for |Δd|:			0.000	0.000
$C_{pp}10$	N28--H1	2.076	0.279	0.021	0.056
	N29--H8	2.559	–	0.100	0.070
	N29--H24	2.164	–	0.011	0.030
	Average for |Δd|:			0.044	0.052
	StDev for |Δd|:			0.049	0.020
$C_{pp}13$	N28--H1	2.128	0.213	–0.023	–0.006
	N29--H26	2.092	0.214	0.040	0.095
	Average for |Δd|:		0.214	0.032	0.051
	StDev for |Δd|:		0.001	0.012	0.063
$C_{pp}23$	N28--H2	2.076	0.278	0.034	0.028
	N29--H24	2.076	0.278	0.034	0.028
	Average for |Δd|:		0.278	0.034	0.028
	StDev for |Δd|:		0.000	0.000	0.000
$C_{pp}14$	N28--H2	2.113	–	0.002	0.035
	N29--H23	2.111	0.243	–0.004	0.038
	Average for |Δd|:			0.003	0.037
	StDev for |Δd|:			0.001	0.002
	Overall Average for |Δd|:		**0.250**	**0.025**	**0.028**
	Overall StDev for |Δd|:		**0.028**	**0.030**	**0.027**
	Overall Average for Δd:		***0.250***	***0.250***	***0.025***
	Overall StDev for Δd:		***0.028***	***0.028***	***0.030***

Table A11 Relative to MP2 values of d(H,H), performance of HF, B3LYP and B97D in terms of Δd(H,H) in top LECs of HLp (all values in Å)

Conformer	Contact	d(N,H)	Δd_{HF}	Δd_{B3LYP}	Δd_{B97D}
C_p03	H8--H15	2.034	0.101	0.086	0.044
	H15--H22	2.133	0.093	0.043	–0.025
	Average for |Δd|:		0.097	0.065	0.035
	StDev for |Δd|:		0.006	0.030	0.013
C_p02	H4--H15	2.124	0.085	0.082	–0.013
	H15--H19	2.335	0.009	–0.006	–0.019
	Average for |Δd|:		0.047	0.044	0.016
	StDev for |Δd|:		0.054	0.054	0.004
C_p01	H5--H14	2.048	0.102	0.111	0.111
	H14--H21	2.299	0.060	0.016	0.016
	Average for |Δd|:		0.081	0.064	0.064
	StDev for |Δd|:		0.030	0.067	0.067
C_p04	H7--H14	2.047	0.091	0.057	0.049
	H14--H18	2.226	0.039	–0.002	–0.007
	Average for |Δd|:		0.065	0.030	0.028
	StDev for |Δd|:		0.037	0.039	0.030
C_p09	H8--H12	2.147	0.014	0.097	0.110
	H14--H21	2.268	0.071	0.046	–0.065
	Average for |Δd|:		0.043	0.072	0.088
	StDev for |Δd|:		0.040	0.036	0.032
C_p08	H5--H11	2.282	0.145	0.166	–0.020
	H15--H22	2.189	0.110	0.070	–0.038
	Average for |Δd|:		0.128	0.118	0.029
	StDev for |Δd|:		0.025	0.068	0.013
C_p06	H8--H15	2.225	0.099	0.106	0.025
	H11--H18	2.225	0.142	0.083	0.016
	Average for |Δd|:		0.121	0.095	0.021
	StDev for |Δd|:		0.030	0.016	0.007
	Overall Average for |Δd|:		**0.083**	**0.069**	**0.040**
	Overall StDev for |Δd|:		**0.041**	**0.045**	**0.034**
	Overall Average for Δd:		***0.083***	***0.068***	***0.013***
	Overall StDev for Δd:		***0.041***	***0.047***	***0.052***

Table A12 Relative to MP2 values of d(H,H), performance of HF, B3LYP and B97D in terms of Δd(H,H) in top LECs of HL_s (all values in Å)

Conformer	Contact	d(N,H)	Δd_{HF}	Δd_{B3LYP}	Δd_{B97D}
C_s04	H8--H14	2.319	0.083	0.063	–0.005
	H15--H22	2.277	0.042	0.156	–0.060
	Average for |Δd|:		0.063	0.110	0.033
	StDev for |Δd|:		0.029	0.066	0.039
C_s03	H7--H12	2.251	0.047	0.013	0.020
C_s01	H5--H14	2.136	0.102	0.092	–0.009
	H15--H19	2.373	0.015	–0.014	0.015
	Average for |Δd|:		0.077	0.101	0.030
	StDev for |Δd|:		0.031	0.003	0.030
C_s02	H4--H15	2.136	0.099	0.099	–0.008
	H14--H23	2.278	0.055	0.103	–0.051
	Average for |Δd|:		0.041	0.060	0.027
	StDev for |Δd|:		0.010	0.066	0.015
C_s06	H7--H15	2.392	0.053	0.134	–0.035
	H8--H12	2.359	0.037	0.040	0.037
	H15--H20	2.328	0.034	0.007	0.010
	Average for |Δd|:		0.041	0.060	0.027
	StDev for |Δd|:		0.010	0.066	0.015
C_s05	H8--H11	2.251	0.064	0.026	0.014
	H14--H23	2.208	0.094	0.066	–0.036
	Average for |Δd|:		0.079	0.046	0.025
	StDev for |Δd|:		0.021	0.028	0.016
	Overall Average for |Δd|:		**0.061**	**0.071**	**0.028**
	Overall StDev for |Δd|:		**0.024**	**0.051**	**0.019**
	Overall Average for Δd:		***0.061***	***0.071***	***–0.011***
	Overall StDev for Δd:		***0.024***	***0.051***	***0.033***

Table A13 Relative to MP2 values of d(H,H), performance of HF, B3LYP and B97D in terms of Δd(H,H) in top LECs and some HECs of H_2L_{ps} (all values in Å)

Conformer	**Contact**	**d(N,H)**	**Δd_{HF}**	**Δd_{B3LYP}**	**Δd_{B97D}**
$C_{ps}17$	H9--H16	2.321	0.104	0.097	–0.042
$C_{ps}41$	H9--H16	2.252	0.190	0.090	–0.010
$C_{ps}10$	H3--H17	2.194	0.267	–0.211	0.012
	H15--H22	2.344	0.062	0.038	–0.088
	Average for |Δd|:		0.165	0.125	0.050
	StDev for |Δd|:		0.145	0.122	0.054
$C_{ps}07$	H12--H20	2.221	0.055	0.042	–0.045
$C_{ps}29$	H5--H12	2.236	0.111	0.075	–0.013
	H11--H21	2.380	0.093	0.123	–0.058
	Average for |Δd|:		0.102	0.099	0.036
	StDev for |Δd|:		0.013	0.034	0.032
	Overall Average for |Δd|:		**0.126**	**0.0966**	**0.0383**
	Overall StDev for |Δd|:		**0.076**	**0.0588**	**0.0290**
	Overall Average for Δd:		***0.126***	***0.0363***	***–0.0349***
	Overall StDev for Δd:		***0.076***	***0.1131***	***0.0337***

Table A14 Relative to MP2 values of d(H,H), performance of HF, B3LYP and B97D in terms of Δd(H,H) in top LECs of H_2L_{pp} (all values in Å)

Conformer	Contact	d(N,H)	Δd_{HF}	Δd_{B3LYP}	Δd_{B97D}
$C_{pp}07$	H4--H11	2.311	–	0.116	–0.016
	H8--H15	2.256	–	0.056	0.006
	Average for |Δd|:			0.086	0.011
	StDev for |Δd|:			0.042	0.007
$C_{pp}12$	H4--H11	2.332	–	0.135	–0.012
	H14--H21	2.345	–	0.121	–0.043
	Average for |Δd|:			0.128	0.028
	StDev for |Δd|:			0.010	0.022
$C_{pp}09$	H8--H15	2.211	–	0.085	0.017
$C_{pp}10$	H7--H14	2.210	–	0.104	–0.037
$C_{pp}13$	H8--H15	2.311	–	0.040	–0.048
	H10--H18	2.187	–	0.081	–0.002
	Average for |Δd|:			0.061	0.025
	StDev for |Δd|:			0.029	0.033
$C_{pp}14$	H4--H11	2.322	–	0.097	–0.022
	H8--H15	2.316	–	0.029	–0.046
	H10--H18	2.178	–	0.100	0.009
	Average for |Δd|:			0.075	0.026
	StDev for |Δd|:			0.040	0.019
$C_{pp}15$	H11--H20	2.229	–	0.084	–0.075
$C_{pp}27$	H11--H21	2.365	0.151	0.108	–0.169
$C_{pp}20$	H11--H20	2.264	0.078	0.038	0.038
	H13--H22	2.511	–	0.091	0.091
	Average for |Δd|:			0.065	0.065
	StDev for |Δd|:			0.037	0.037
$C_{pp}17$	H5--H14	2.263	–	0.038	–0.107
	Overall Average for |Δd|:		**0.115**	**0.083**	**–0.026**
	Overall StDev for |Δd|:		**0.052**	**0.033**	**0.060**
	Overall Average for Δd:		***0.115***	***0.083***	***–0.026***
	Overall StDev for Δd:		***0.052***	***0.033***	***0.060***

Table A15 Relative to MP2 ρ_{CP} values, performance of HF, B3LYP and B97D in terms of $\Delta\rho_{CP}$ for NH•••N interactions in top LECs of HL_p (all values in a.u.)

Conformer	Interaction	ρ_{CP}	$\Delta\rho_{CP}^{HF}$	$\Delta\rho_{CP}^{B3LYP}$	$\Delta\rho_{CP}^{B97D}$
C_p03	N28•••H29	0.0517	–0.0215	–0.0025	0.0102
	N26•••H2	0.0241	–0.0057	–0.0016	–0.0047
	N27•••H2	0.0153	–0.0040	0.0007	0.0036
	Average($\lvert\delta\rho_{CP}\rvert$):		0.0104	0.0016	0.0062
	StDev($\lvert\delta\rho_{CP}\rvert$):		0.0096	0.0009	0.0035
C_p02	N28•••H2	0.0507	–0.0213	–0.0021	0.0115
	N26•••H1	0.0287	–0.0071	–0.0022	–0.0043
	Average($\lvert\delta\rho_{CP}\rvert$):		0.0142	0.0021	0.0079
	StDev($\lvert\delta\rho_{CP}\rvert$):		0.0101	0.0001	0.0051
C_p01	N28•••H29	0.0521	–0.0224	–0.0010	0.0128
	N26•••H1	0.0269	–0.0064	–0.0019	–0.0042
	N27•••H1	0.0154	–0.0025	0.0000	0.0005
	Average($\lvert\delta\rho_{CP}\rvert$):		0.0104	0.0010	0.0059
	StDev($\lvert\delta\rho_{CP}\rvert$):		0.0105	0.0010	0.0063
C_p05	N27•••H2	0.0614	–0.0305	–0.0035	0.0118
	N28•••H1	0.0136	–	–0.0007	0.0008
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0021	0.0063
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0020	0.0078
C_p04	N28•••H2	0.0518	–0.0234	–0.0028	0.0120
	N26•••H29	0.0258	–0.0070	–0.0008	–0.0037
	N27•••H29	0.0126	–	–0.0005	0.0000
	Average($\lvert\delta\rho_{CP}\rvert$):		0.0152	0.0013	0.0053
	StDev($\lvert\delta\rho_{CP}\rvert$):		0.0115	0.0013	0.0061
C_p09	N27•••H2	0.0601	–0.0278	0.0007	0.0115
	N28•••H1	0.0308	–0.0197	0.0001	0.0117
	Average($\lvert\delta\rho_{CP}\rvert$):		0.0237	0.0004	0.0116
	StDev($\lvert\delta\rho_{CP}\rvert$):		0.0057	0.0004	0.0002
C_p08	N27•••H1	0.0559	0.0238	–0.0037	0.0143
C_p06	N27•••H29	0.0599	–0.0305	–0.0050	0.0046
	N28•••H2	0.0336	–	–0.0007	0.0117
	Average($\lvert\delta\rho_{CP}\rvert$):		–0.0305	0.0028	0.0081
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0031	0.0050
	Overall Average($\lvert\delta\rho_{CP}\rvert$):		**0.0169**	**0.0017**	**0.0074**
	Overall StDev($\lvert\delta\rho_{CP}\rvert$):		**0.0102**	**0.0014**	**0.0049**
	Overall Averager($\delta\rho_{CP}$):		***–0.0137***	***–0.0015***	***0.0056***
	Overall StDev($\delta\rho_{CP}$):		***0.0144***	***0.0016***	***0.0070***

Table A16 Relative to MP2 ρ_{CP} values, performance of HF, B3LYP and B97D in terms of $\Delta\rho_{CP}$ for NH•••N interactions in top LECs of HL_s (all values in a.u.)

Conformer	Interaction	ρ_{CP}	$\Delta\rho_{CP}^{HF}$	$\Delta\rho_{CP}^{B3LYP}$	$\Delta\rho_{CP}^{B97D}$
C_s04	N26•••H17	0.0521	–0.0252	–0.0045	0.0104
	N29•••H16	0.0221	–	–0.0016	–0.0057
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0031	0.0081
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0020	0.0034
C_s03	N26•••H16	0.0526	0.0256	–0.0046	0.0096
	N29•••H17	0.0213	–	–0.0017	–0.0049
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0032	0.0073
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0020	0.0033
C_s01	N26•••H17	0.0542	–0.0271	–0.0061	0.0093
	N29•••H16	0.0205	–	–0.0011	–0.0048
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0029	0.0048
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0028	0.0046
C_s02	N26•••H16	0.0540	–0.0266	–0.0055	0.0096
	N29•••H17	0.0220	–	–0.0017	–0.0056
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0030	0.0051
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0022	0.0048
C_s06	N27•••H16	0.0285	–0.0112	–0.0028	–0.0037
	N29•••H17	0.0232	–	–0.0003	–0.0042
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0016	0.0040
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0018	0.0004
C_s09	N27•••H16	0.0255	–0.0086	–0.0028	–0.0036
	N29•••H17	0.0246	–0.0082	–0.0017	–0.0054
	Average($\lvert\delta\rho_{CP}\rvert$):		0.0084	0.0022	0.0045
	StDev($\lvert\delta\rho_{CP}\rvert$):		0.0003	0.0008	0.0013
C_s07	N27•••H17	0.0298	–0.0114	–0.0036	–0.0047
	N29•••H16	0.0259	–0.0082	–0.0015	–0.0055
	Average($\lvert\delta\rho_{CP}\rvert$):		0.0098	0.0025	0.0051
	StDev($\lvert\delta\rho_{CP}\rvert$):		0.0022	0.0015	0.0006
C_s05	N26•••H11	0.0418	–0.0250	–0.0069	0.0168
	N26•••H17	0.0093	–0.0009	–0.0015	–0.0011
	N27•••H16	0.0210	–	–0.0025	–
	Average($\lvert\delta\rho_{CP}\rvert$):		0.0129	0.0036	0.0089
	StDev($\lvert\delta\rho_{CP}\rvert$):		0.0170	0.0029	0.0111
	Overall				
	Average($\lvert\delta\rho_{CP}\rvert$):		**0.0162**	**0.0030**	**0.0066**
	Overall StDev($\lvert\delta\rho_{CP}\rvert$):		**0.0097**	**0.0019**	**0.0038**
	Overall				
	Averager($\delta\rho_{CP}$):		***–0.0115***	***–0.0030***	***0.0004***
	Overall StDev($\delta\rho_{CP}$):		***0.0154***	***0.0019***	***0.0077***

Table A17 Relative to MP2 ρ_{CP} values, performance of HF, B3LYP and B97D in terms of $\Delta\rho_{CP}$ for NH•••N interactions in top LECs of H_2L_{ps} (all values in a.u.)

Conformer	**Interaction**	ρ_{CP}	$\Delta\rho_{CP}^{HF}$	$\Delta\rho_{CP}^{B3LYP}$	$\Delta\rho_{CP}^{B97D}$
$C_{ps}01$	N27•••H29	0.034	–	–0.0024	0.0094
	N28•••H3	0.025	–	0.0003	0.0183
	N28•••H29	0.017	–	–0.0019	–
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0015	0.0138
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0011	0.0063
$C_{ps}02$	N27•••H3	0.022	–	–0.0008	–0.0032
	N28•••H17	0.029	–0.011	–0.0006	–0.0044
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0007	0.0038
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0002	0.0008
$C_{ps}05$	N27•••H17	0.022	–	–0.0003	–0.0045
	N28•••H29	0.025	–0.009	–0.0001	–0.0043
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0002	0.0044
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0002	0.0001
$C_{ps}17$	N27•••H29	0.022	–	0.0002	–0.0031
	N28•••H17	0.026	–0.009	0.0004	–0.0058
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0003	0.0044
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0002	0.0019
$C_{ps}41$	N27•••H29	0.027	–	0.0009	0.0012
	N28•••H3	0.036	–	0.0012	0.0125
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0011	0.0069
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0002	0.0079
$C_{ps}10$	N27•••H17	0.023	–	0.0014	–0.0012
	N28•••H17	0.020	–0.002	0.0017	–
	N28•••H3	0.023	–0.015	0.0019	0.0160
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0017	0.0086
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0002	0.0105
$C_{ps}09$	N27•••H29	0.023	–	0.0022	–0.0050
	N28•••H17	0.026	–0.009	0.0024	–0.0040
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0023	0.0045
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0002	0.0007
$C_{ps}34$	N27•••H29	0.0197	–	0.0032	–0.0022
	N28•••H17	0.0266	–0.009	0.0034	–0.0060
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0033	0.0041
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0002	0.0027
	Overall				
	Average($\lvert\delta\rho_{CP}\rvert$):		**0.0081**	**0.0014**	**0.0063**
	Overall StDev($\lvert\delta\rho_{CP}\rvert$):		**0.0049**	**0.0010**	**0.0051**
	Overall				
	Averager($\delta\rho_{CP}$):		***–0.0092***	***0.0007***	***0.0009***
	Overall StDev($\delta\rho_{CP}$):		***0.0039***	***0.0016***	***0.0082***

Table A18 Relative to MP2 ρ_{CP} values, performance of HF, B3LYP and B97D in terms of $\Delta\rho_{CP}$ for NH•••N interactions in top LECs of H_2L_{pp} (all values in a.u.)

Conformer	Interaction	ρ_{CP}	$\Delta\rho_{CP}^{HF}$	$\Delta\rho_{CP}^{B3LYP}$	$\Delta\rho_{CP}^{B97D}$
$C_{pp}07$	N28•••H2	0.0252	–	–0.0009	–0.0017
	N29•••H24	0.0246	–	–0.0011	–0.0033
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0010	0.0025
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0001	0.0011
$C_{pp}12$	N28•••H1	0.0260	–	–0.0006	–0.0018
	N29•••H24	0.0227	–	–0.0015	–0.0040
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0011	0.0029
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0006	0.0015
$C_{pp}09$	N28•••H2	0.0245	–	–0.0008	–0.0026
	N29•••H24	0.0245	–	–0.0008	–0.0026
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0008	0.0026
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0000	0.0000
$C_{pp}02$	N28•••H1	0.0248	–	–0.0010	–0.0025
	N29•••H24	0.0248	–	–0.0010	–0.0025
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0010	0.0025
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0000	0.0000
$C_{pp}10$	N28•••H1	0.0275	–	–0.0012	–0.0030
	N29•••H24	0.0228	–	–0.0007	–0.0013
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0009	0.0022
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0004	0.0012
$C_{pp}13$	N28•••H1	0.0244	–	0.0241	0.0218
	N29•••H26	0.0265	0.0176	0.0259	0.0250
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0250	0.0234
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0012	0.0023
$C_{pp}23$	N28•••H2	0.0272	–	–0.0018	–0.0013
	N29•••H24	0.0272	–	–0.0018	–0.0013
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0018	0.0013
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0000	0.0000
$C_{pp}14$	N28•••H2	0.0253	–	–0.0002	–0.0017
	N29•••H23	0.0256	–	0.0001	–0.0019
	Average($\lvert\delta\rho_{CP}\rvert$):			0.0002	0.0018
	StDev($\lvert\delta\rho_{CP}\rvert$):			0.0000	0.0002
	Overall				
	Average($\lvert\delta\rho_{CP}\rvert$):			**0.0040**	**0.0049**
	Overall StDev($\lvert\delta\rho_{CP}\rvert$):			**0.0082**	**0.0073**
	Overall				
	Averager($\delta\rho_{CP}$):			***0.0023***	***0.0009***
	Overall StDev($\delta\rho_{CP}$):			***0.0089***	***0.0088***

Table A19. Relative to MP2 ρ_{CP} values, performance of HF, B3LYP and B97D in terms of $\Delta\rho_{CP}$ for CH•••HC interactions in top LECs of HL_s (all values in a.u.)

