A Computational Study of Acidic Ionic Liquids for Cellobiose Hydrolysis in Ionic Liquids

Jessica Lisé Nel

Dissertation presented for the degree of

Master of Science

in the Department of Chemistry

UNIVERSITY OF CAPE TOWN

September 2019

Under the supervison of

Dr Gerhard A. Venter

I know the meaning of plagiarism and declare that all of the work in the dissertation, save for that which is properly acknowledged, is my own.

Signed by candidate

Jessica Lisé Nel

September 2019

ABSTRACT

The current environmental situation, with respect to global warming and the ever–approaching depletion of fossil fuel sources, places significance on the development of green fuel and platform chemical production methods. In this context, processes that utilise biomass sources as feedstock, are of great interest. Cellulose, which is the most abundant biopolymer in nature, is a renewable low–cost carbon resource derived from harvest residues and sources like wood and straw. Glucose generation from cellulose requires a saccharide conversion, whereby the *β*-(1,4)-glycosidic bond linkages in the cellobiose polymer repeating units are cleaved. Problems arise in the hydrolysis of cellulose as experimental and theoretical studies have shown cellulose to have very low solubility in water and most other general molecular solvents. This results in the use of harsh pre-treatments at high temperatures and pressures to extract cellulose from lignocellulosic material and strong acids catalysts ($pK_a < -3.2$). Room temperature ionic liquids (RTILs) provide potentially environmentally friendly alternative. It has been shown that ILs can dissolve cellulose under relatively benign conditions and can possibly be adapted into a one-pot-like process of hydrolysis using acid-functionalised IL catalysts.

This dissertation investigated the effect of various ionic liquids on the thermodynamics of cellobiose acid hydrolysis, as both a catalyst and as a solvent, using computational means. An appropriate thermodynamic cycle protocol, a DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP [M05-2X/6-31+G** (SMD)] proton exchange cycle, was established through benchmarking for the prediction of Brønsted acid-functionalised ionic liquid pK_a values in ionic liquids. The sulfonyl-functionalised acidic IL was shown to be the most acidic IL resulting in a lower protonation free energy. Solvation in ionic liquids resulted in higher protonation and barrier height free energies relative to solvation in water.

ACKNOWLEDGEMENTS

I would like to thank the National Research Foundation of South Africa (Grant Number: 106067) for funding this project, as well as the Centre for High Performance Computing (CHPC) for use of their resources.

On a personal note, I could not have done this project without the help and support of so many wonderful people. First and foremost is my amazing supervisor, Dr Gerhard Venter, who reignited my love of chemistry and learning. I thank you for making me a better student, researcher and person, and expanding my thinking on so many things in life. You truly are an extraordinary educator and it has been an absolute privilege to be your student.

I would also like to thank the SCRU Lab and all my lab mates, past and present. You guys make the tough times easier and the good times funnier. Good luck with the hunt for the new baker! I would like to extend a special thank you to Tharindu Senapathi, who has become so much more than a colleague to me. You were a voice of reason when I needed it most. Your words of wisdom are always appreciated and your constant encouragement a lifesaver. I swear I will consider doing a PhD one day, just for you.

To my friends and family, words cannot even describe how much your love and support meant to me. Simone Wills, Simone Renga, Camille Hall, Nolwazi Gcwensa, Matt Del Grande and Reabetswe Dube, thank you for putting up with me. Thank you for calling and messaging and eventually dragging me out of the house. I will organise those meet-ups and hang outs now, I promise!

Jean-Jacques Lenferna de la Motte and Hannah-Ruth Engelbrecht, no I did not forget you. You guys get your own paragraph because quite frankly you are family. Thank you for bringing me back to reality and believing in me. JJ, you changed me for the better, and Hannah, you are the best wusband a girl could ask for. You are the best part of everything I have done in Cape Town and my life is leagues better for having you both in it.

Finally, to my mom and dad, Andrea Le Roux and Jo Nel, this would not have happened without you. Your love is endless and support all encompassing. All that I am is thanks to you and live every day hoping I can make you that little bit prouder.

TABLE OF CONTENTS

LIST OF FIGURES AND SCHEMES

LIST OF TABLES

LIST OF ABBREVIATIONS

$[(HCO_2)^2C_2C_1im]^+$/[IL-COOH]$^+$: 1-(2-carboxyethyl)-3-methyl-1-imidazol-3-ium

$[(HO)^2C_2C_1im]^+$/[IL-OH]$^+$: 1-(2-hydroxyethyl)-3-methyl-1-imidazol-3-ium

$[(HSO_3)^4C_4C_1im]^+$/[IL-SO_3H]$^+$: 3-methyl-1-(4-sulfobutyl)-1-imidazol-3-ium

$[BF_4]^-$: tetrafluoroborate

$[C_1SO_4]^-$: methyl hydrogen sulfate or methylsulfate

$[C_4C_1im]^+$: 1-butyl-3-methyl-1*H*-imidazol-3-ium

$[C_nC_1im]^+$: 1-alkyl-3-methylimidazolium

$[C_nC_1pyrr]^+$: 1-alkyl-1-methylpyrrolidinium

$[C_nP]^+$: alkylphosphonium

$[C_nSO_4]^-$: alkylsulfate

$[HSO_4]^-$: hydrogen sulfate

$[NTf_2]^-$: bis(trifluoromethylsulfonyl)azanide or bis(trifluoromethylsulphonyl)imide