Conformer	**Interaction**	$\boldsymbol{\rho_{CP}}$	$\Delta\rho_{CP}^{HF}$	$\Delta\rho_{CP}^{B3LYP}$	$\Delta\rho_{CP}^{B97D}$
C_s01	H5•••H14	0.0096	–0.0020	–0.0016	–0.0001
C_s02	H4•••H15	0.0096	–0.0020	–0.0017	–0.0002
C_s14	H5•••H14	0.0097	–0.0020	–0.0017	–0.0002
C_s13	H5•••H14	0.0097	–0.0020	–0.0017	–0.0002
	Overall Average($\lvert\delta\rho_{CP}\rvert$):		**0.0020**	**0.0017**	**0.0002**
	Overall StDev($\lvert\delta\rho_{CP}\rvert$):		**0.0000**	**0.0000**	**0.0001**
	Overall Averager($\delta\rho_{CP}$):		***–0.0020***	***–0.0017***	***–0.0002***
	Overall StDev($\delta\rho_{CP}$):		***0.0000***	***0.0000***	***0.0001***

Table A20. Relative to MP2 ρ_{CP} values, performance of HF, B3LYP and B97D in terms of $\Delta\rho_{CP}$ for CH•••HC interactions in HECs of HL_{pp} (all values in a.u.)

Conformer	**Interaction**	$\boldsymbol{\rho_{CP}}$	$\Delta\rho_{CP}^{HF}$	$\Delta\rho_{CP}^{B3LYP}$	$\Delta\rho_{CP}^{B97D}$
C_{pp}15	H11•••H20	0.0079	–	–0.0017	0.0007
C_{pp}27	H11•••H21	0.0061	–0.0019	0.0113	0.0017
C_{pp}20	H11•••H20	0.0075	–0.0019	–0.0012	0.0011
C_{pp}17	H5•••H14	0.0075	–	0.0154	0.0011
	Overall Average($\lvert\delta\rho_{CP}\rvert$):		**0.0019**	**0.0074**	**0.0012**
	Overall StDev($\lvert\delta\rho_{CP}\rvert$):		**0.0000**	**0.0071**	**0.0004**
	Overall Averager($\delta\rho_{CP}$):		***–0.0019***	***0.0059***	***0.0012***
	Overall StDev($\delta\rho_{CP}$):		***0.0000***	***0.0087***	***0.0004***

Appendix B

Supplementary Information for Chapter 4

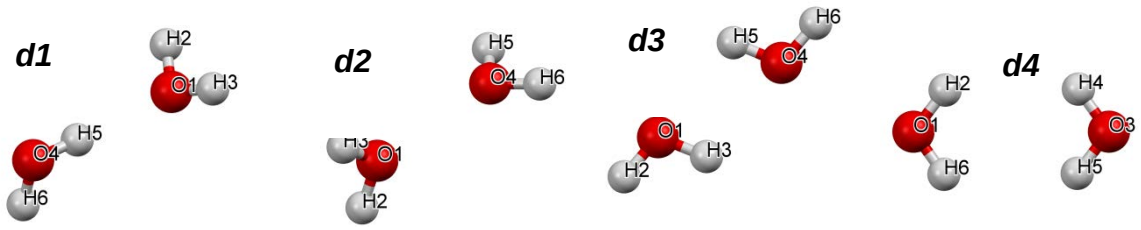


Figure B1. Ball-and-stick representation of water dimers arranged to simulate various intramolecular interactions.

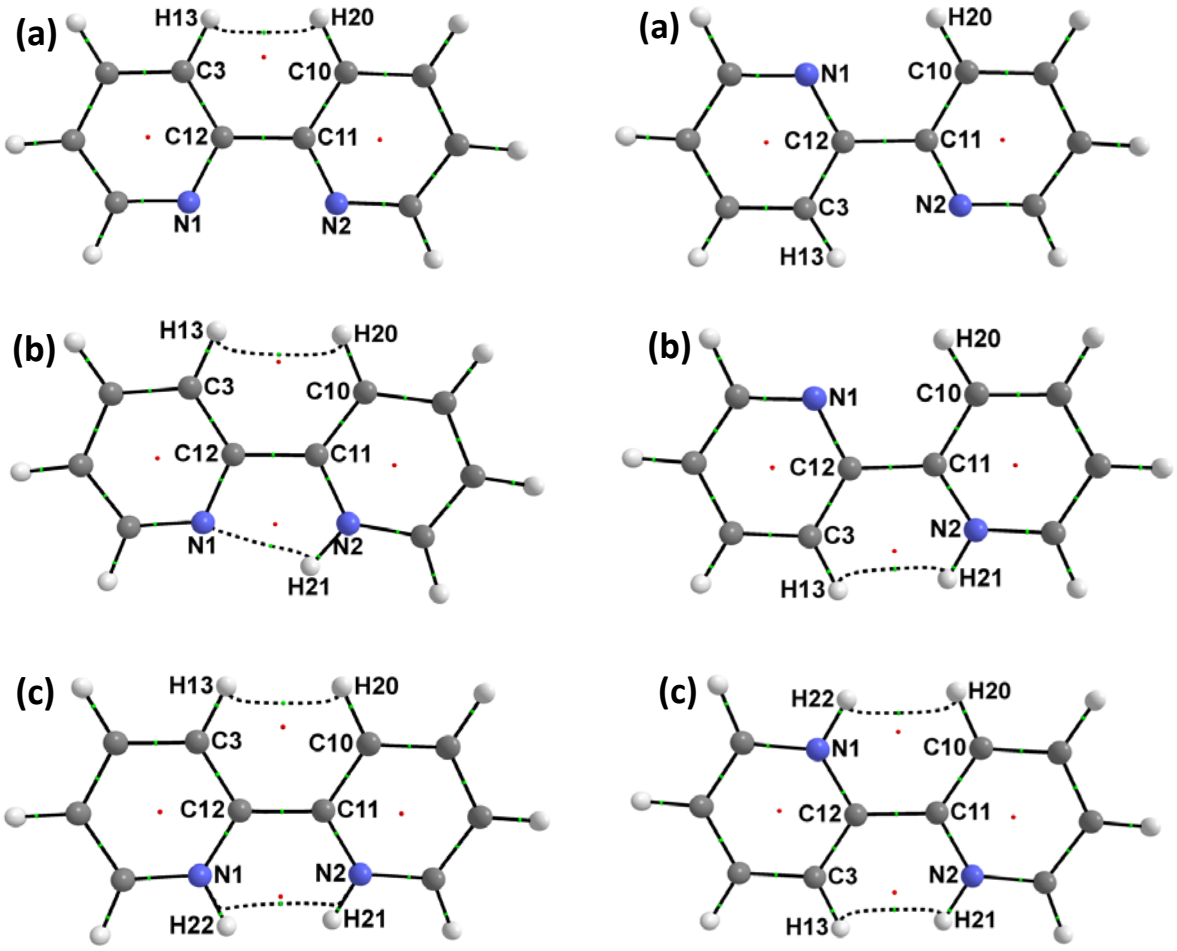


Figure B2. Molecular graphs of the s-*cis* and s-*trans* forms of a) bipyridine, L, b) HL and c) H_2L.

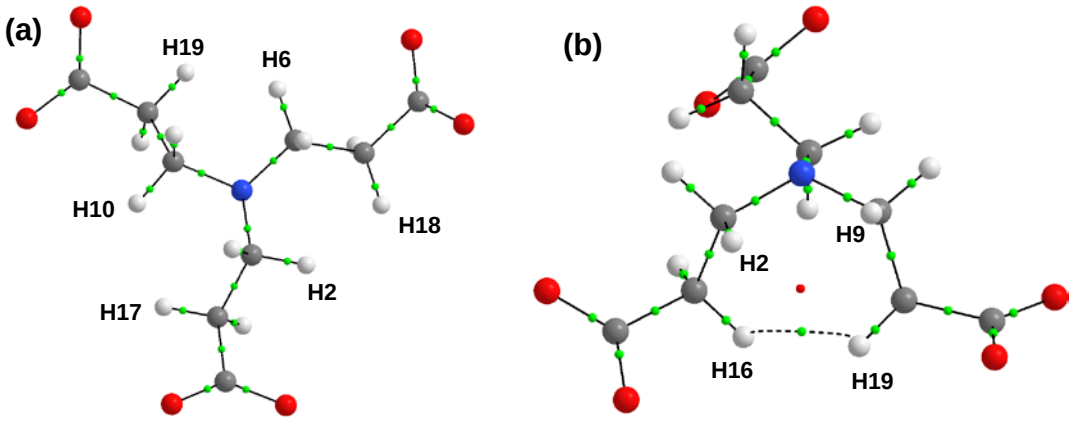


Figure B3. Molecular graphs of a) the lowest and b) highest energy conformer of NTPA.

Table B1. Relative electronic energies ($\Delta E = E_{Conf} - E_{LEC}$ in kcal/mol) of top fifteen lowest energy conformers found at MP2 level of theory; bond length, ρ_{CP} (a.u.) and IQA interaction energy (at HF level of theory in kcal/mol) of leading intramolecular NH•••N interaction in –HL_p,and –HL_s conformers of 2,2,2-tet.

Conf.	**ΔE**	**d(N,H)**	**ρ_{CP}(NH•••N)**	**$E_{int}^{A,B}$**
C_s04	0.00	1.745	0.052	–120.2
C_p02	0.00	1.750	0.051	–
C_s01	0.14	1.728	0.054	–119.3
C_s03	0.64	1.743	0.053	–
C_p01	0.70	1.740	0.052	–126.0
C_s02	0.71	1.728	0.054	–120.6
C_p03	0.71	1.741	0.052	–127.2
C_p04	1.52	1.740	0.052	–125.1
C_p05	2.09	1.680	0.061	–
C_p06	2.47	1.691	0.060	–135.8
C_p08	2.57	1.717	0.056	–
C_s05	2.72	1.845	0.042	–
C_s22	2.90	1.719	0.056	–122.8
C_s06	3.58	2.052	0.029	–99.7
C_s13	3.59	1.694	0.059	–122.2
C_s14	3.60	1.692	0.059	–122.5
C_p09	3.63	1.689	0.060	–138.9
C_p20	3.75	1.708	0.056	–
C_s07	3.98	2.032	0.030	–
C_p23	4.61	1.661	0.064	–136.8
C_p29	4.95	1.694	0.060	–133.1
C_s09	5.05	2.111	0.026	–
C_s08	5.12	2.111	0.025	–
C_p26	5.15	1.671	0.063	NA
C_p15	5.16	1.690	0.060	–131.4
C_p24	5.17	1.654	0.065	–138.2
C_s32	5.19	2.083	0.027	–98.8
C_s10	5.61	2.063	0.028	–
C_s24	5.73	2.053	0.029	–

Appendix C

Supplementary Information for Chapter 5

PART 1

Comments on conformational search procedure and competition reactions considered.

For both 3,2,3-tet and 2,2,2-tet, when the implicit solvation model was to be implemented, we varied each rotatable bond in steps of 60^0 each and retained thirty lowest energy LECs. These were fully optimized using the implicit solvation model to describe solvent environment in order to locate representative set of LECs. In implementing the discrete continuum solvation model for 2,2,2-tet, we used a similar conformational search protocol to that which was implemented in the case of implicit solvation but only retained twenty LECs based on their relative MMFF(aq) energies because our previous conformational analysis of 2,2,2-tet indicated that its representative LECs would most likely be found among the ten LECs obtained from an MMFF(aq) based search of its conformational space.[1] However, since 3,2,3-tet has a greater number of rotatable bonds than 2,2,2-tet due to its longer alkyl chain length and there has been no conformational analysis work reported for it, we carried out the conformational search for its LECs in the presence of explicit solvent molecules in two stages. This was to enable us investigate which torsional angle increment (60° or 120°) is best for locating its low energy conformers:

- In the first stage, each rotatable bond was allowed to change in 60^0 steps and thirty unique lowest energy conformers were retained based on their relative MMFF(aq) energies.

- For the second stage rotation of bonds was allowed in 120^0 steps and thirty unique lowest energy conformers were also retained based on their relative MMFF(aq) energies.

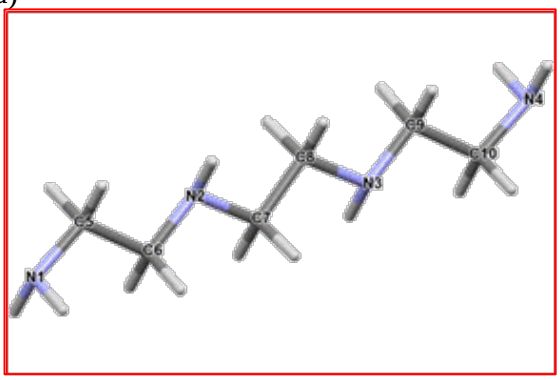
Conformers obtained from both stages were combined and for each tautomer examined, only the thirty unique LECs out of sixty generated, were selected for further optimizations. After full optimization, a thorough examination and analysis of the final structures and energies of LECs generated for 3,2,3-tet showed that in general most of its lowest in energy conformers were from the MM-based conformational search when each rotatable bond was allowed to vary in steps of 120°. Hence, as pointed out in our previous work on 2,2,2-tet, to identify representative LECs of any given aliphatic polyamine, it is sufficient to vary its rotatable bonds in increments of 120° each.

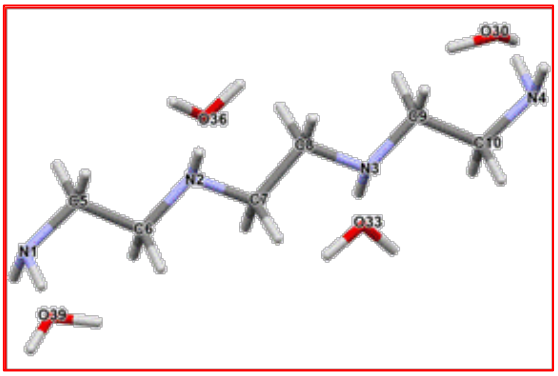
1. Adeyinka, A.S.; Cukrowski, I.; *J. Mol Model.* **2015**, 21, 162.

PART 2

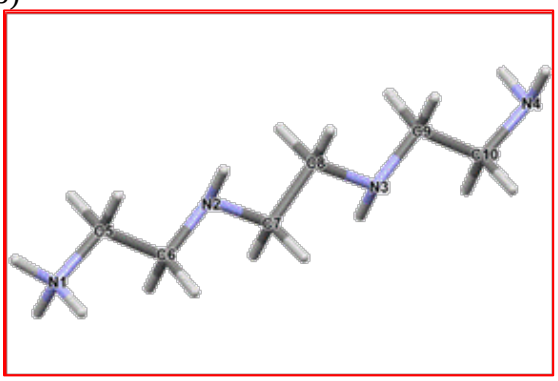
Input structures used for the conformational search

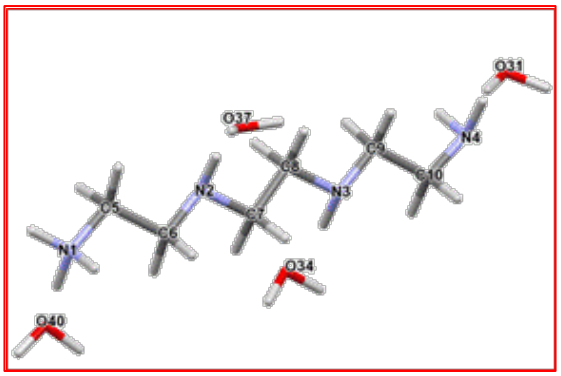
(a)



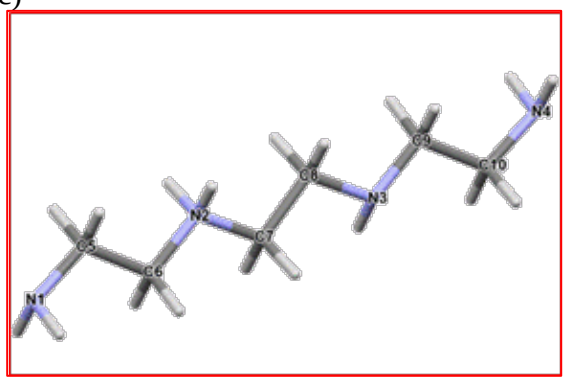


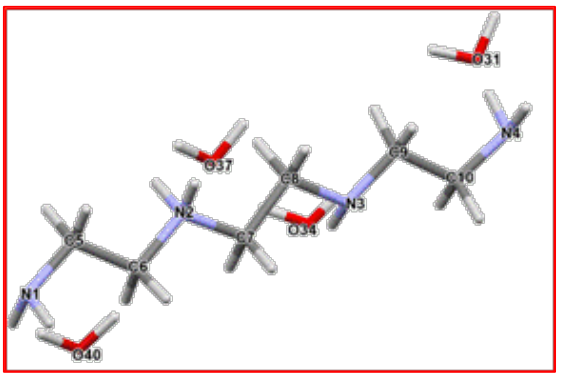
(b)



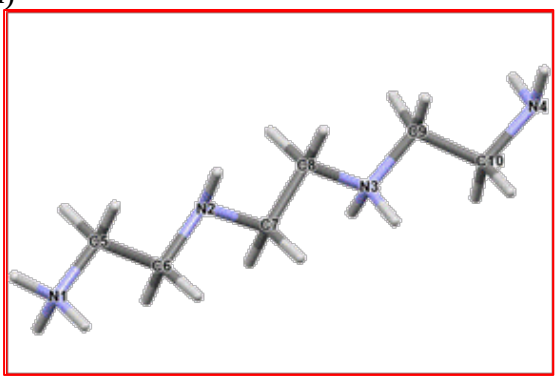


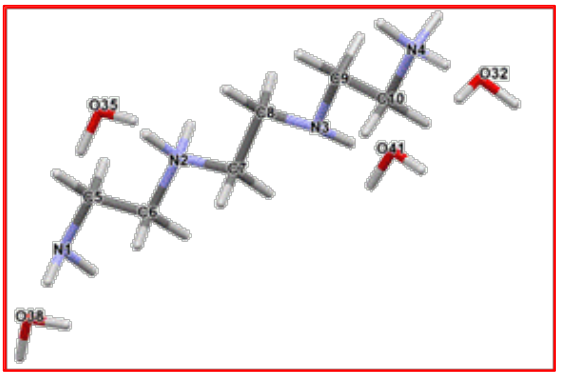
(c)





(d)





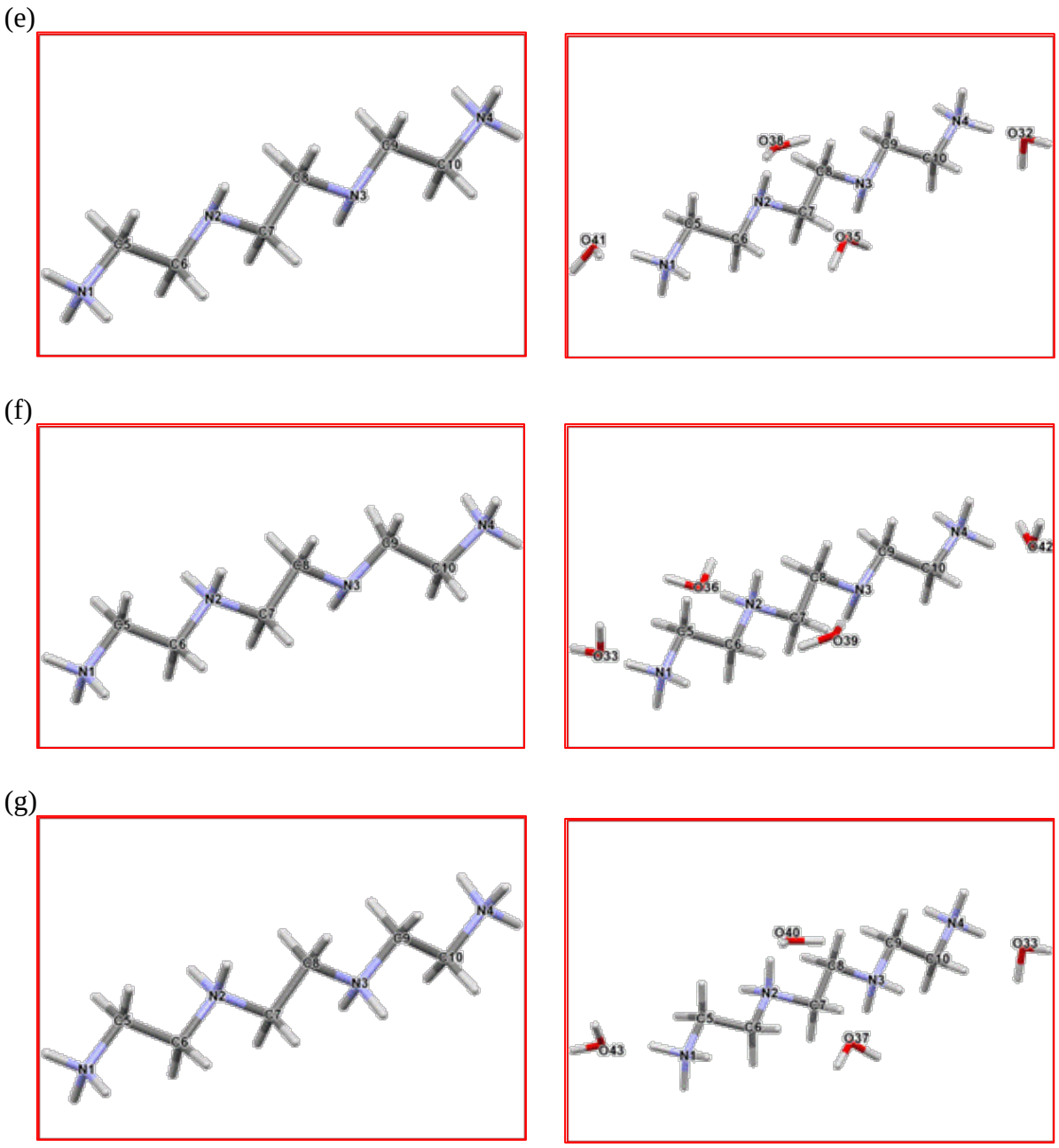
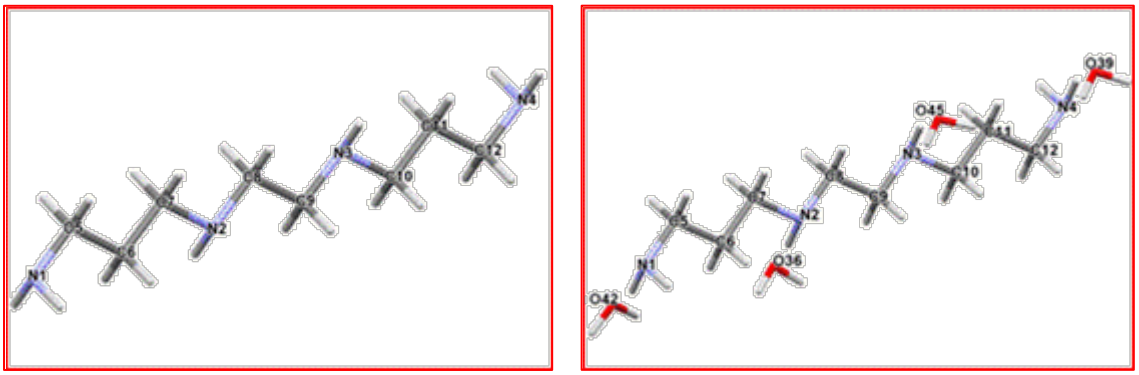
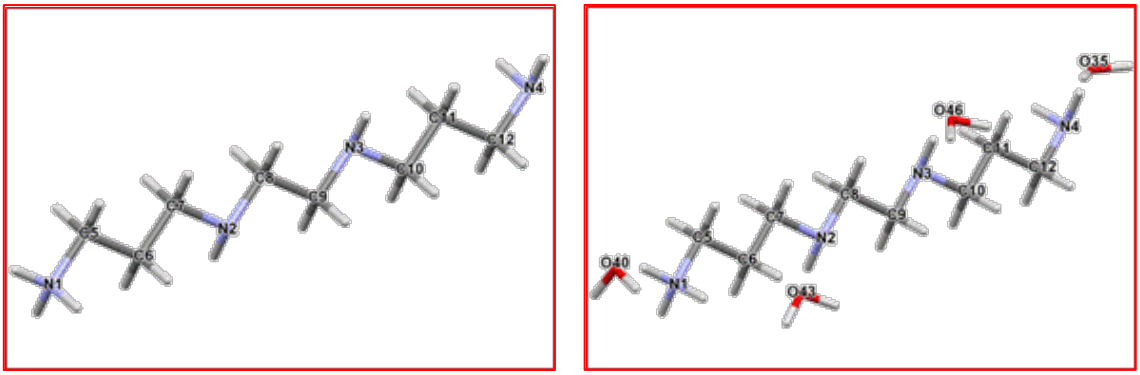


Figure C1. Capped-stick representation of linear structures of H_nL^{n+} forms of 2,2,2-tet with and without explicit water molecules used as inputs for conformational search by MM, also showing atoms' numbering: part (a) – L; part (b) – HL_{N1} (HL_p); part (c) – HL_{N2} (HL_s); part (d) – H_2L_{N1N3} (H_2L_{ps}), part (e) – H_2L_{N1N4} (H_2L_{pp}); part (f) – H_3L_{N1N2N4} (H_3L_{psp}) and part (g) – H_4L.

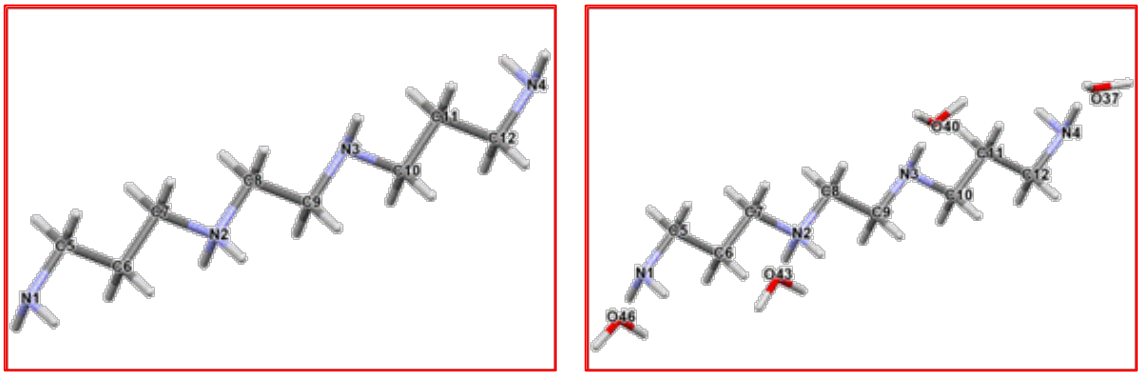
(a)



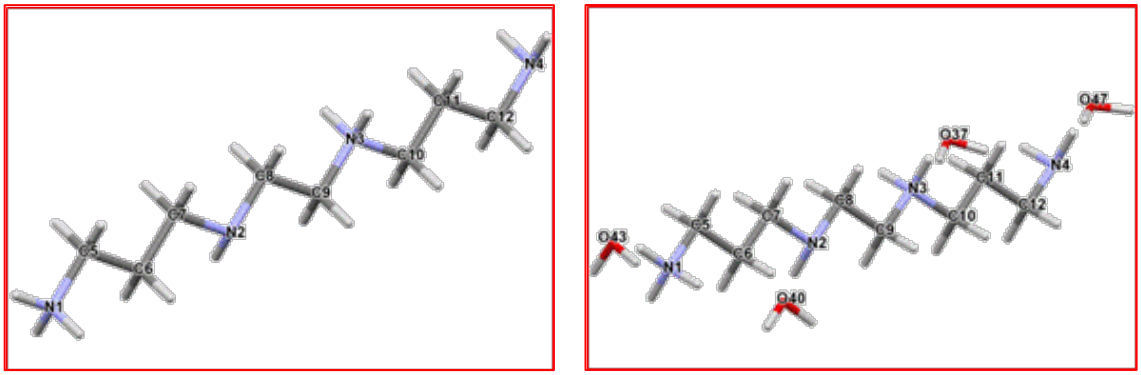
(b)



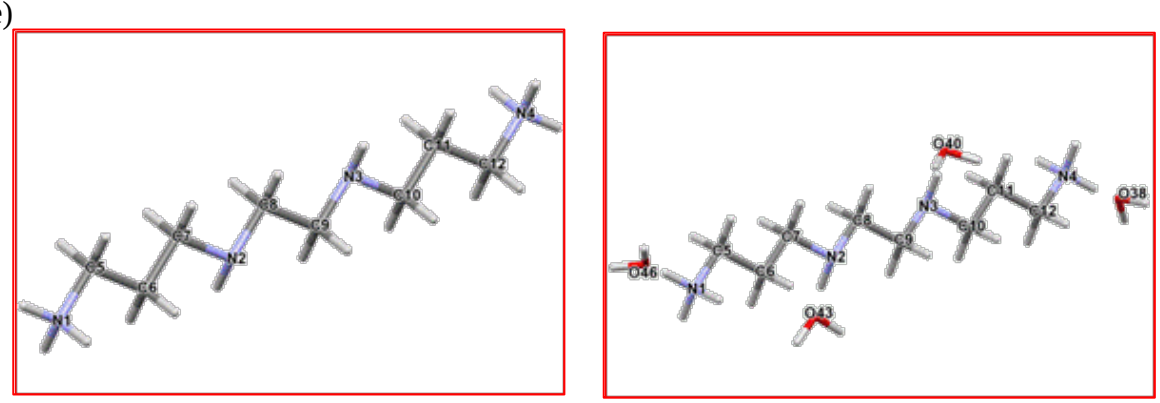
(c)



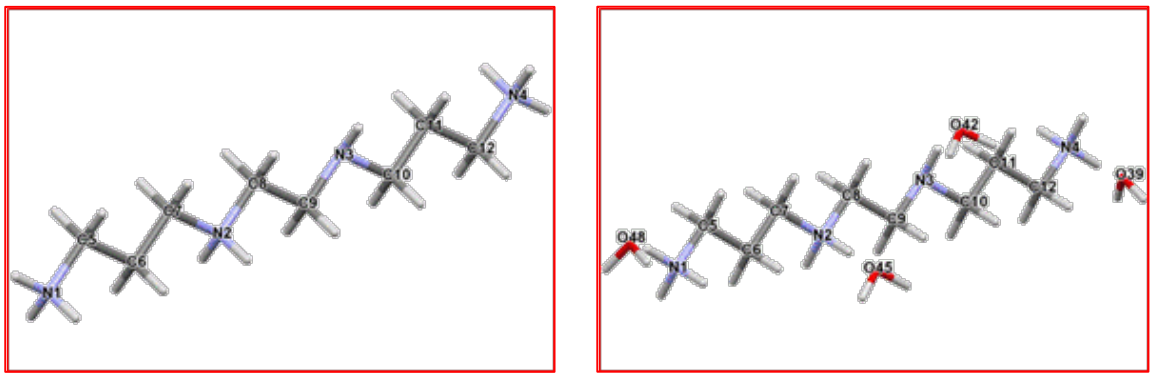
(d)



(e)



(f)



(g)

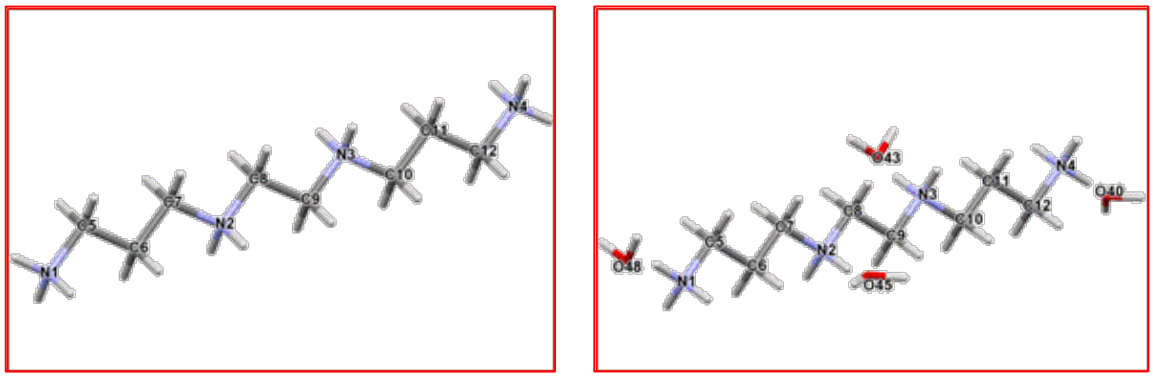
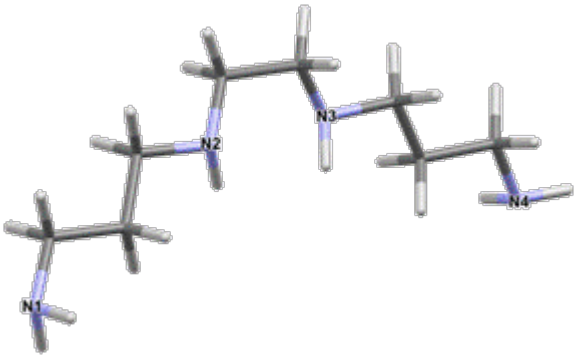


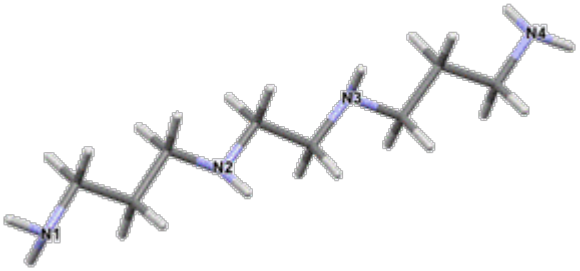
Figure C2. Capped-stick representation of linear structures of H_nL^{n+} forms of 3,2,3-tet with and without explicit water molecules respectively used as inputs for conformational search by MM, also showing atoms' numbering; -part (a) L , -part (b) $HL_{N1}(HL_p)$, -part (c) $HL_{N2}(HL_s)$, part (d) H_2L_{N1N3} (H_2L_{ps}), -part (e) H_2L_{N1N4}(H_2L_{pp}), -part (f) H_3L_{N1N2N4} (H_3L_{psp}) and -part (g) H_4L.

PART 3

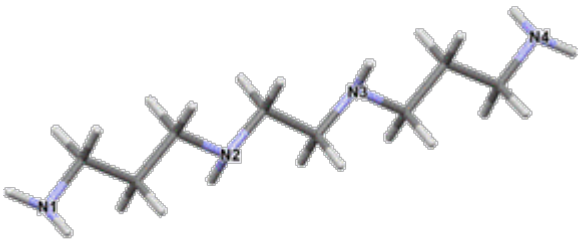
Lowest energy conformers discovered in the continuum solvation model, PCM.



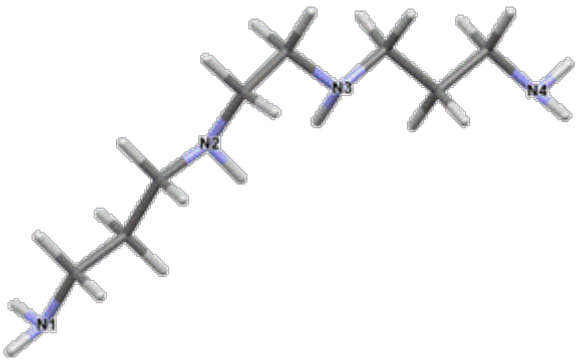
C_L05



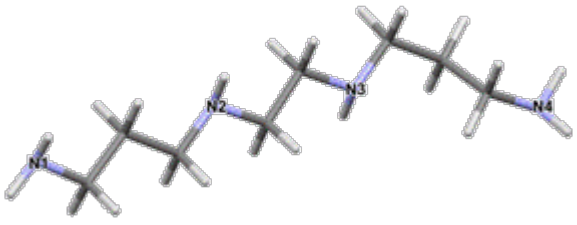
C_L02



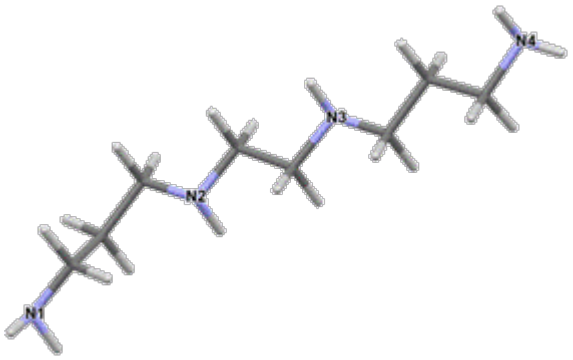
C_L01



C_L04

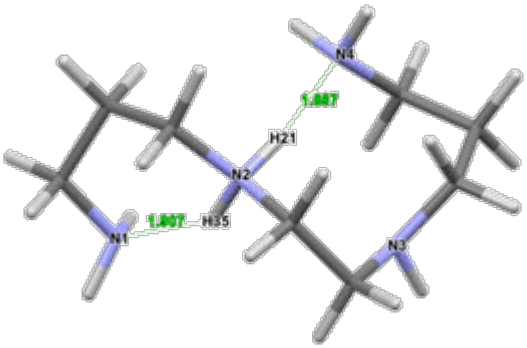


C_L06

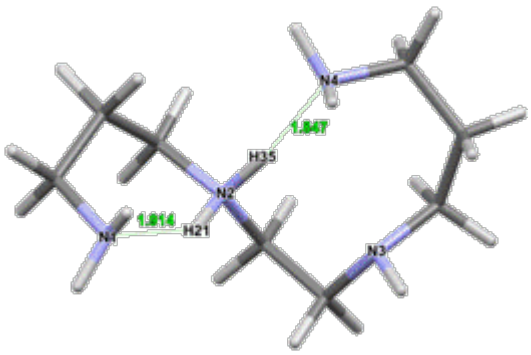


C_L03

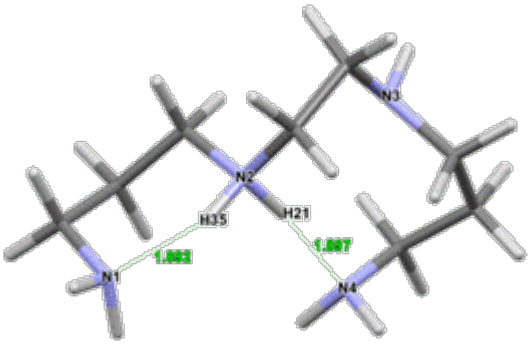
Figure C3. Structures of all lowest in energy conformers of the free ligand of 3,2,3-tet used to calculate protonation constants with continuum solvation model, PCM