AIL: acidic ionic liquid

BO: Born-Oppenheimer

CBS: complete basis set

CC: coupled cluster

CDS: cavitation, dispersion and solvent structural

CI: configuration interaction

COSMO: conductor-like screening model

C-PCM: conductor-like polarizable continuum

CS_2: carbon disulfide

DFT: density functional theory

DMAc: *N*,*N*-dimethylacetamide

GGA: generalized gradient approximation

GIL: generic ionic liquid

Gn: Gaussian-n

HF: Hartree-Fock

HLC: higher level correction

IL: ionic liquid

INT: intermediate

KS: Kohn-Sham

LCAO: linear combination of atomic orbitals

LDA: local density approximation

LiCl: lithium chloride

LoT: level of theory

LPNO: local pair natural orbital

LSDA: local spin density approximation

MAD: mean absolute deviation

MAE: mean absolute error

MC: monomer cycle

MD: molecular dynamics

MM: molecular mechanics

MO: molecular orbital

MP: Møller-Plesset

NaOH: sodium hydroxide

NBO: natural bond orbital

NMMO: *N*-methylmorpholine-*N*-oxide

NMR: nuclear magnetic resonance

PAO: projected atomic orbital

PB: Poisson-Boltzmann

PCM: polarizable continuum solvent model

PGP: partial generic parameters

PNO: pair natural orbital

PX: proton exchange

QM: quantum mechanics

RMSD: root-mean-square deviation

RTIL: room temperature ionic liquid

SAS: solvent accessible surface

SD: standard deviation

SMD: solvation model based on solute electron density or solvation model based on density

TC: thermodynamic cycle

TRS: total reducing sugar

TS: transition state

XC: exchange-correlation

GLOSSARY

$\Delta G^{\circ\rightarrow *}$: standard state correction term. Free energy change of n moles of an ideal gas from 1 atm standard state concentration to 1 mol L^{-1} (1 M).

ΔG_{solv}: the free energy of transfer of 1 mole of solute from a fixed position in an ideal gas state to a fixed position in an infinitely dilute solution

K_a: acid dissociation equilibrium constant

$\widetilde{R}$: universal gas constant in units of L atm K^{-1} mol^{-1} (0.082057 L atm K^{-1} mol^{-1})

μ_{gas}: gas phase chemical potential

μ_{soln}: chemical potential in solution

μ°: the standard molar chemical potential of some ideal gas

6-31+G**: Pople split-valence double-zeta basis set with p polarization functions for hydrogen d polarization and diffuse s and p functions on non-hydrogen atoms

6-31G*: Pople split-valence double-zeta basis set with added d polarization

B3LYP: Becke's three parameter exchange functional with the correlation functional of Lee, Yang and Parr

CBS-QB3: CBS-Q modified to use B3LYP hybrid DFT optimized geometries and frequencies

cc-pVTZ: Dunning correlation-consistent, polarized valence, triple-zeta basis set

DLPNO-CCSD(T): domain based local pair natural orbital coupled cluster method with single-, double-, and perturbative triple excitations

G3B3(MP2): G3 theory with reduced perturbation theory and DFT B3LYP geometries

G4(MP2): G4 theory with reduced perturbation theory

$\hat{H}$: Hamiltonian operator

h: Plank's constant

HA: acid of interest

HRef: reference acid

k: Boltzmann's constant

m_e: electron mass

M_k: mass of nucleus k

nS: cluster formed of independent solvent monomers

R: universal gas constant in units of J K^{-1} mol^{-1} (8.314472 (15) J K^{-1} mol^{-1})

Ref^-: reference acid conjugate base

S_n: cluster of interacting solvent monomers/cluster of only solvent molecules

S_N2: bimolecular nucleophilic substitution

T: temperature in units of Kelvin

ε: dielectric constant

1.

Introduction

The current environmental situation, with respect to global warming and the ever-approaching depletion of fossil fuel sources, places significance on the development of green fuel and platform chemical production methods.[1,2] In this context, bio-refinery models, that utilise biomass sources as feed-stocks,[3] are of great interest.

Starchy biomass or sugarcane related vegetation are efficient and cost-effective sources of glucose, a versatile platform chemical base or bio-fuel source. However, their use infringes on agricultural production and land usage.[3–5] Alternatively, cellulose, which is the most abundant component of lignocellulose derived from harvest residues and non-food sources like wood and straw,[3] is essentially a renewable low-cost carbon resource.[6,7] However, their use in glucose generation requires a saccharide conversion, whereby the β-(1,4)-glycosidic bond linkages of the basic cellobiose units making up the long saccharide chains are cleaved.[7,8] This can be achieved through pyrolysis, enzymatic and acidic hydrolysis processes.[5]

Problems arise in the hydrolysis of cellulose where firstly, experimental and theoretical studies have shown cellulose to have very low solubility in water and most other general molecular solvents.[7,9] This results in the use of harsh pre-treatments at high temperatures and pressures.[10,11] Additionally, a substantial energy barrier is observed for the uncatalyzed reaction under aqueous conditions necessitating the use of strong acids, $pK_a < -3$.[8] Overall, this renders the process costly, inefficient and counteractive to its overall environmentally-friendly purpose.

Recently, a favourable medium of cost efficiency and environmental “friendliness” has been found in the use of ionic liquids (ILs). Typically consisting of organic cations and inorganic or organic anions,[12] ILs are considered eco-friendly, designer solvents.[10] In relation to cellulose, they allow for a reduction in activation energy compared to water and facilitate the use of milder conditions.[13–15] Additionally, they improve the biorefinery processing through effective dissolution and separation of lignocellulosic matter and possible application into a one-pot-like process of hydrolysis.[10] This has been proposed in the use of Brønsted-acid functionalised ILs, whereby the ILs serve as both the solvent and the catalyst.[11,16,17]

Thus, this study aims to theoretically investigate the effect ionic liquids have on the thermodynamics of cellobiose acid hydrolysis, both as a solvent and as a catalyst. Specifically, through the determination of a comparable measure of acidity in ionic liquids for different Brønsted acid functionalized ionic liquids used in cellulose hydrolysis; and assessment of the effect of ionic liquid solvation on the cellobiose acid hydrolysis mechanism using implicit models, such as the solvation model based on density (SMD).[18]

1.1 Lignocellulosic Biomass

Lignocellulose is mainly composed of a chemically and physically bonded mixture of the polymeric carbohydrates hemicellulose and cellulose, and the aromatic polymer lignin (Figure 1.1),[10] and as previously stated can be derived from non-food source plant matter.[3]

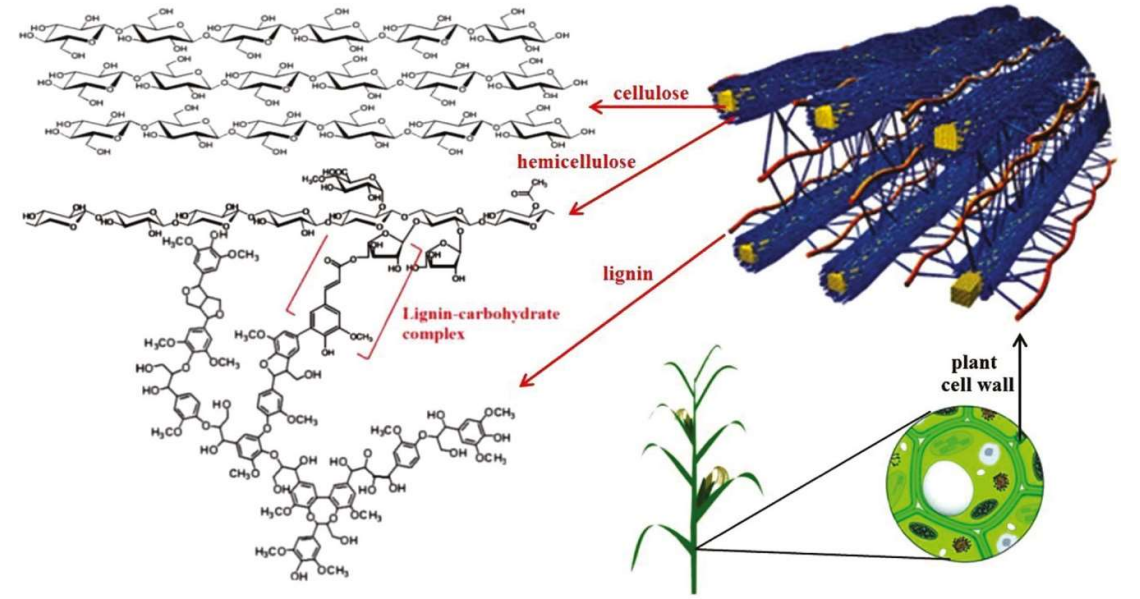


Figure 1.1 Structure of lignocellulose expanded to show the main components.[19]

Lignin is composed of randomly arranged cross-linked phenylpropanoid entities, which induce rigidity in the biomass structure through hydrogen bonding and covalent linkages. It is an amorphous polymer that serves as a linker between hemicellulose and cellulose.[10] Lignin can be used to directly produce chemicals like polyurethane, polyesters and biologically active polyphenols, as well as derived products such as low-value heating fuels, binders and dispersants.[20]

Hemicellulose is similarly amorphous,[2] it is a branched heteropolysaccharide with a β-(1,4)-glycosidic linked backbone, whose saccharide constitution varies based on source plant matter species group and climatic conditions.[10,21] While this varied saccharide composition indicates high diversity in potential processing products, it also complicates the processing of hemicellulose sources. Nevertheless, chemical and biological studies have shown the ability to produce furfural, lactones and xylitol.[22]

Conversely, the other polysaccharide component, cellulose, is a semi-crystalline homopolysaccharide, that is linear and fibrous in nature.[22,23] The general structure is comprised of a few hundred to thousands of D-glucose subunits joined into a chain by β-(1,4)-glycosidic ether bond linkages,[21] the repeating unit of which is cellobiose (Figure 1.2). The structure most commonly found in plant matter, known as cellulose I_β,[24] is one of seven polymorphs of cellulose that are known presently, defined by their intermolecular hydrogen bonding network (Figure 1.3).[3]