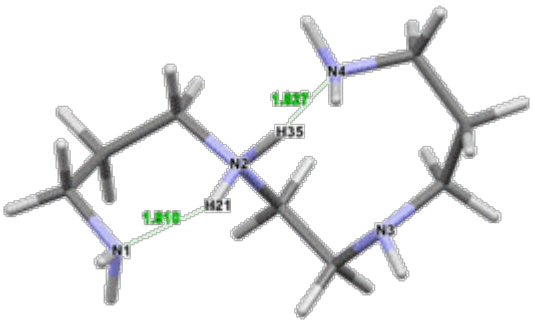
C_s01



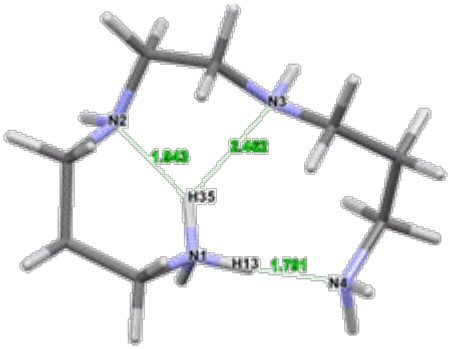
C_s03



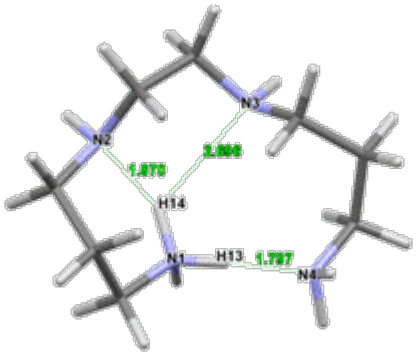
C_s02



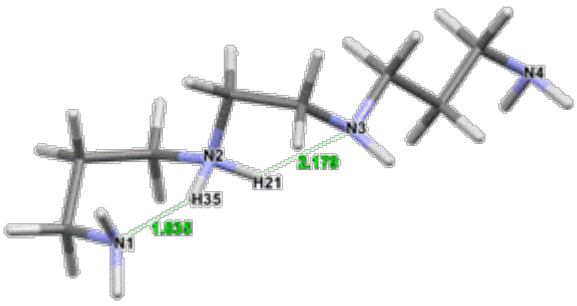
C_s04



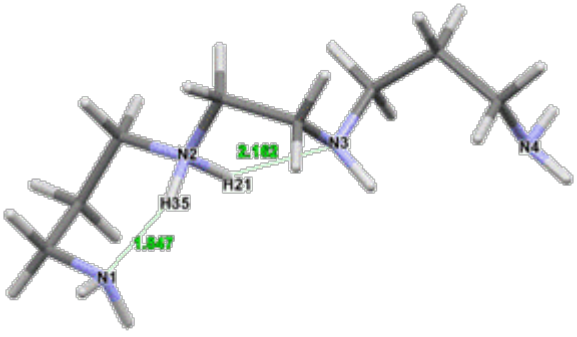
C_p01



C_p03

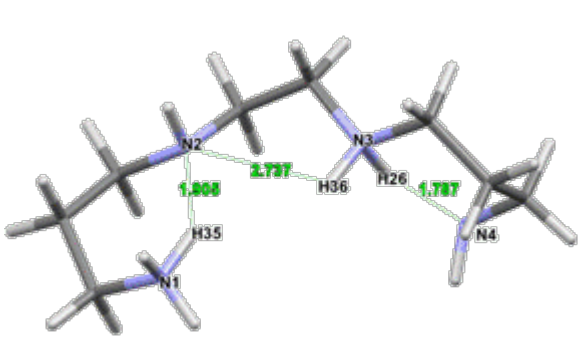


C_s10

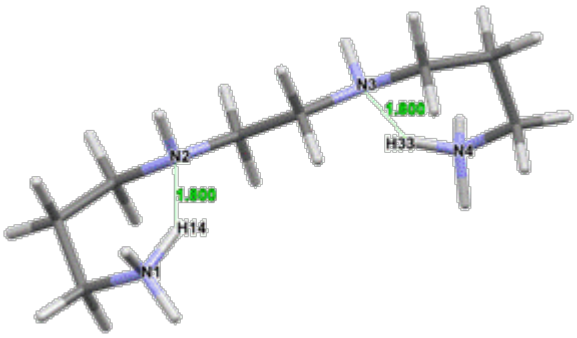


C_s09

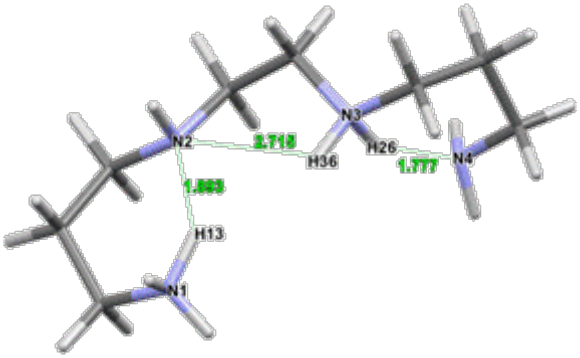
Figure C4. Structures of all lowest in energy conformers for HL form of 3,2,3-tet used to calculate protonation constants with continuum solvation model, PCM



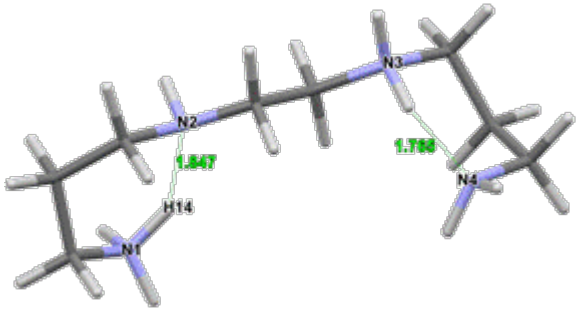
$C_{ps}03$



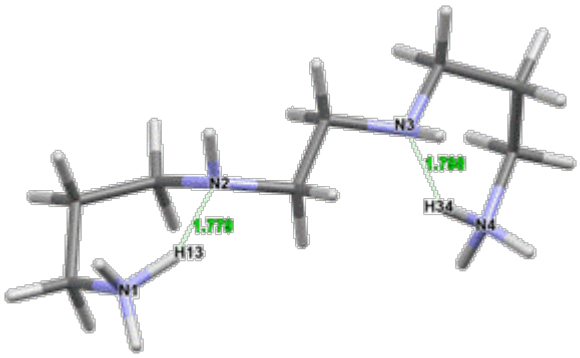
$C_{pp}01$



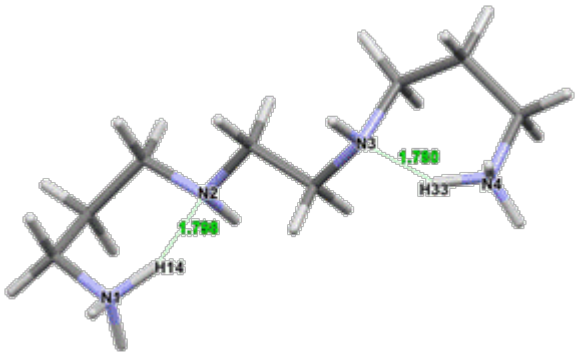
$C_{ps}01$



$C_{ps}14$

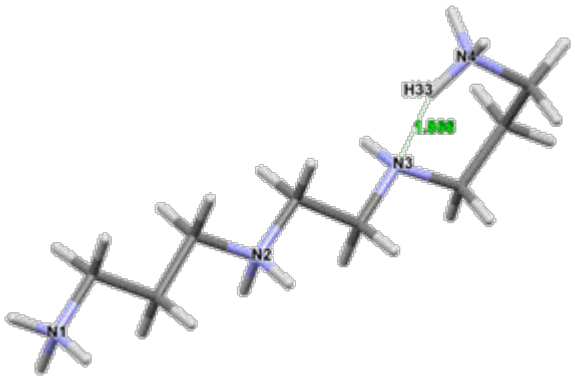


$C_{pp}05$

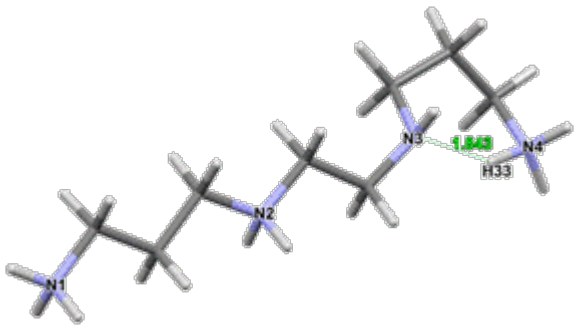


$C_{pp}02$

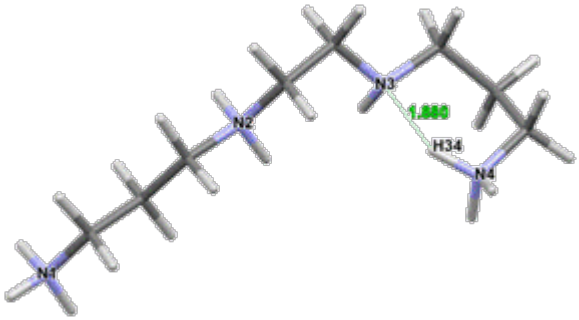
Figure C5. Structures of all lowest in energy conformers for H_2L form of 3,2,3-tet used to calculate protonation constants with continuum solvation model, PCM.



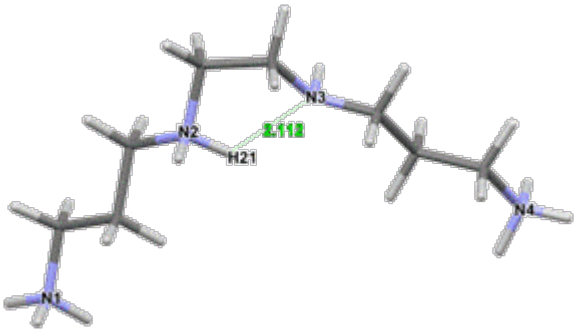
$C_{tp}07$



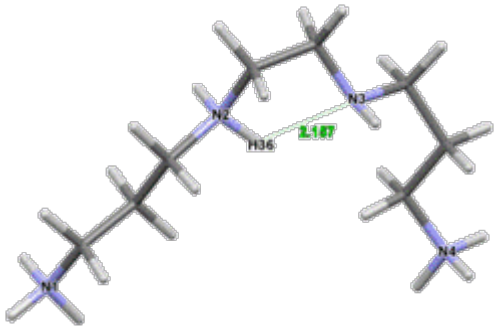
$C_{tp}09$



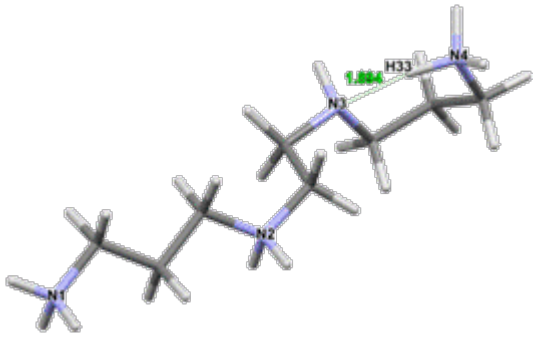
$C_{tp}05$



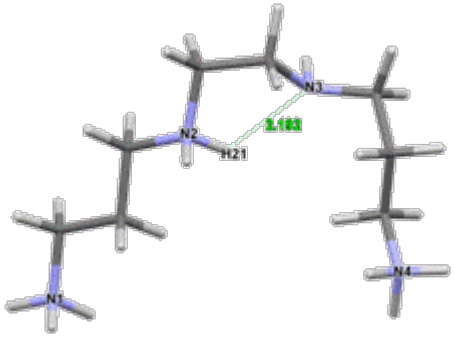
$C_{tp}01$



$C_{tp}02$

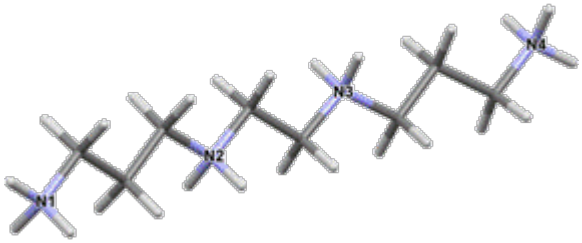


$C_{tp}11$

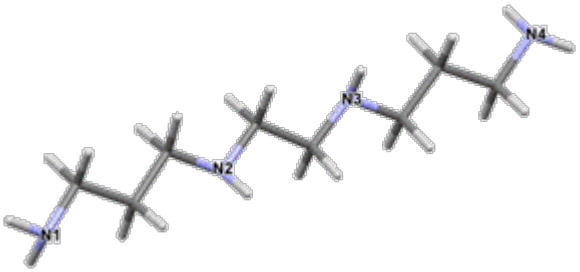


$C_{tp}04$

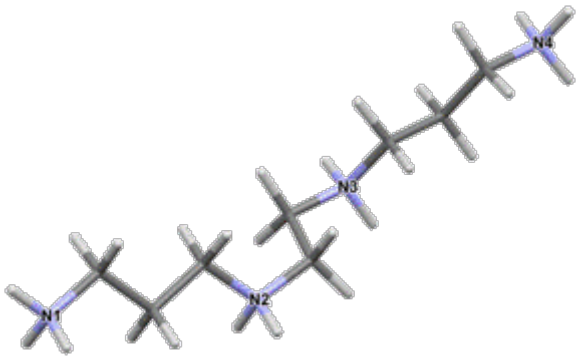
Figure C6. Structures of all lowest in energy conformers for H_3L form 3,2,3-tet used to calculate protonation constants with continuum solvation model, PCM.



$C_{fp}01$

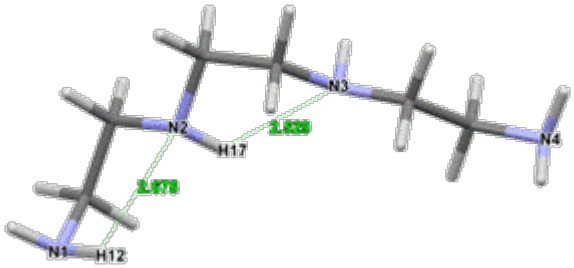


$C_{fp}02$

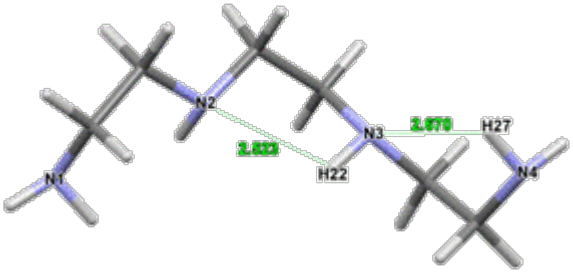


$C_{fp}04$

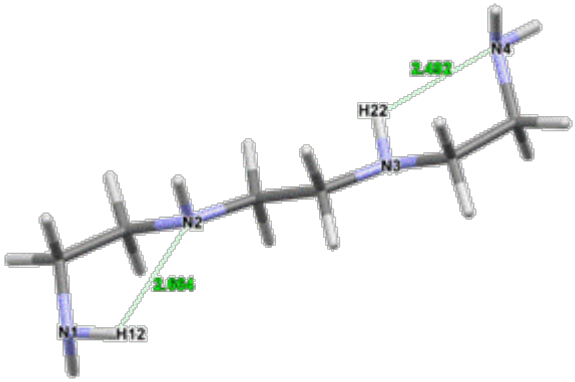
Figure C7. Structures of all lowest in energy conformers for H_4L form of 3,2,3-tet used to calculate protonation constants with continuum solvation model, PCM.



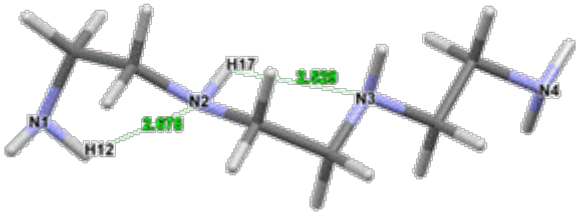
C_L01



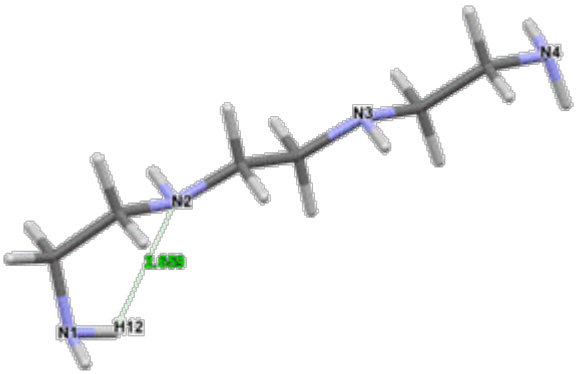
C_L04



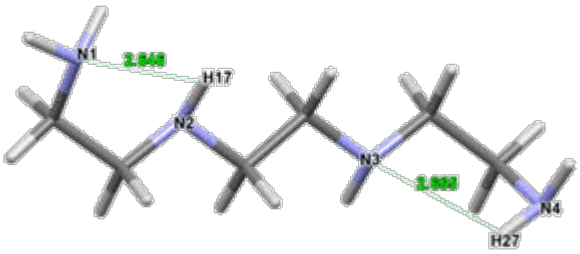
C_L02



C_L03

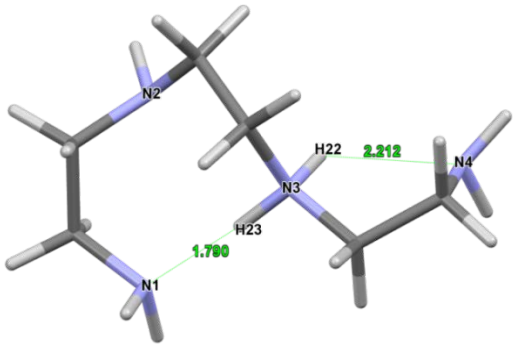


C_L05

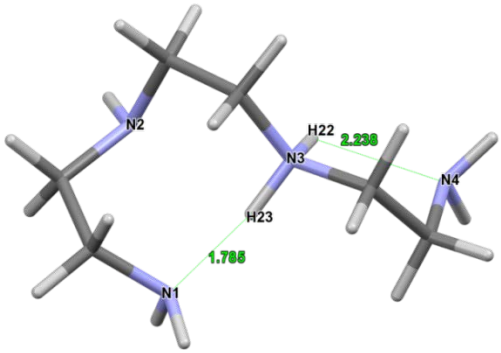


C_L06

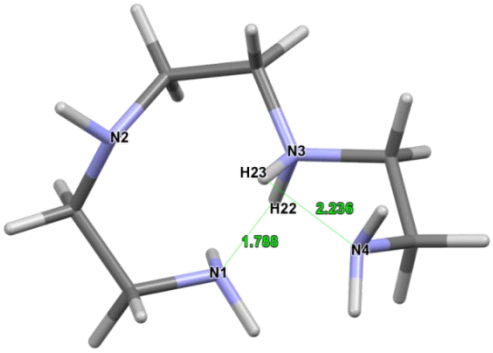
Figure C8. Structures of all lowest in energy conformers for the free ligand of 2,2,2-tet used to calculate protonation constants with continuum solvation model, PCM