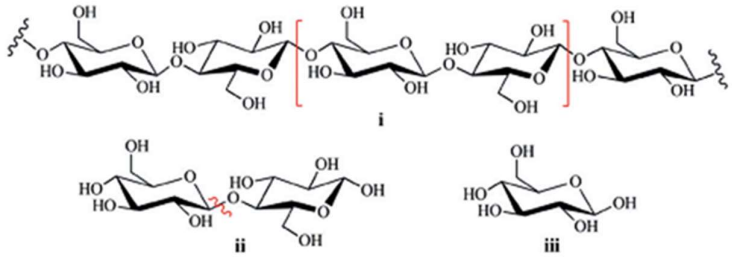
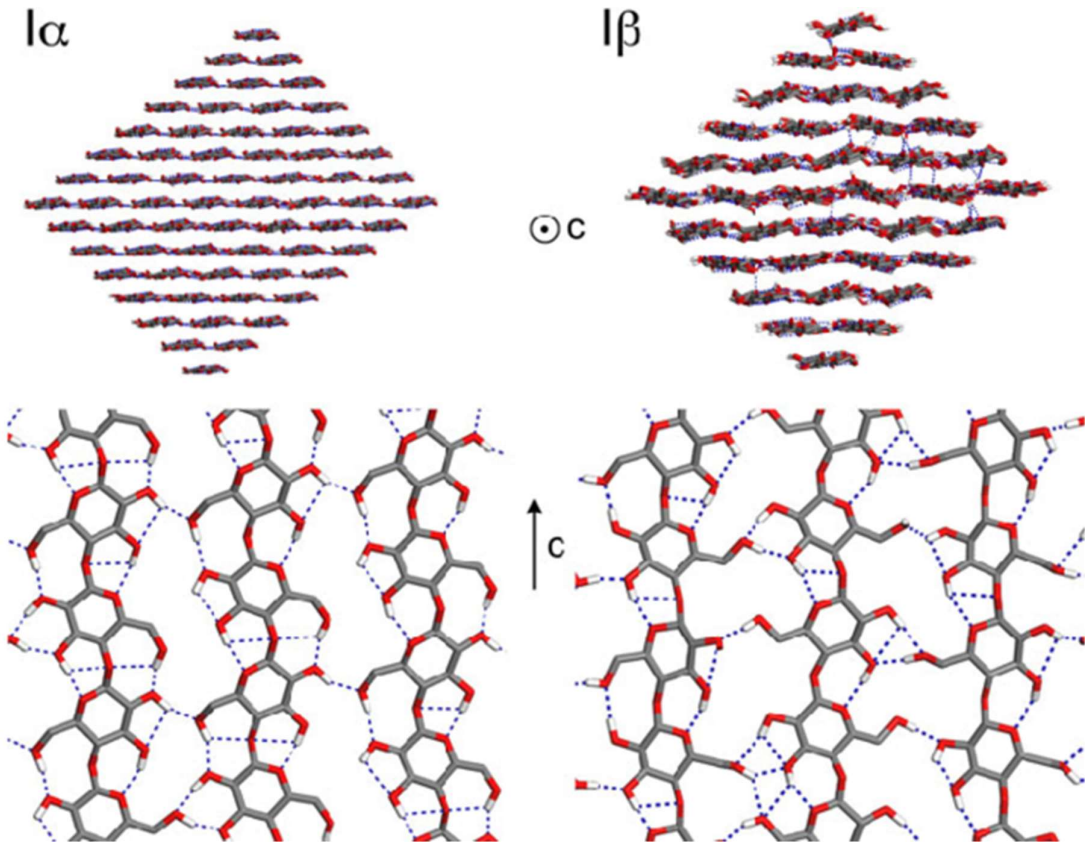
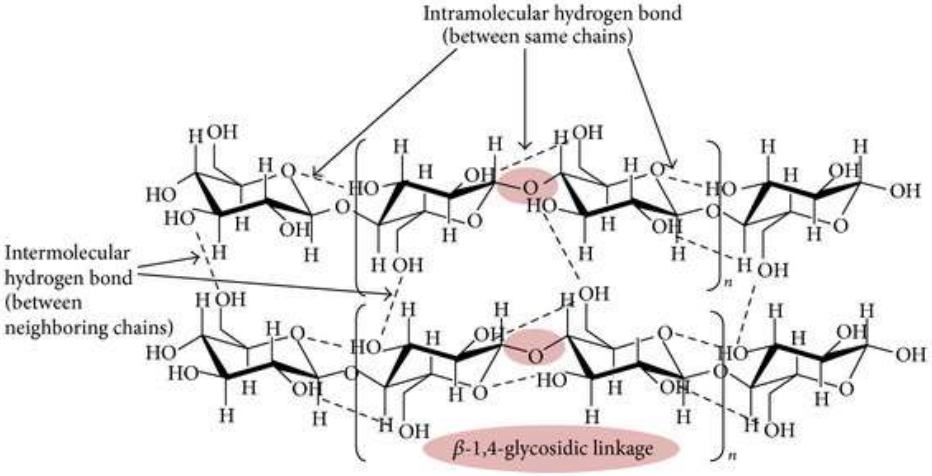


Figure 1.2 Representation of the general structure of a cellulose **(i)** chain, **(ii)** basic cellobiose repeating unit and **(iii)** D-glucose subunit.[7]

Side-by-side intermolecular hydrogen bonding between the cellulose chains allow the formation of sheets (Figure 1.3).[24] The sheets then stack on top of one another, kept together through van der Waals forces and weak intersheet hydrogen bonding.[24] Intrachain hydrogen bonding between the pyranic oxygen and the vicinal hydroxyls of the rings further adds to the extensive hydrogen bonding network (Figure 1.3),[24] confining the cellulose chains to a flat ribbon-like conformation.[3]



(i)



(ii)

Figure 1.3 Depiction of **(i)** variations in hydrogen bonding for the two cellulose polymorphs commonly found in plant matter and **(ii)** specific intra- and intermolecular hydrogen bonding networks found in I_β crystalline structure.[25,26]

Another contributing feature to the rigidity of cellulose is electronic effects, such as the anomeric effect occurring in the acetal moiety.[8,24] This is due to the orbital interactions in the O-C-O linkage of the β-(1,4)-glycosidic bond, which in cellulose manifests in two forms: the first is the endo-anomeric effect, which elongates the C-O bond in the glycosidic linkage, and the second and dominant exo-anomeric effect, which shortens the bond. The shortening of the bond has been proven to further hinder the hydrolysis.[8,27]

It is this dense hydrogen bonding network and rigid structure that makes cellulose highly resistant to chemical and biological transformations, complicating its applicability in biorefinery processes.