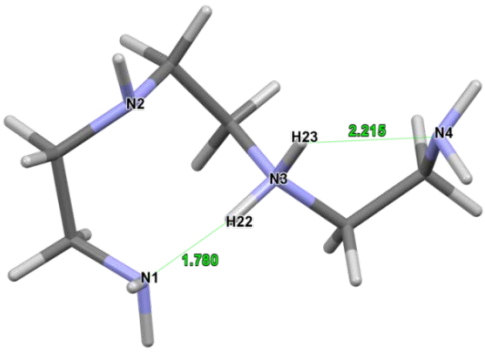
C_s01



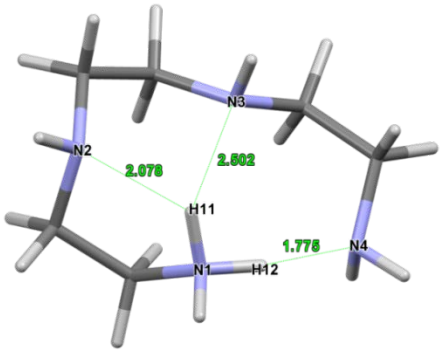
C_s03



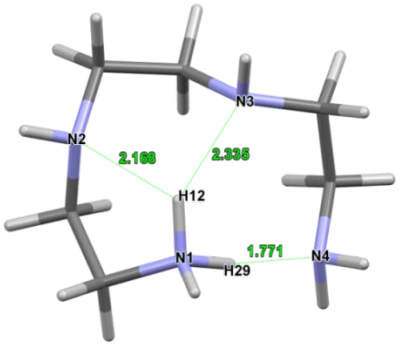
C_s02



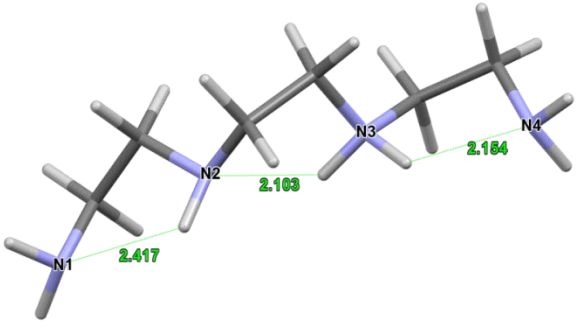
C_s04



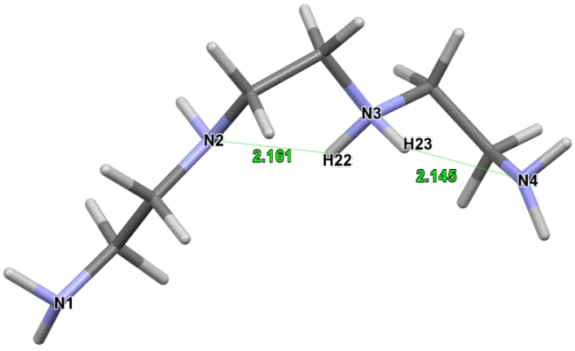
C_p02



C_p01

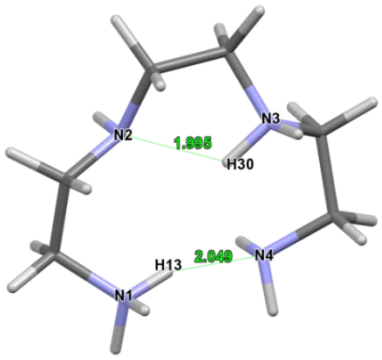


C_s05

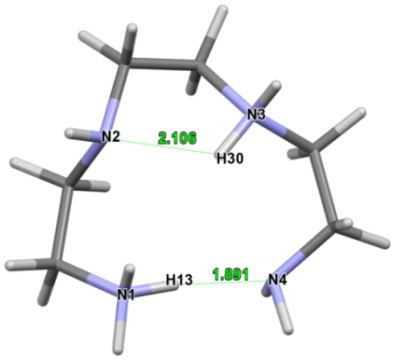


C_s06

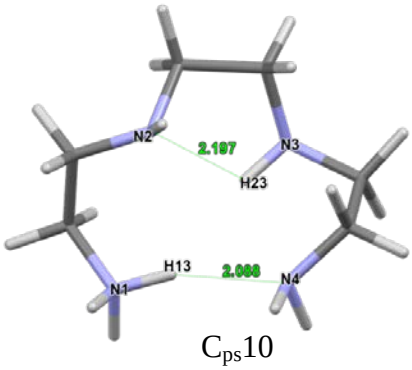
Figure C9. Structures of all lowest in energy conformers for HL form of 2,2,2-tet used to calculate protonation constants with continuum solvation model, PCM.



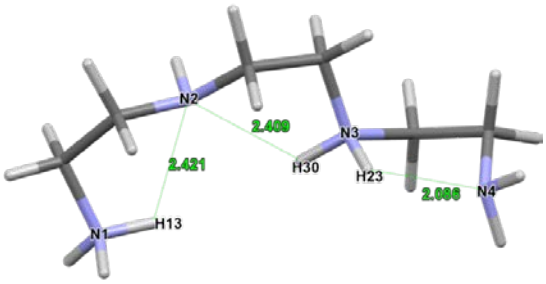
$C_{ps}01$



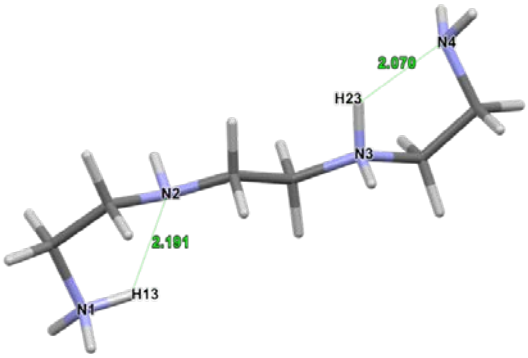
$C_{ps}09$



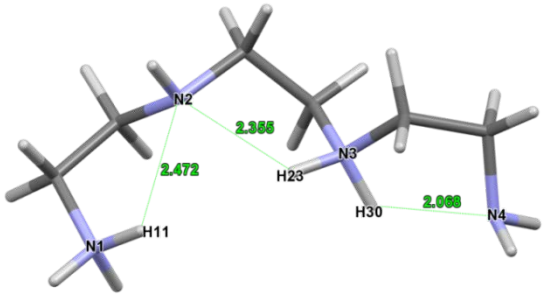
$C_{ps}10$



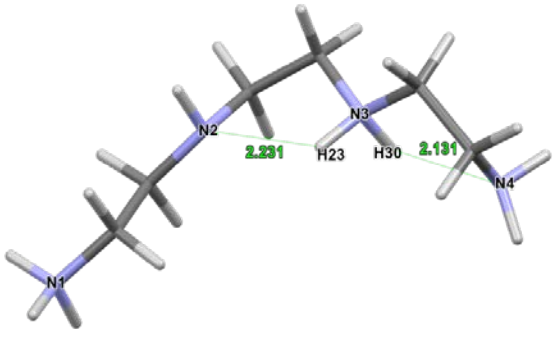
$C_{ps}02$



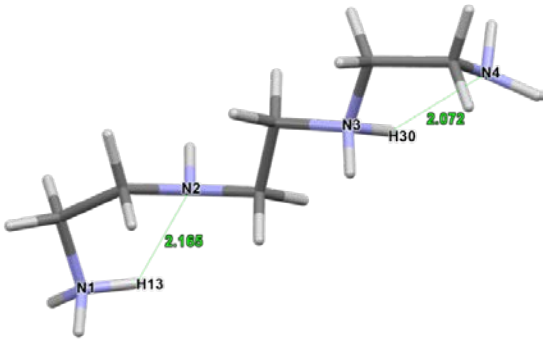
$C_{ps}04$



$C_{ps}06$

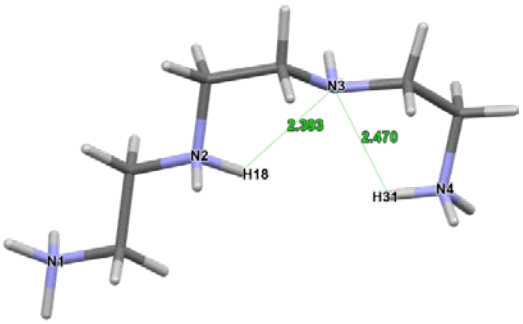


$C_{ps}05$

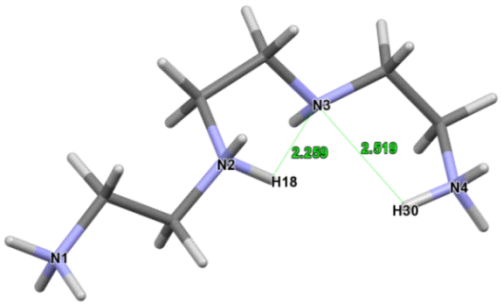


$C_{ps}03$

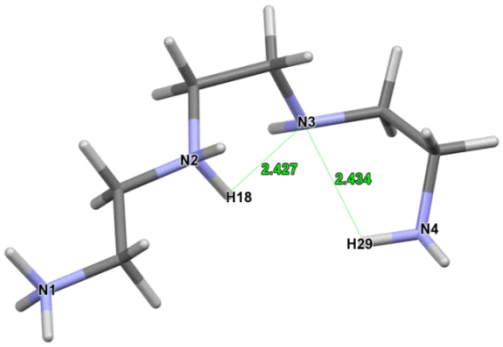
Figure C10. Structures of all lowest in energy conformers for H_2L form of 2,2,2-tet used to calculate protonation constants with continuum solvation model, PCM



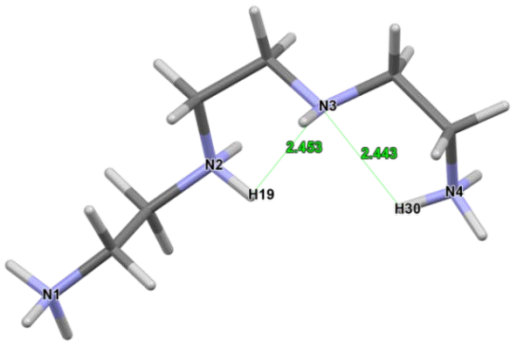
$C_{tp}01$



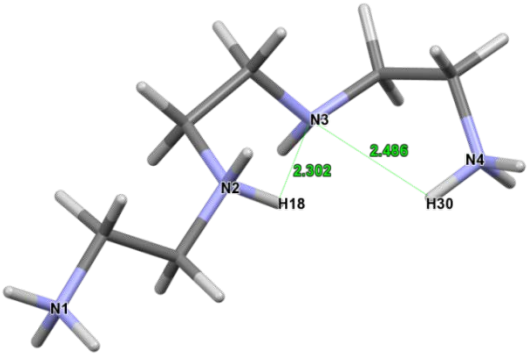
$C_{tp}02$



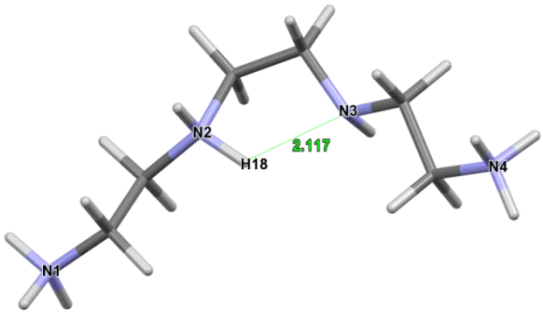
$C_{tp}08$



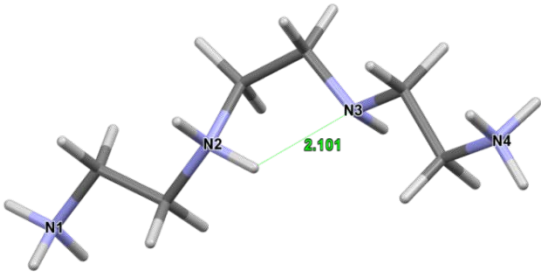
$C_{tp}07$



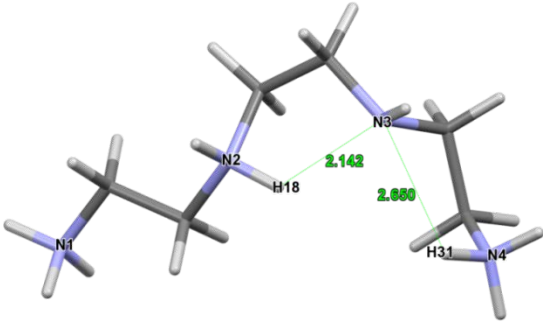
$C_{tp}06$



$C_{tp}03$

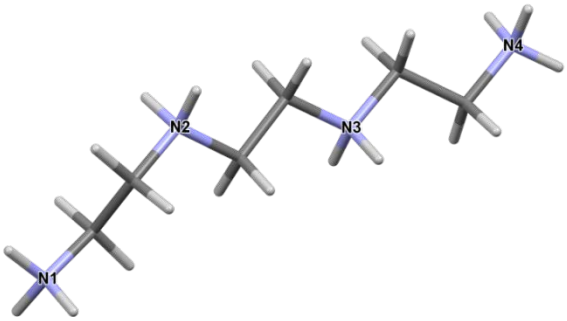


$C_{tp}05$

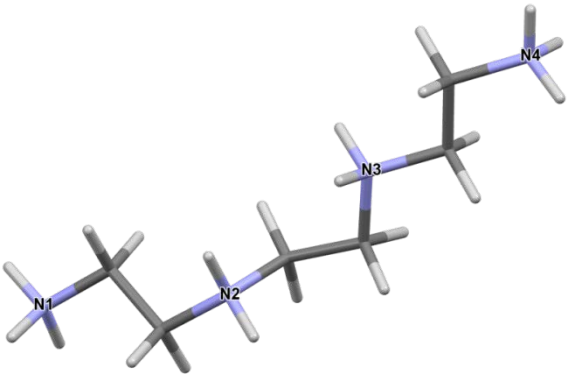


$C_{tp}04$

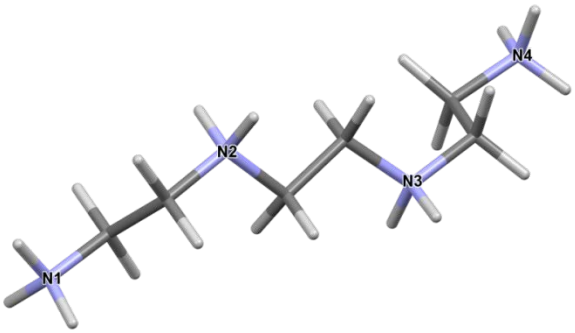
Figure C11. Structures of all lowest in energy conformers for H_3L form of 2,2,2-tet used to calculate protonation constants with continuum solvation model. PCM.



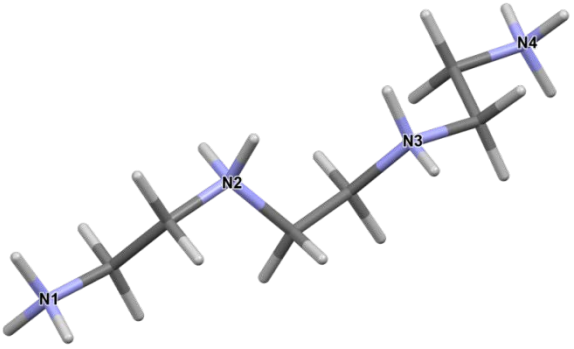
$C_{fp}02$



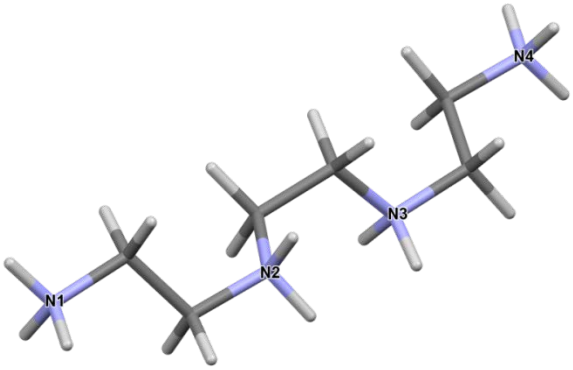
$C_{fp}05$



$C_{fp}03$



$C_{fp}06$

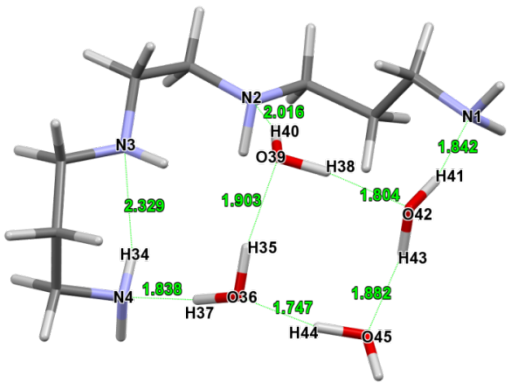


$C_{fp}07$

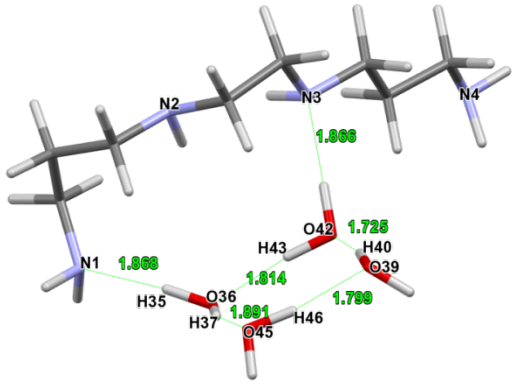
Figure C12. Structures of all lowest in energy conformers for H_4L form of 2,2,2-tet used to calculate protonation constants with continuum solvation model, PCM.

PART 4

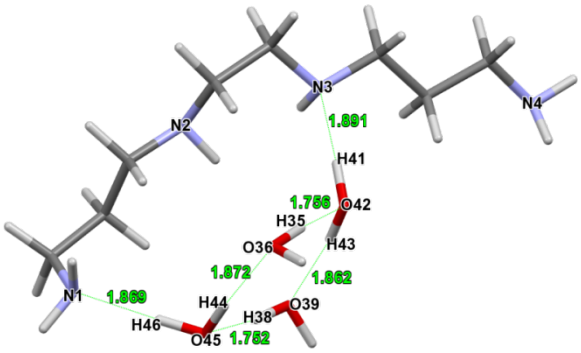
Lowest energy conformers discovered in the discrete-continuum solvation model.



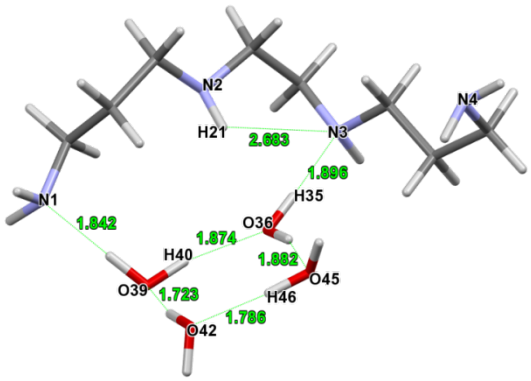
C_L21



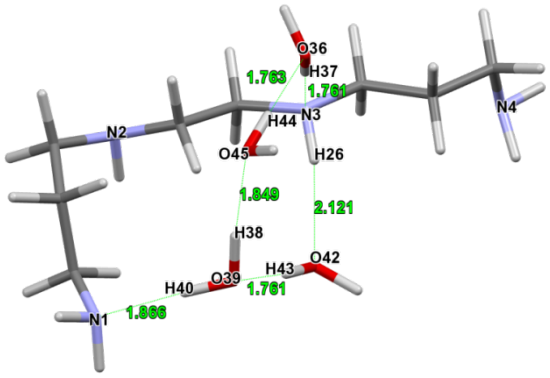
C_L19



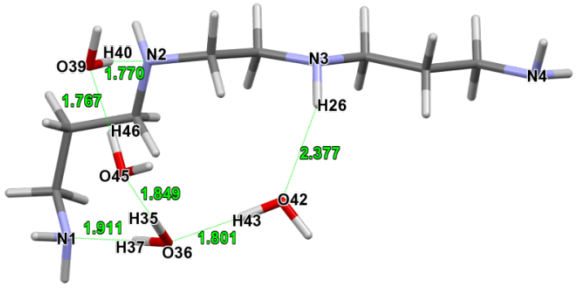
C_L01



C_L09

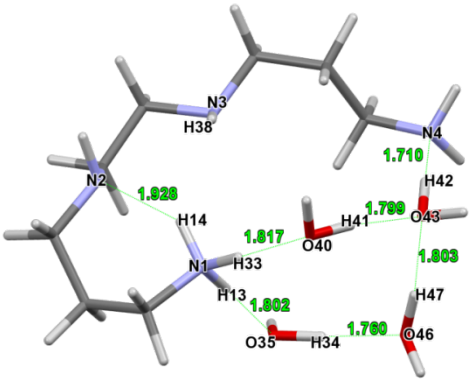


C_L07

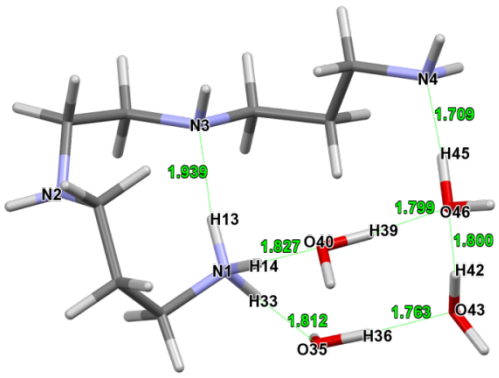


C_L22

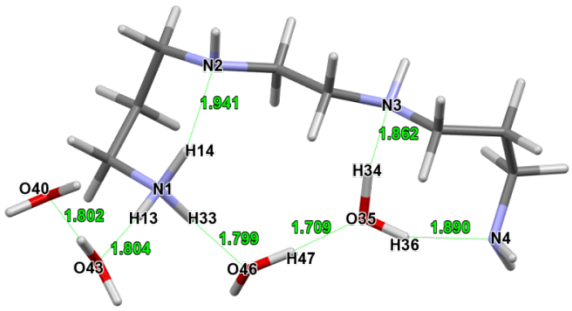
Figure C13. Structures of all lowest in energy conformers of the free ligand of 3,2,3-tet used to calculate protonation constants with discrete-continuum solvation model.