1.2 Ionic Liquids

One of the major problems encountered when using cellulose in a biorefinery model is its low solubility in water and many common organic solvents. Many alternate solvent systems have been tested, such as aqueous *N*-methylmorpholine-*N*-oxide (NMMO),[13] or lithium chloride/*N*,*N*-dimethylacetamide (LiCl/DMAc),[28] however, these systems all have notable disadvantages, many of which render them environmentally damaging, negating the purpose of biorefinery processing.[3,13,29,30] An alternative is presented in the use of ionic liquids, which generally efficiently dissolve cellulose without derivatisation.

1.2.1 Background

Ionic liquids are generally defined as salts in the liquid phase but can also be described by a variety of terms such as molten salts, liquid organic salts, fused salts or liquid electrolytes.[31] A further restriction defines room temperature ionic liquids as molten salts with a melting point below 100 °C.[32] They are typically comprised of a nitrogen or phosphorus containing organic cation and a large inorganic or organic anion (Figure 1.4).

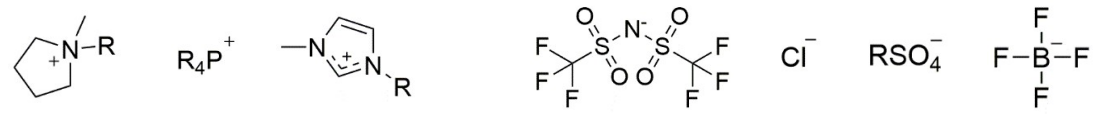


Figure 1.4 Generalised structures of commonly used ionic liquid **(i)** cations and **(ii)** anions.[3,33]

The high interest in ionic liquid solvents as replacements for molecular solvents is due in part to the compositional versatility of ILs, but largely due to the non-volatile nature of ILs under ambient conditions, which generally leads to lower accidental exposure risks due to insignificant vapour pressures and reduced ecotoxicity. This property of ILs allows for the possible design of safer, environmentally friendly solvents.[34] Furthermore, ILs also possess many other attractive physicochemical properties as well, such as low melting points, large stable liquid ranges and good thermal and electrochemical stability.[35]

In terms of cellulose dissolution, the focus is predominantly on their ability to form strong hydrogen bonds.[28] As was detailed in Section 1.2, cellulose's low solubility in most molecular solvents has been attributed to its supramolecular structure. Ionic liquids, while neutral overall, consist of two charged entities which are weakly bound and thus can act independently; allowing each component to interact with a solute structure in a manner that favours their hydrogen bonding capabilities.[31,33,36] Hydrogen bonds that can form between the cellulose chain and either the cation or anion of the ionic liquid may be significantly stronger than some of the intermolecular or intramolecular hydrogen bonds already existing within the cellulose chain. This allows for effective disruption of the hydrogen bonding network in cellulose induced by the presence of the ionic liquid solvent.

Alternative molecular solvents, such as water and methanol, have also been shown to partially disrupt cellulose's intramolecular network through formation of hydrogen bonding.[23] The compositional variability of ionic liquids, however, allows for the use of anions that are strong hydrogen-bond acceptors. These anions can then form hydrogen bonds that are energetically more favourable than those between the hydroxyls present in cellulose or hydrogen-bond acceptor functionalities present in molecular solvents.

1.2.2 Functionalised Ionic Liquids

ILs can also be functionalised to expand their applicability, allowing for possible use as catalysts in reactions for which they already are a preferred solvent. Of interest to cellulose biorefinery processes are acid functionalised ionic liquids or acidic ionic liquids (AILs), characterised as low melting ionic salts with acidic properties.[32] These properties are due to the presence of Brønsted and/or Lewis functionalities in either or both solvent ions (Figure 1.5).[32]

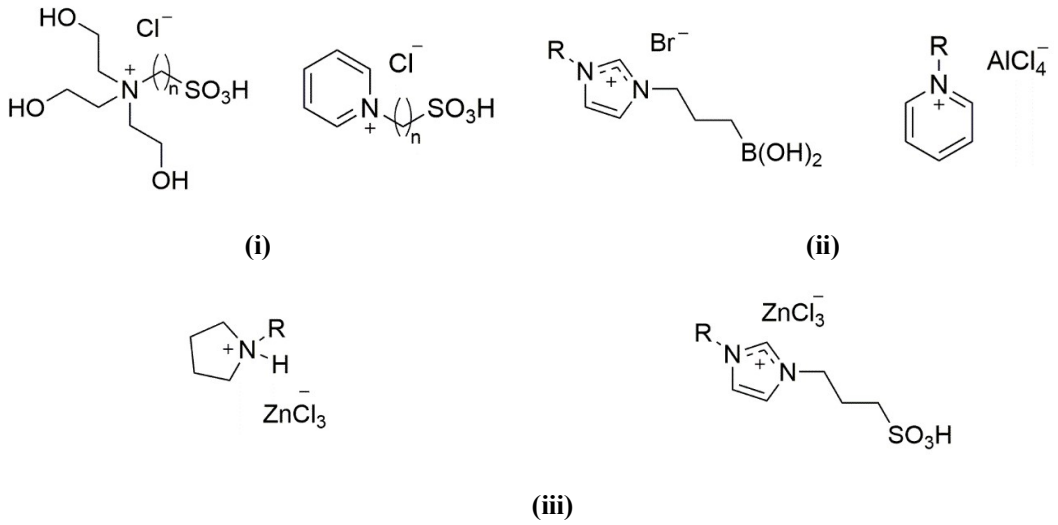
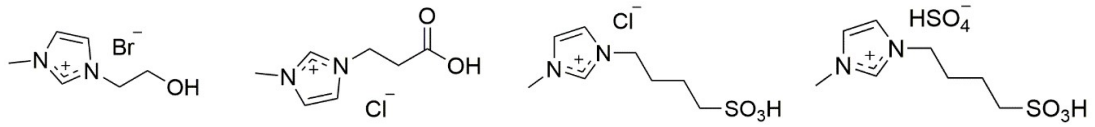


Figure 1.5 Examples of AILs with **(i)** Brønsted, **(ii)** Lewis and **(iii)** Brønsted-Lewis functionalities.[32]

Acidity, according to Lewis acid-base theory, is defined in terms of nonbonding electron-pair donation or acceptance, such that acids are electron-pair acceptors and bases are electron-pair donors and thus Lewis AILs result from an electron deficiency (Figure 1.5).[32,37] Likewise, Brønsted AILs are proton donors in accordance with the Brønsted-Lowry theory of acids and bases, which defines acidity in terms of proton donation or acceptance. In most cellulose studies, acidity is limited to a singular functionality of a Brønsted nature, localised to the cation (Figure 1.6).



[(HO)^{2}C$_2$C$_1$im][Br] [(HCO$_2$)^{2}C$_2$C$_1$im][Cl] [(HSO$_3$)^{4}C$_4$C$_1$im][Cl] [(HSO$_3$)^{4}C$_4$C$_1$im][HSO$_4$]

Figure 1.6 Examples of Brønsted acid functionalised AILs used in cellulose hydrolysis and solubility studies.[10,11,38]

Envisaged as a more efficient hydrolysis process, combining the dissolution capabilities of ILs with the hydrolysis capabilities of mineral acids, Amarasekara and Owereh investigated cellulose hydrolysis using Brønsted acid functionalised AILs in [C$_4$C$_1$im][Cl].[11] Their findings highlighted the first application of SO$_3$H functionalised ILs, under mild reaction conditions.

Evaluation of alternative acidic functionalities or cationic cores, such as COOH and triethylammonium, respectively, found that SO$_3$H functionalised cations considerably

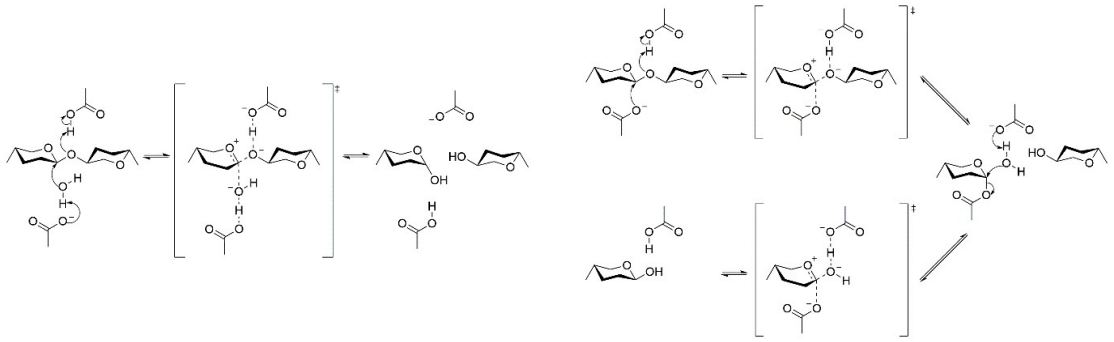
improved hydrolysis rates and total reduced sugar yields.[16,17] However, many of the studies evaluated performance of the acidic IL as a catalyst only, using a non-reactive IL as a solvent, such as [C_4C_1im][Cl], thus not fully assessing the duality of the acidic solvent.[7,17] Furthermore, there are only a few published reports on the detailed mechanistic effects of acidic IL catalyst, thus limiting further modification based on a clearly elucidated method of action.[7,10]

1.3 Cellulose Hydrolysis: Mechanism to Medium

Effective dissolution of cellulose overcomes a significant hurdle to its use in biorefinery processes by allowing access to the glycosidic linkages of cellulose chains for hydrolysis to reducing sugar units. This hydrolysis can occur by a variety of means such as enzymes, solid acid catalysts, metal salts and recently functionalised ILs; which all aim to cleave the glycosidic ether linkage between glucose subunits.[5,17,39–41] This follows a general path of protonation of a glycosidic oxygen, breaking of a C-O bond and a nucleophilic attack by water at the anomeric carbon. Mechanistic studies have determined that this path can vary in site of protonation, presence of intermediates and transition states.

1.3.1 Enzymatic Hydrolysis

Literature tends to favour the biomimetic pathways, which are modelled after the mechanism of action of glycoside hydrolase enzymes.[4] This gives rise to two specific mechanisms, the retaining or double displacement mechanism and the inverting mechanism (Scheme 1.1),[6] differing with regards to product stereochemistry.



(i) **(ii)**

Scheme 1.1 Approximate representations of glycoside hydrolase enzymatic **(i)** inverting and **(ii)** retaining mechanisms using carboxylic acid residues.[6]

1.3.1.1 Inverting

An inverting mechanism inverts the configuration at the anomeric carbon, as shown in Scheme 1.1 (i).[6] This concerted mechanism proceeds through one transition state in which six bonds are broken or formed, in unison.[7] The glycosidic oxygen is protonated by the acid moiety while the basic moiety deprotonates the water, increasing its nucleophilicity, allowing it to perform a concerted attack on the anomeric carbon. The simultaneous attack of the water and protonation of the glycosidic oxygen causes the C-O bond cleavage and formation of a glucose entity with inverted anomeric configuration.[6] In terms of reactants, this involves the use of a charged system with two catalytic entities, an acid and a base.