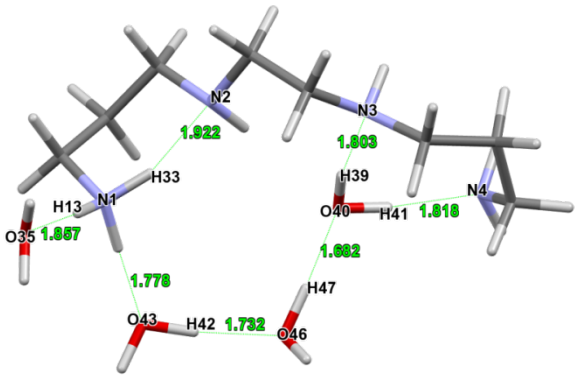
C_p24



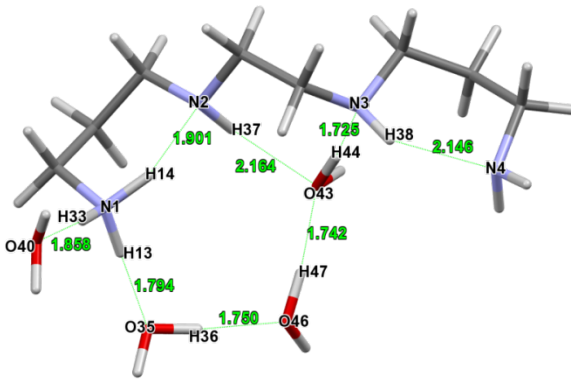
C_p26



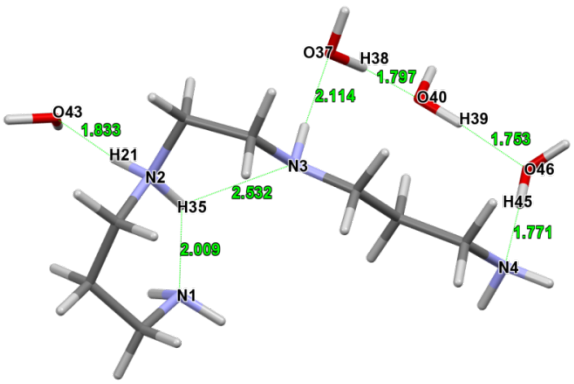
C_p07



C_p25

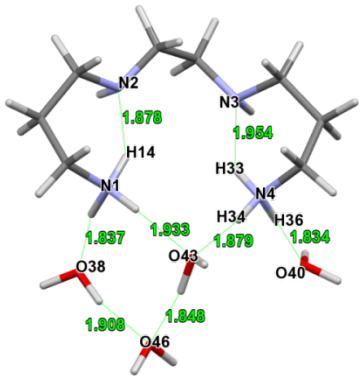


C_p27

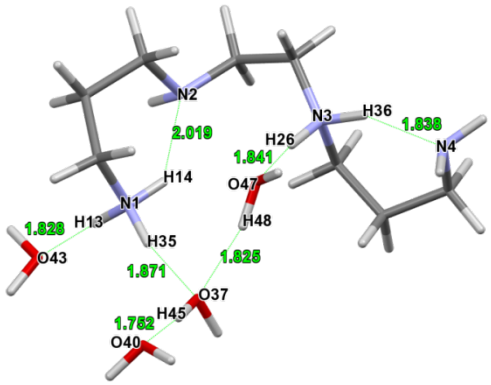


C_s07

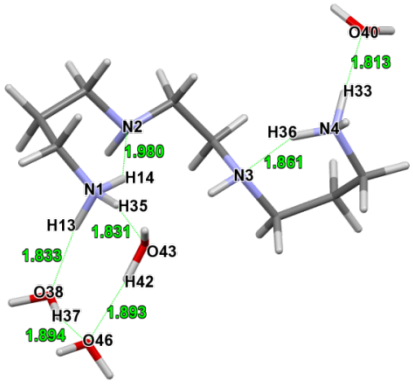
Figure C14. Structures of all lowest in energy conformers for HL form of 3,2,3-tet used to calculate protonation constants with discrete-continuum solvation model.



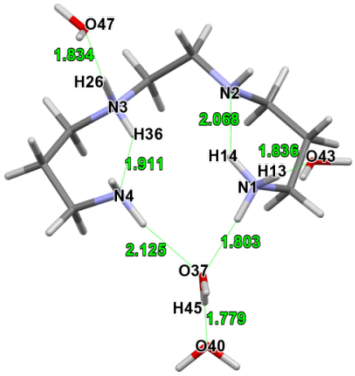
$C_{pp}18$



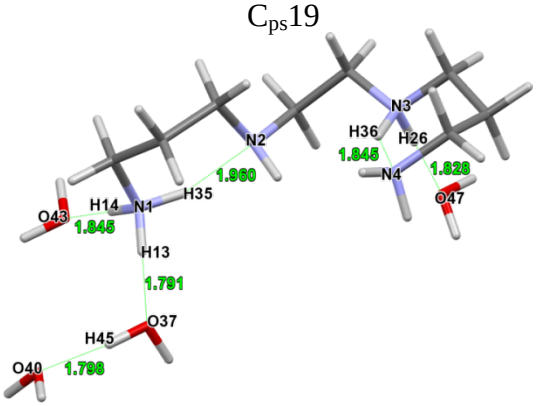
$C_{ps}06$



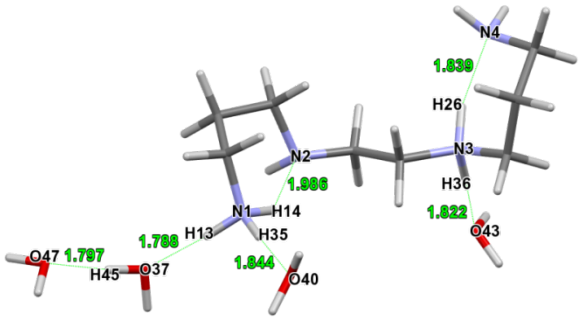
$C_{ps}19$



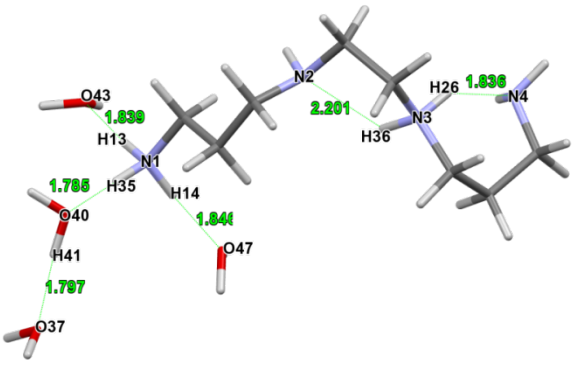
$C_{ps}01$



$C_{ps}10$

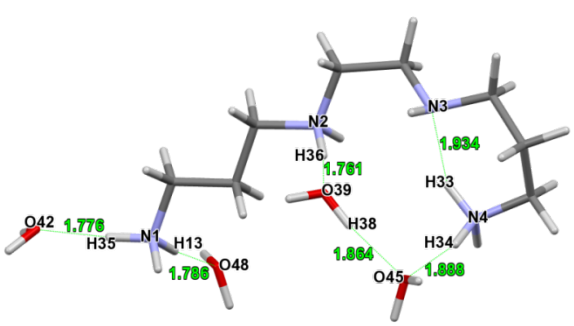


$C_{ps}28$

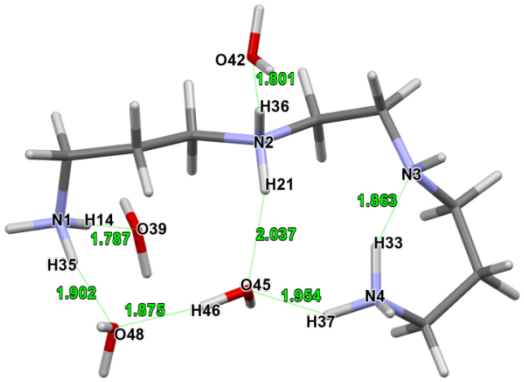


$C_{ps}13$

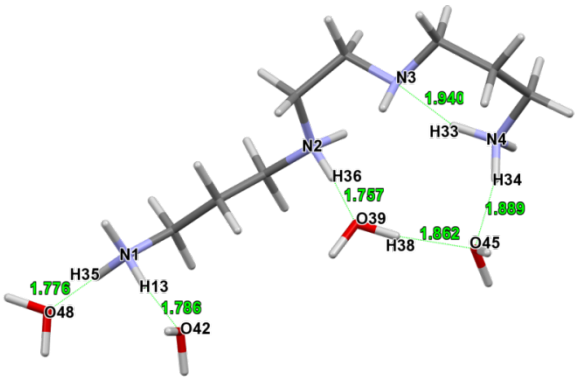
Figure C15. Structures of all lowest in energy conformers of H_2L form of 3,2,3-tet used to calculate protonation constants with discrete-continuum solvation model.



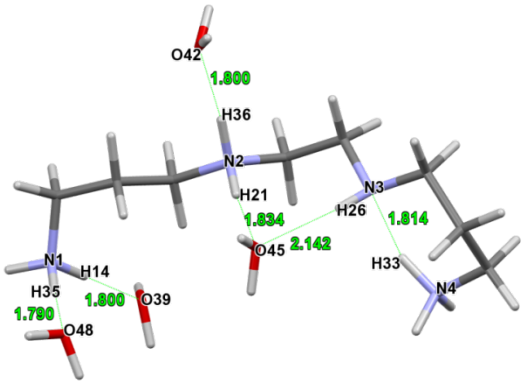
$C_{tp}01$



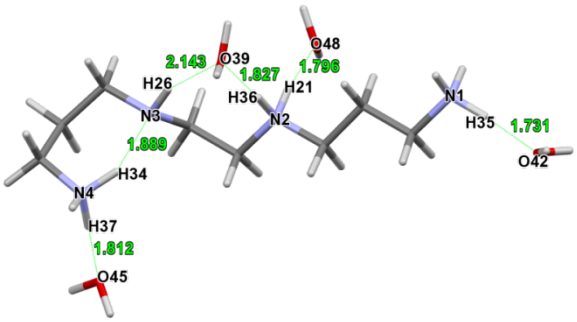
$C_{tp}25$



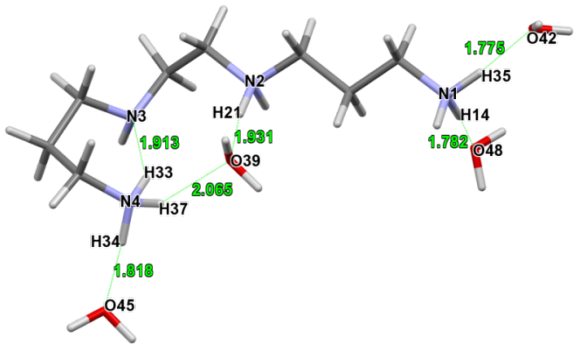
$C_{tp}02$



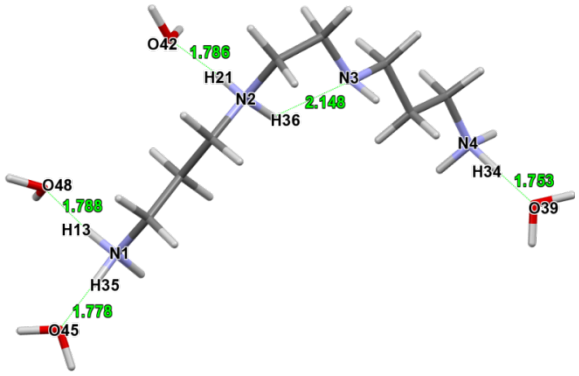
$C_{tp}23$



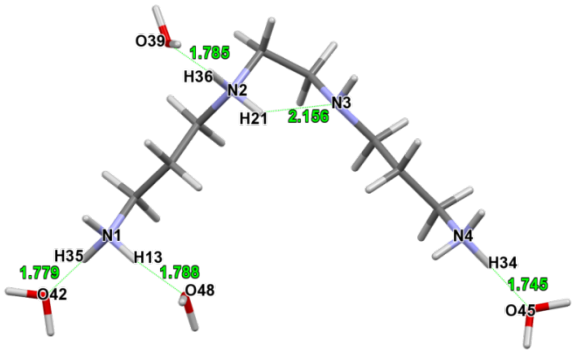
$C_{tp}09$



$C_{tp}03$

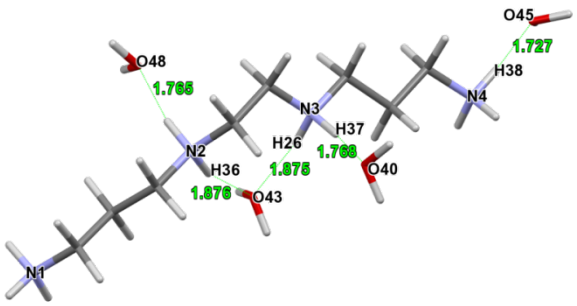


$C_{tp}21$

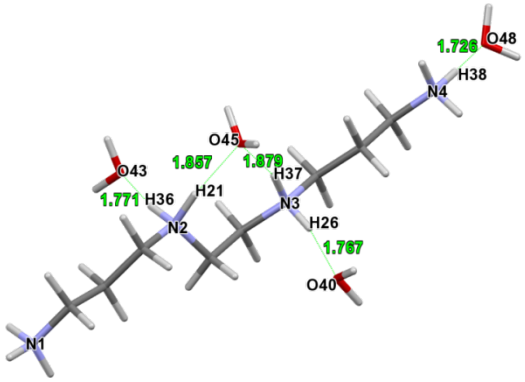


$C_{tp}04$

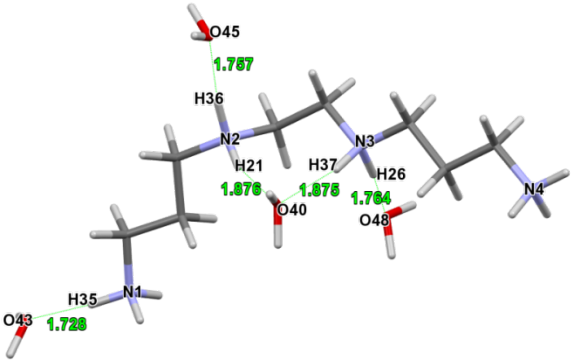
Figure C16. Structures of all lowest in energy conformers for H_3L form of 3,2,3-tet used to calculate protonation constants with discrete-continuum solvation model.



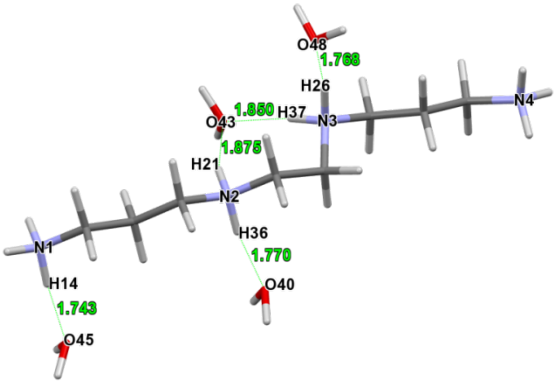
$C_{fp}01$



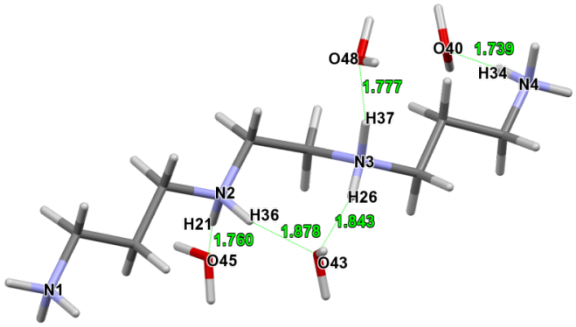
$C_{fp}09$



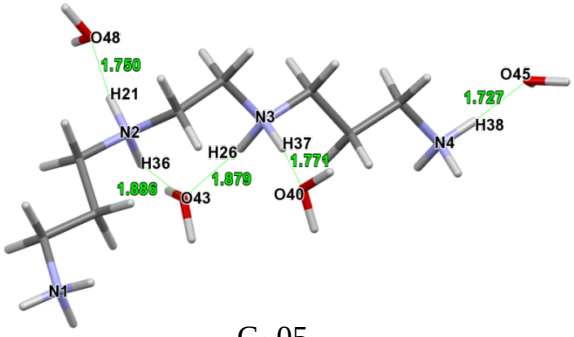
$C_{fp}04$



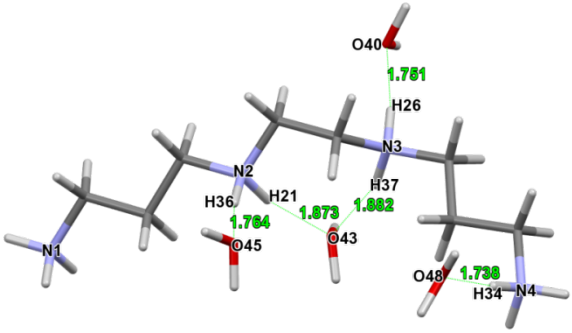
$C_{fp}25$



$C_{fp}24$

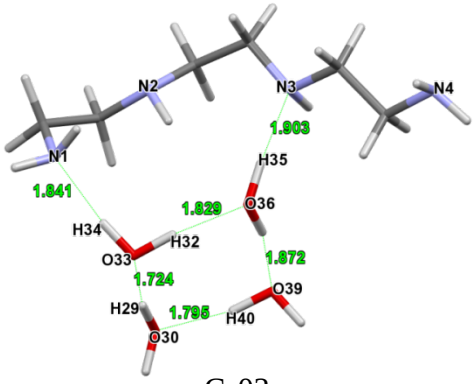


$C_{fp}05$

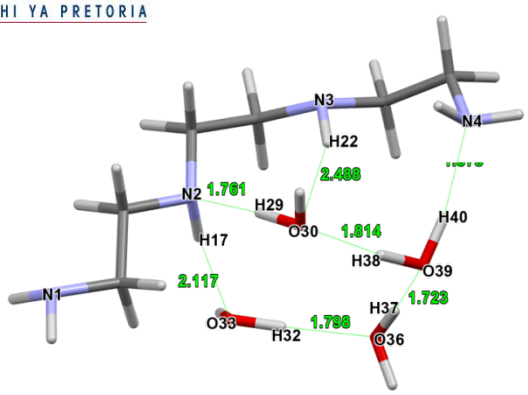


$C_{fp}06$

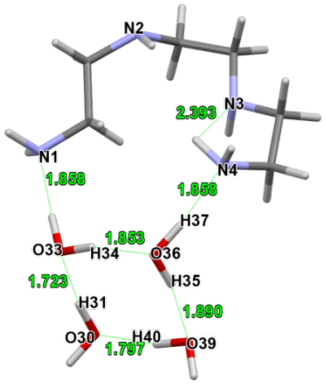
Figure C17. Structures of all lowest in energy conformers of H_4L form of 3,2,3-tet used to calculate protonation constants with discrete-continuum solvation model.