In relation to functionalized ionic liquids, a catalytic cation would exist in a zwitterionic form with a deprotonated basic functional group, to activate a water molecule. This is the case in a theoretical study done by Li et al., in which the mechanistic details of cellulose conversion using functionalized ionic liquids were investigated using a methyl glucose model.[7] Only biomimetic mechanisms were investigated whereby [$(HSO_3)^4C_4C_1$im][Cl] cations and anions catalysed the glycosidic cleavage. The SO_3H functionality acted as proton donor and its conjugate base, SO_3^-, acted as a proton acceptor, activating the water molecule. Furthermore, an ionic liquid solvation medium (SMD-PGP, [C_4C_1im][Cl]) was used. It was found that under these conditions the inverting mechanism is energetically more favourable than the alternative retaining mechanism.[7] However, this mechanism is not often favoured in cellulose hydrolysis studies and has not been extensively investigated, specifically in terms of ionic liquid solvation, both implicitly and explicitly.

1.3.1.2 Retaining

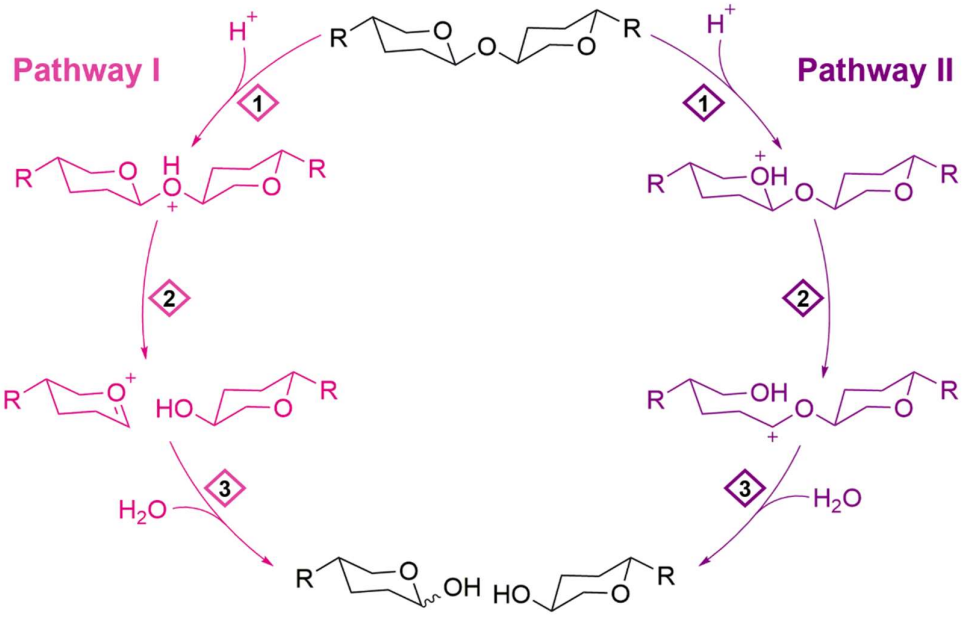
Alternatively, the anomeric carbon configuration is maintained through a double displacement or Koshland mechanism, as shown in Scheme 1.1 (ii).[42] This is a stepwise pathway, consisting of an inverted intermediary structure and two oxocarbenium-like transition states. Protonation of the glycosidic oxygen and concomitant cleavage of the glycosidic bond result in a glucose-enzyme or acid moiety-linked intermediate.[5] Exemplified in the first transition state, four bonds are simultaneously broken or formed, as the acidic moiety protonates the oxygen atom and the anomeric carbon undergoes a nucleophilic attack from the basic entity.[6] The acid catalyst reforms by abstracting a proton from water, initiating a nucleophilic substitution whereby the basic or enzymatic entity is replaced by a hydroxyl group in an S_N2 like mechanism. This

similarly proceeds through a oxocarbenium-like transition state, inverting the anomeric configuration again and thus retaining the original configuration.[43] In functionalized ionic liquids, Li et al. proposed that the anomeric carbon undergoes nucleophilic attack from the IL anion, not a basic moiety.[7]

The reaction to generate the intermediate is termed *glycosylation*, followed by *deglycosylation* where simultaneous deprotonation of water and nucleophilic substitution at the anomeric centre produce glucose.[43–45]

1.3.2 Catalytic Hydrolysis

While biomimetic mechanisms have proven experimentally viable and applicable to other systems such as ionic liquids, theoretical analyses of the mechanisms have shown the transition state's endocyclic oxygen and anomeric carbon bond length to be slightly double bond-like in character.[7,46] This links to alternative mechanisms that have been proposed for cellulose acid hydrolysis, illustrated in Scheme 1.2. These mechanisms undergo similar steps with acid protonation of an oxygen and cleavage of the glycosidic C-O linkage. However, due to the absence of a basic entity to block the inversion of stereochemistry, planar carbocation intermediates are formed, producing both sugar anomers.[3]



Scheme 1.2 Proposed glycoside acid-catalysed hydrolysis mechanistic pathways.[3]

In pathway I, a cyclic carbocation intermediate is proposed. This mechanism, where formation of the intermediate requires conformational changes in the ring, resulting in a ring-flip to a half-chair conformation (Step 2, pathway I, Scheme 1.2), is favoured in theoretical studies of cellobiose analogues in multiple reaction environments or systems.[3,23,24,43,44] However, variations in the mechanism have been noted between implicit and explicit solvation models.[8] In implicit solvation, the protonated glycoside first undergoes a conformational change, forming a conformationally altered protonated intermediate, through three possible transition states (Figure 1.7).[24] This conformational shift activates the glycosidic bond towards dissociation and allows formation of the cyclic intermediate.[24,27]