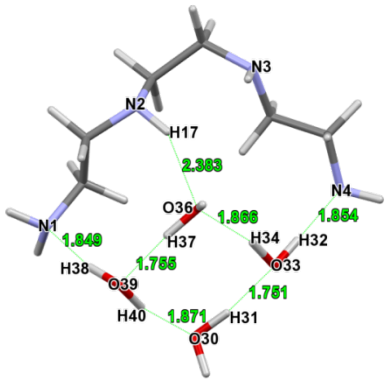
C_L03



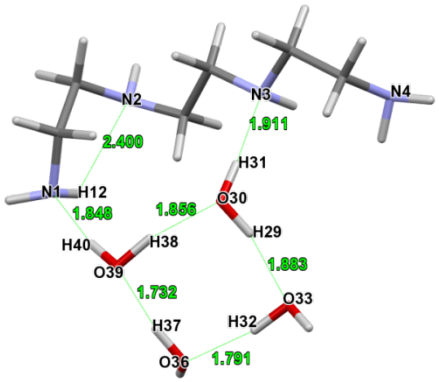
C_L01



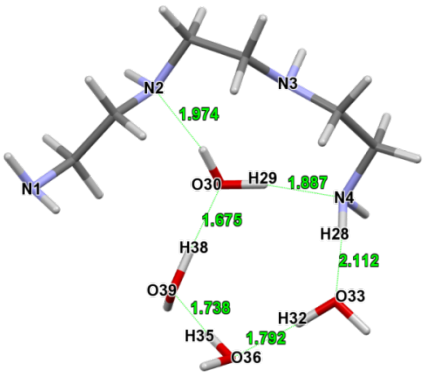
C_L05



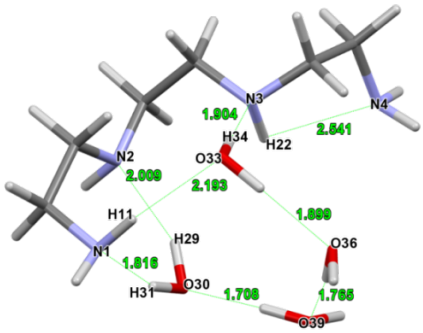
C_L04



C_L06

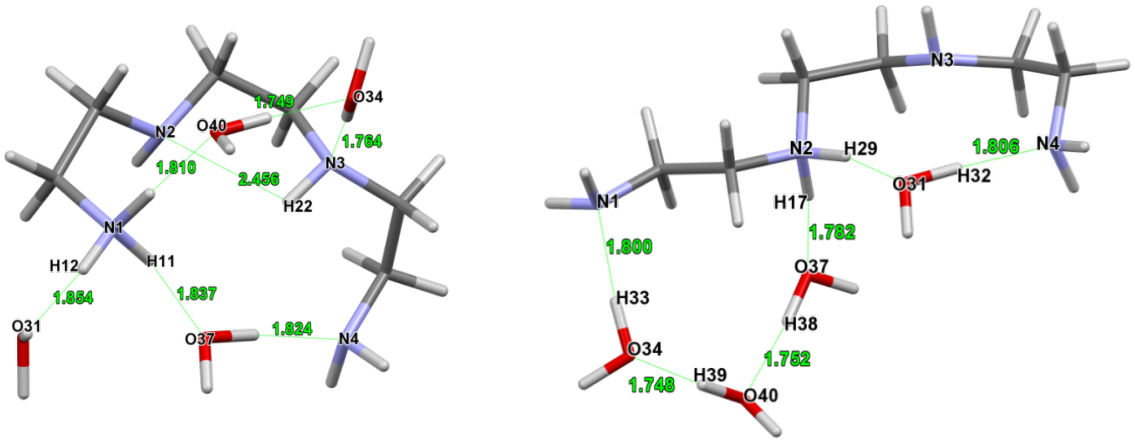


C_L09



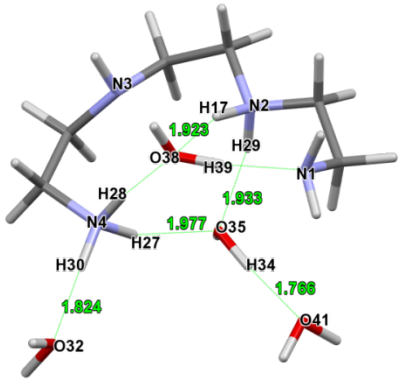
C_L07

Figure C18. Structures of all lowest in energy conformers of the free Ligand of 2,2,2-tet used to calculate protonation constants with discrete-continuum solvation model.

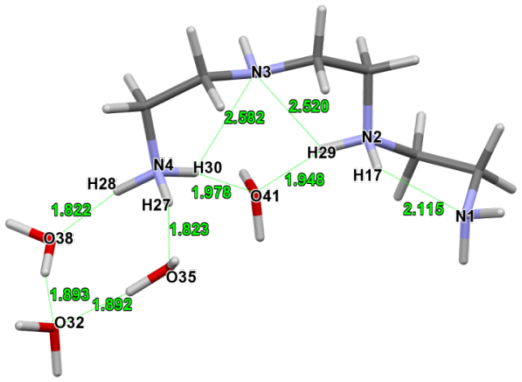


C_p01 C_s07

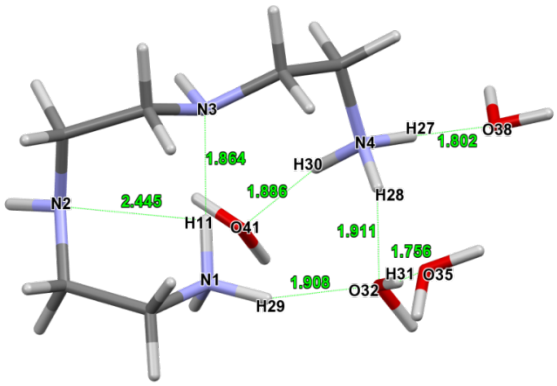
Figure C19. Structures of all lowest in energy conformers for HL form of 2,2,2-tet used to calculate protonation constants with discrete-continuum solvation model.



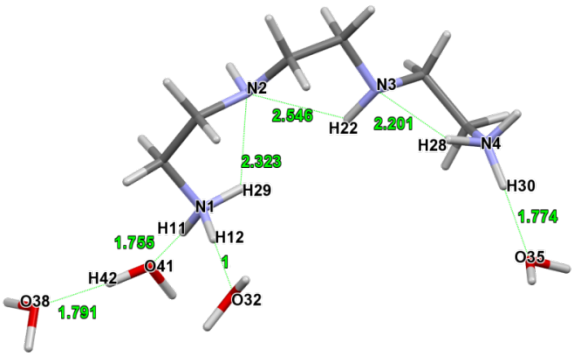
$C_{ps}02$



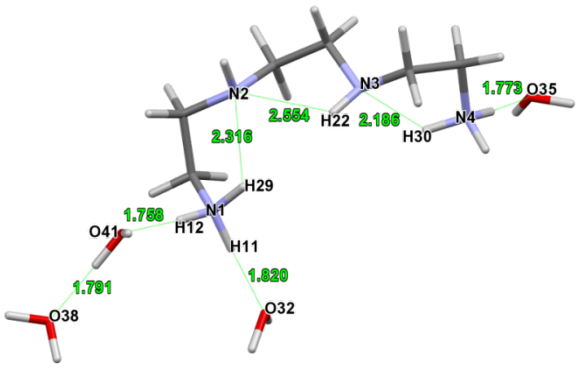
$C_{ps}03$



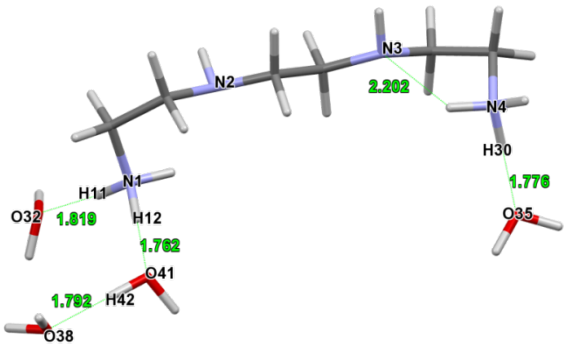
$C_{pp}01$



$C_{pp}07$

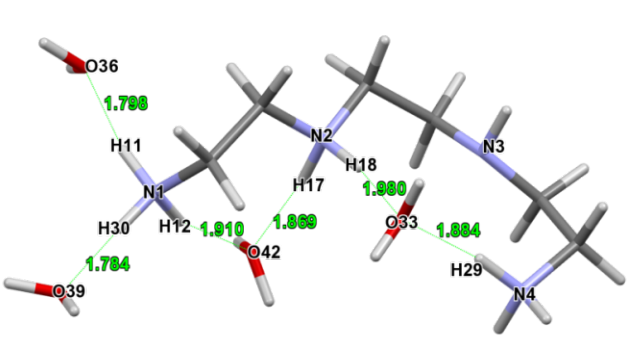


$C_{pp}06$

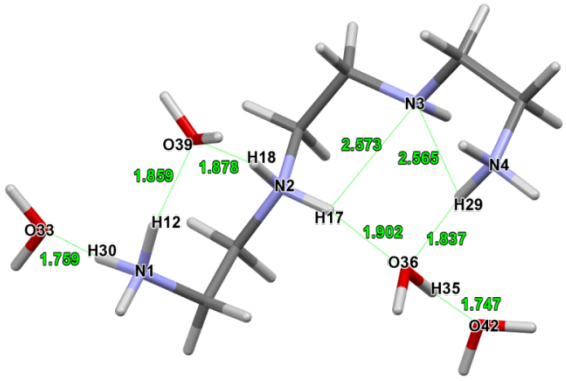


$C_{pp}08$

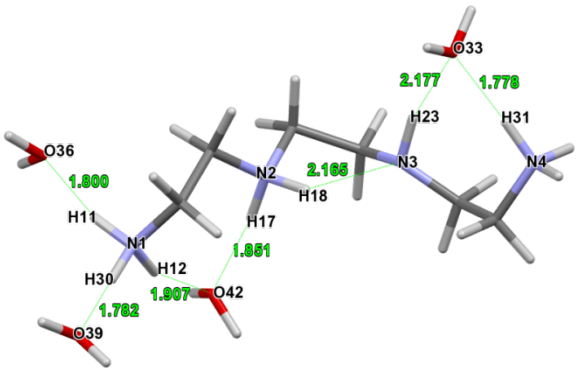
Figure C20. Structures of all lowest in energy conformers for H_2L form of 2,2,2-tet used to calculate protonation constants with discrete-continuum solvation model.



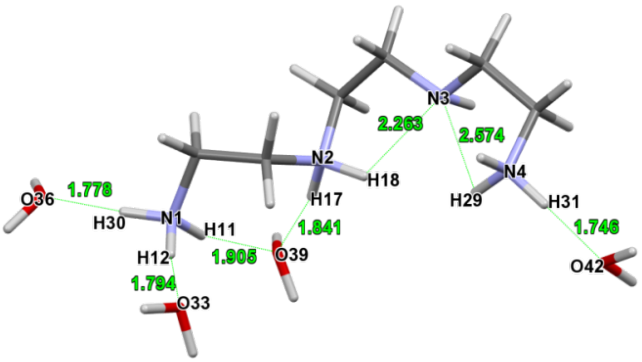
$C_{tp}02$



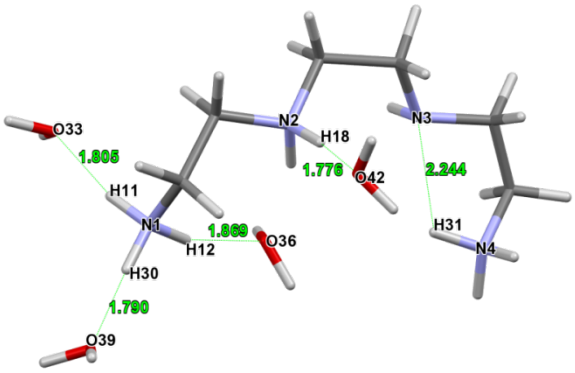
$C_{tp}10$



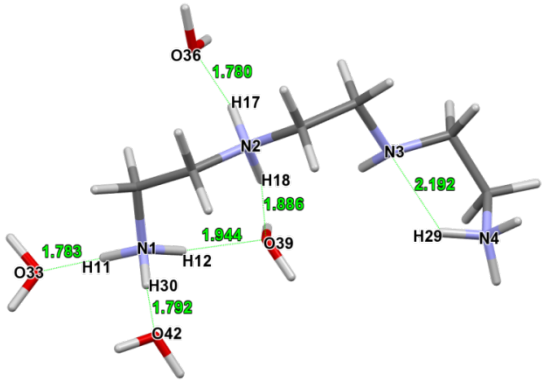
$C_{tp}06$



$C_{tp}01$

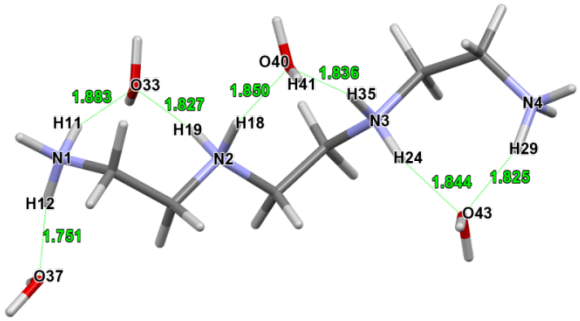


$C_{tp}03$

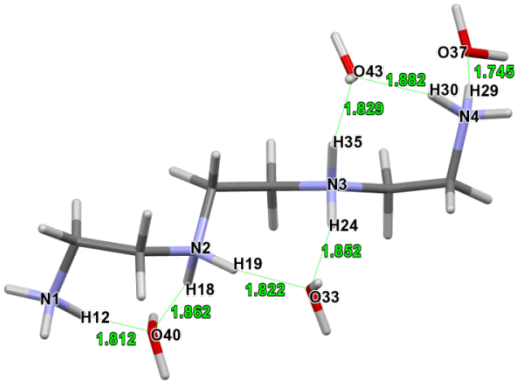


$C_{tp}09$

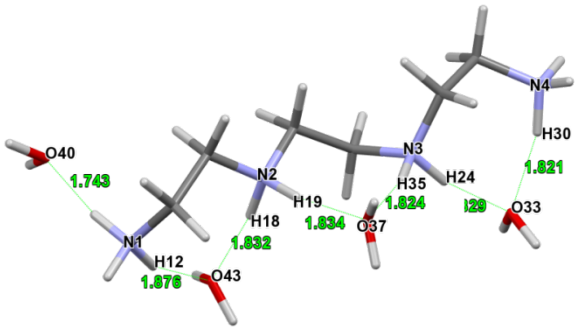
Figure C21. Structures of all lowest in energy conformers for H_3L form of 2,2,2-tet used to calculate protonation constants with discrete-continuum solvation model.



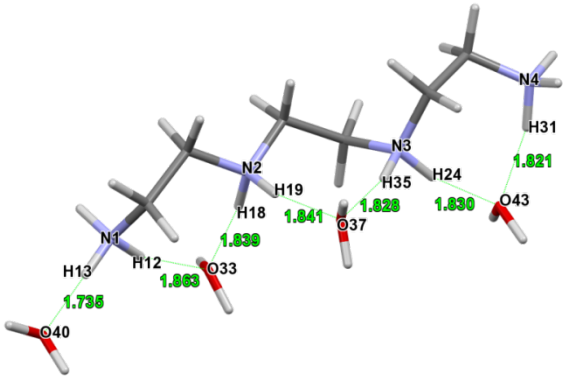
$C_{fp}01$



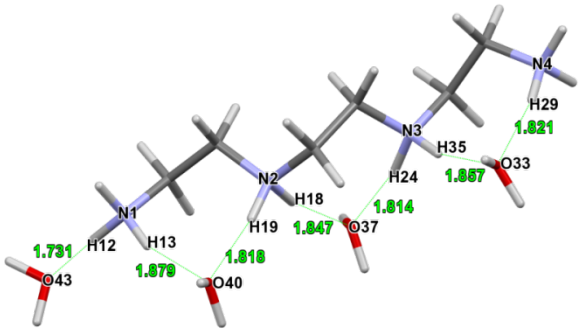
$C_{fp}02$



$C_{fp}04$



$C_{fp}06$



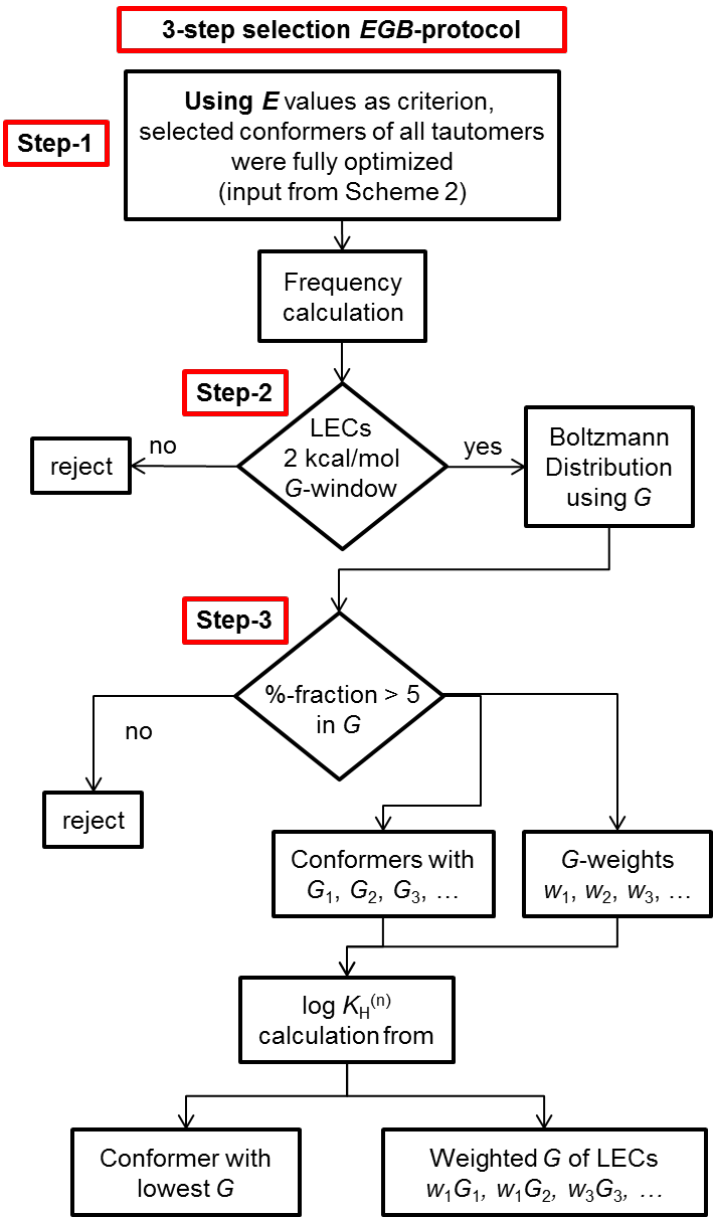
$C_{fp}07$

Figure C22. Structures of all lowest in energy conformers for H_4L form of 2,2,2-tet used to calculate protonation constants with discrete-continuum solvation model.

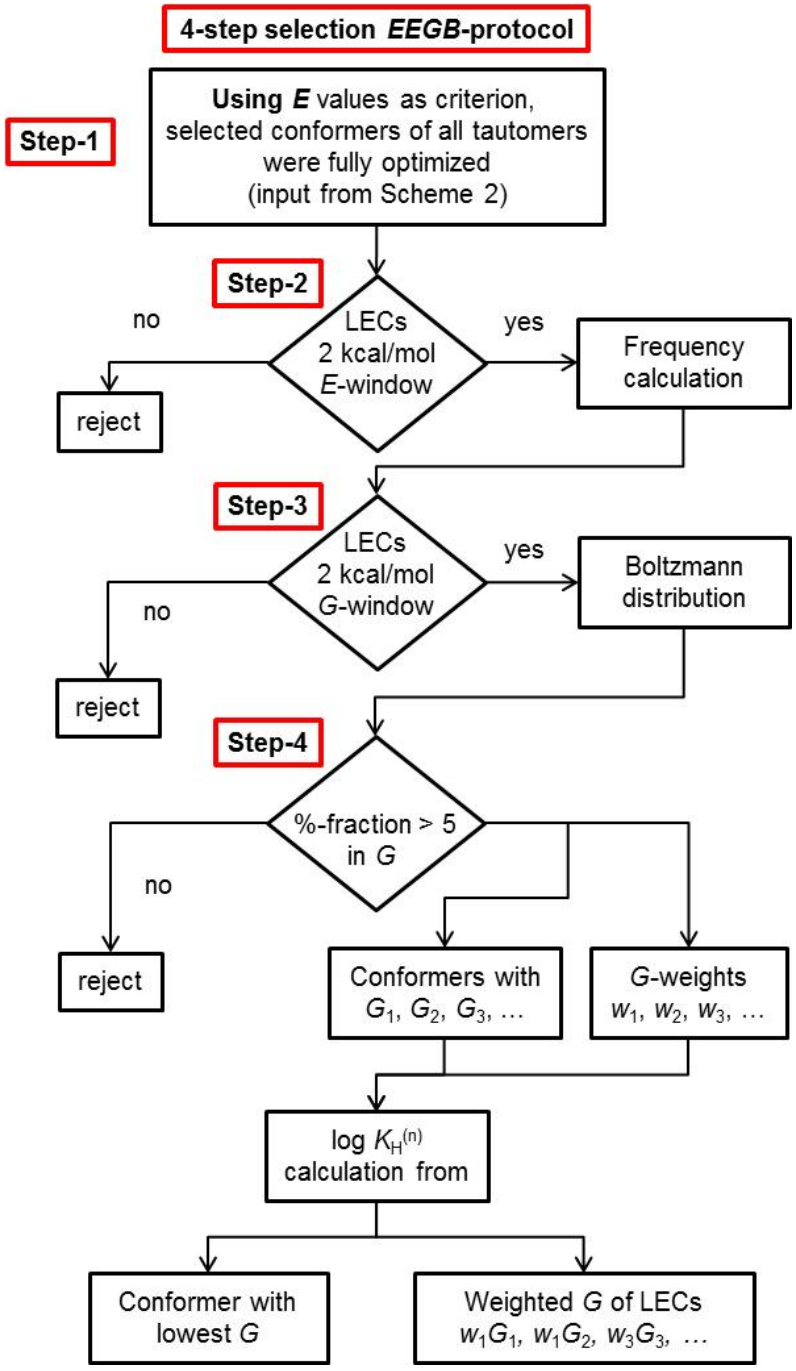
Table C1. Examples of competition reactions in which reference molecule used had either smaller or larger, or similar number of protons relative to the molecule being investigated.

Reaction	$\log K_H^{(n)}$	Δ
$HL^{(1)} + HL^{(2)} = H_2L^{(1)} + L^{(2)}$	15.56	6.56
$H_2L^{(1)} + HL^{(2)} = H_3L^{(1)} + L^{(2)}$	15.47	8.89
$H_3L^{(1)} + HL^{(2)} = H_4L^{(1)} + L^{(2)}$	14.33	11.06
$L^{(1)} + H_2L^{(2)} = HL^{(1)} + HL^{(2)}$	3.01	–6.74
$H_2L^{(1)} + H_2L^{(2)} = H_3L^{(1)} + HL^{(2)}$	8.72	2.14
$H_3L^{(1)} + H_2L^{(2)} = H_4L^{(1)} + HL^{(2)}$	7.58	4.31

Δ = computed – experimental $\log K_H^{(n)}$



Scheme C1. Time most demanding and least accurate 3-step selection *EGB*-protocol tested for protonation constants calculations of polyamines.



Scheme C2. Time efficient and well-performing 4-step selection *EEGB*-protocol tested for protonation constants calculations of polyamines.

Appendix D

Supplementary Information for Chapter 5

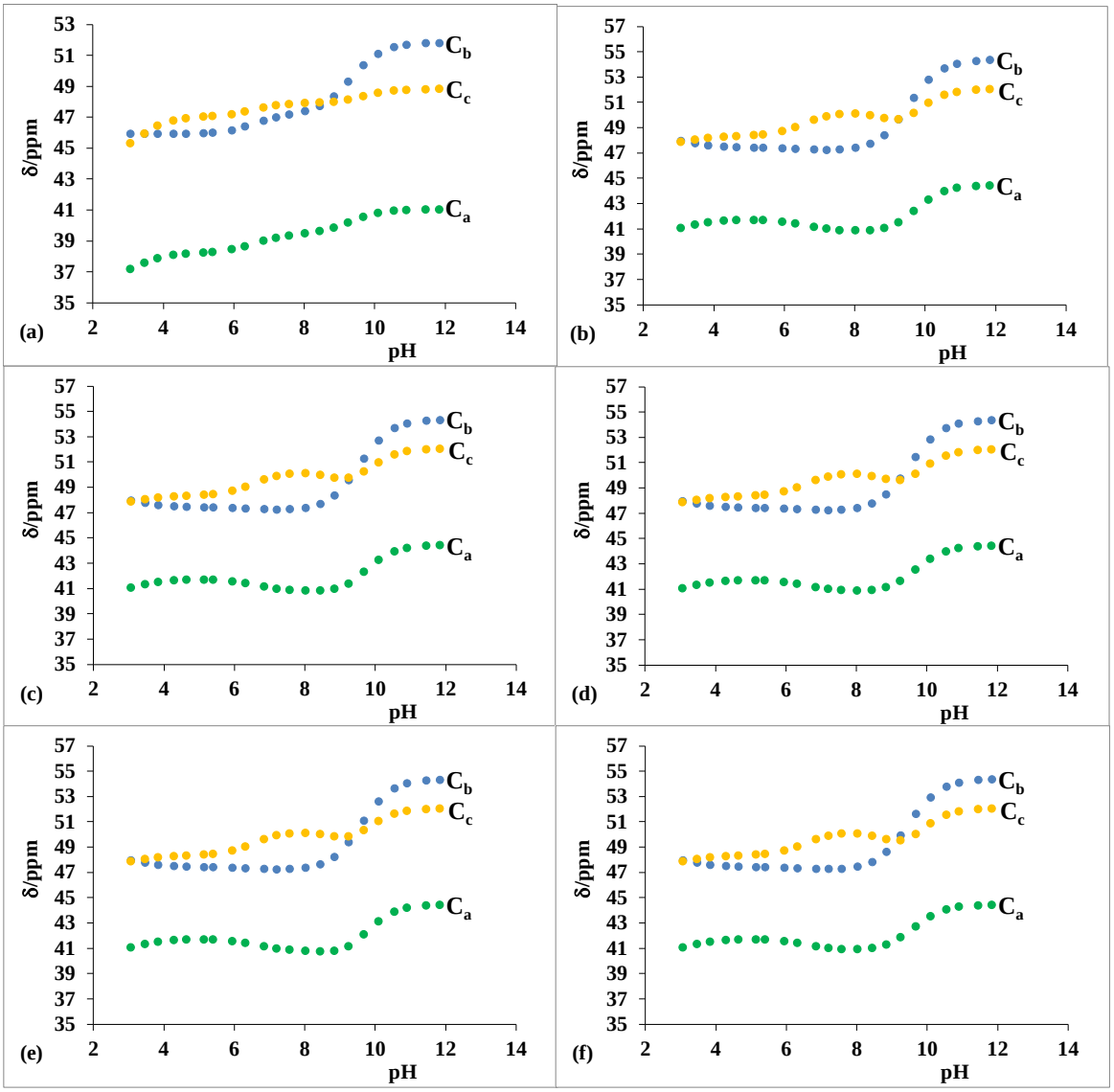


Figure D1. ^{13}C NMR chemical shifts of *trien* as a function of pH using low energy conformers, B3LYP/4H2O+PCM/UFF (a) experiment (b) HL_p: HL_s (1:1) (c) HL_p: HL_s (2:1) (d) HL_p: HL_s (1:2) (e) HL_p only (f) HL_s only.

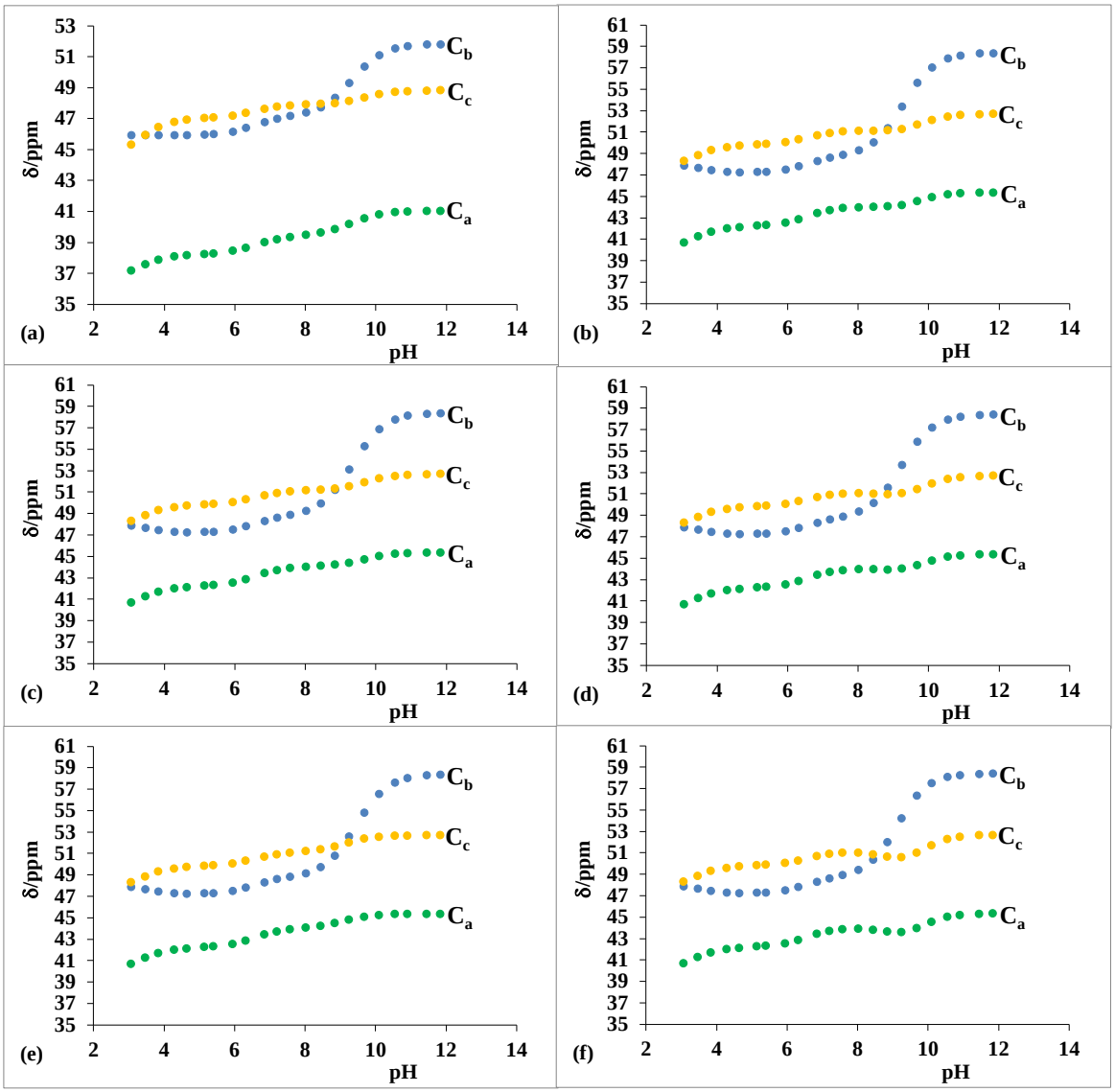
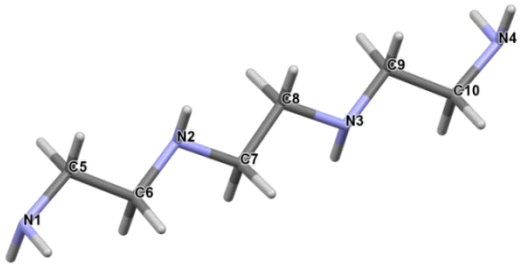
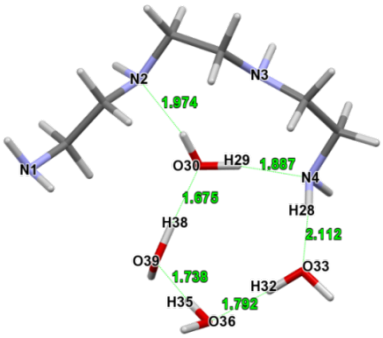


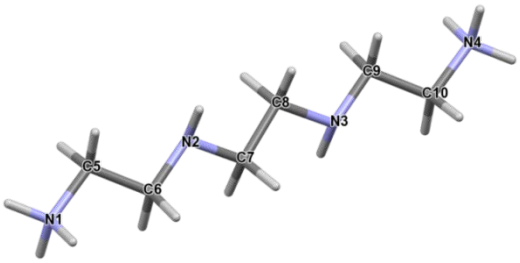
Figure D2. ^{13}C NMR chemical shifts of *trien* as a function of pH using fully linear conformers, B3LYP/PCM/UFF.(a) experiment (b) HL_p: HL_s (1:1) (c) HL_p: HL_s (2:1) (d) HL_p: HL_s (1:2) (e) HL_p only (f) HL_s only,



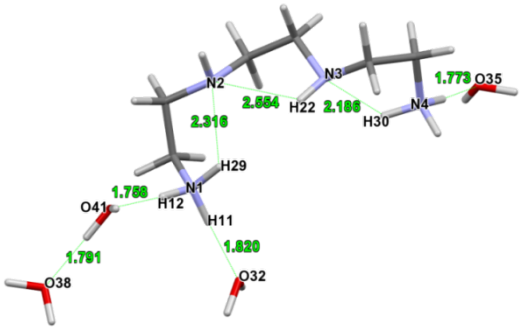
C_L



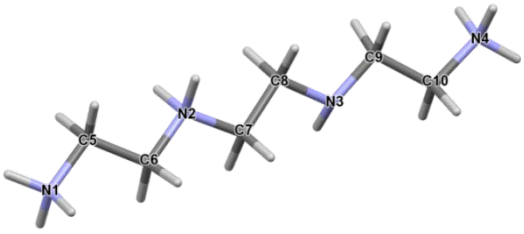
C_L09



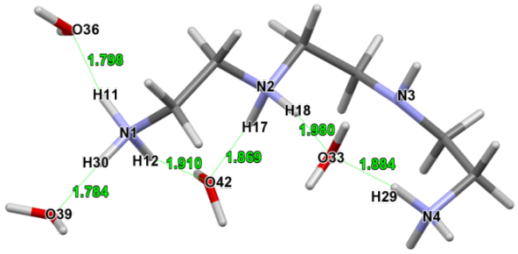
C_{pp}



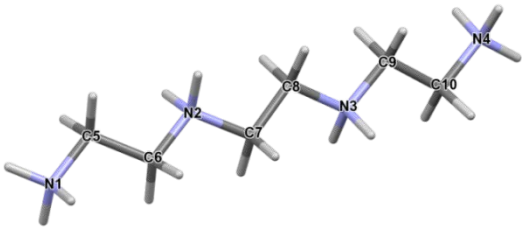
$C_{pp}06$



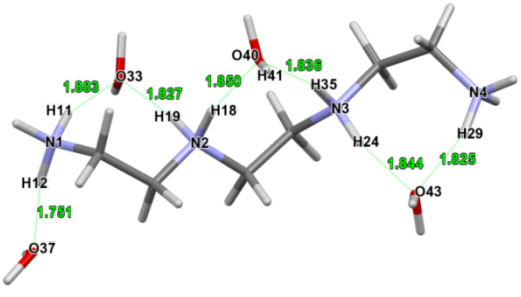
C_{tp}



$C_{tp}02$



C_{fp}



$C_{fp}01$

Figure D3. Structures of fully linear and Low energy conformers of *trien*

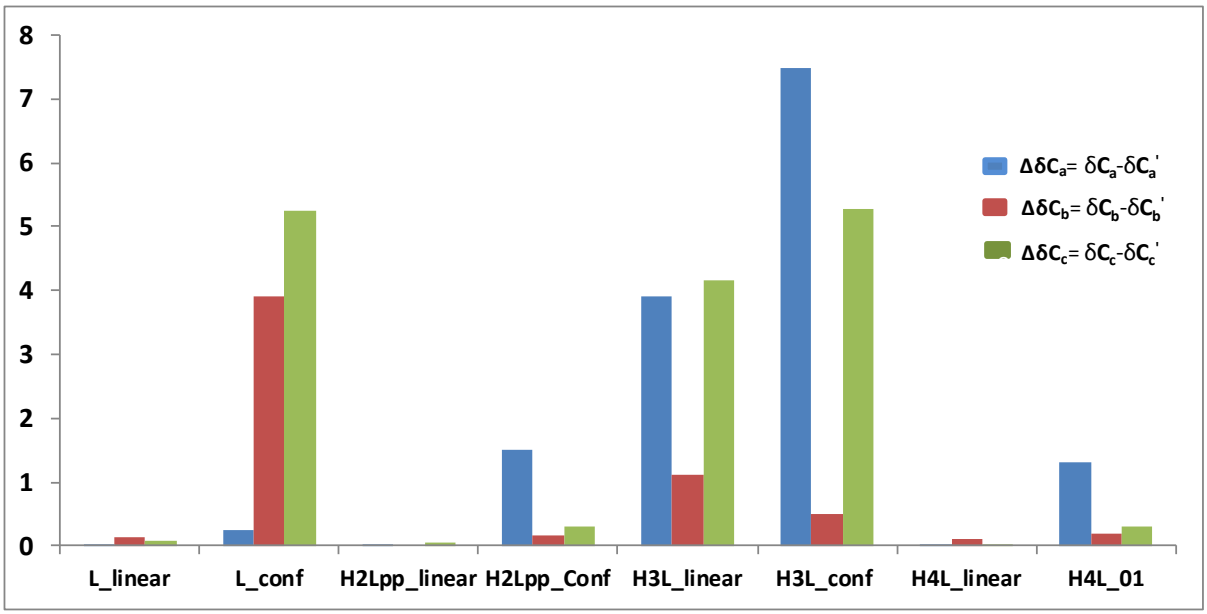


Figure D4. Difference (Δδ) in ^{13}C chemical shift of Equivalent carbon atoms in L, H_2L_{pp}, H_3L and $\mathbf{H_4L}$ tautomers of *trien*.