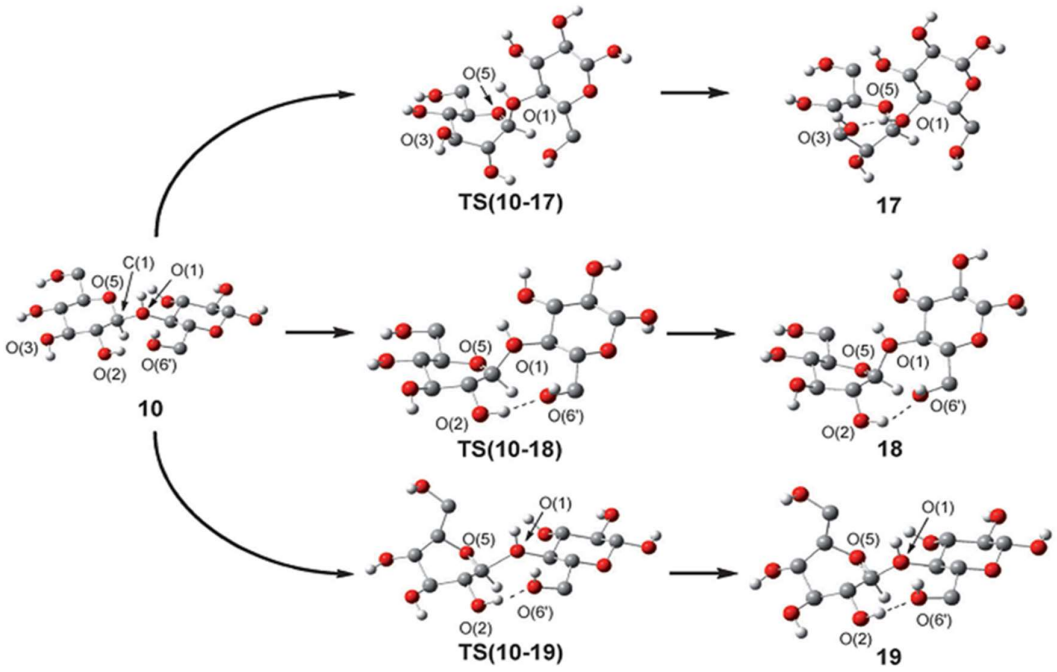
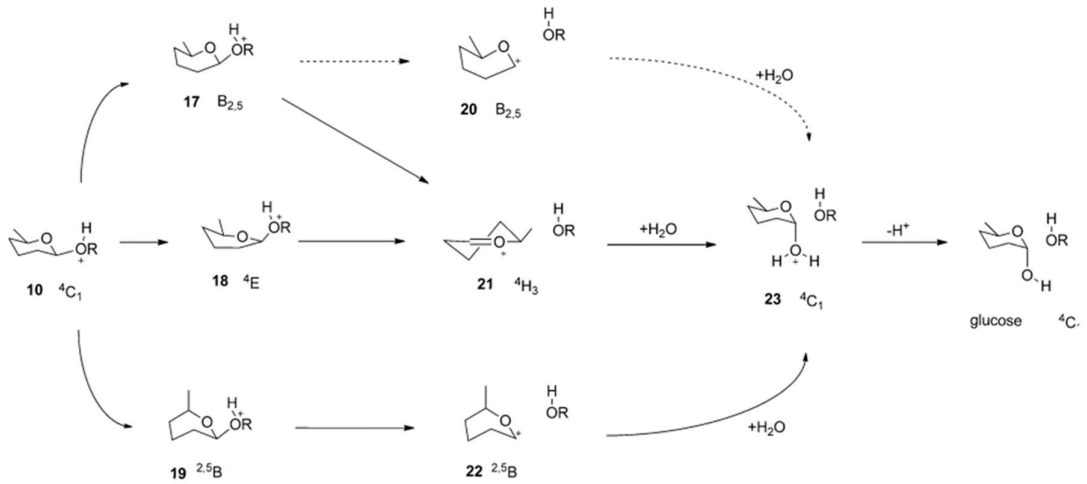


Figure 1.7 Transition states found for the conformational shift of the reactant when protonated.[24]

A stepwise path proceeding through envelope and half-chair intermediates, was found to be most energetically favourable in CPCM water by Loerbroks et al. in comparison to the other conformational alternatives investigated, shown below (Scheme 1.3).[24]



Scheme 1.3 Stepwise mechanisms proposed by Loerbroks et al. that proceed through different cyclic carbocation intermediate conformations.[24]

In a later paper by Loerbroks et al., where implicit and explicit cellulose hydrolysis in water were compared, it was found that in explicit solvation, protonated intermediates were unstable and those with a non-chair conformational change were not found at all.[8] Thus, a concerted mechanism with fewer stationary points and similar non-chair transition states as seen in implicit solvent, as well as some novel structures, is possible.[8] However, the mechanistic choice was found to depend heavily on water proximity to the anomeric carbon. Consequently, the stepwise mechanism was determined to be more likely due to the position of water in relation to the ring being further away than found in explicit solvent calculations, due to the hydrophobicity of the ring.[3]

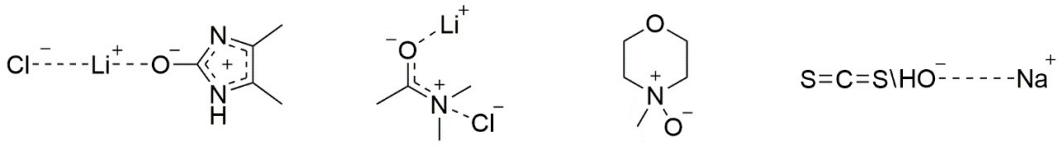
An alternative mechanism, where acid catalysed protonation occurs at the pyranic oxygen, (pathway II, Scheme 1.2), is also possible.[47] It proceeds through an acyclic intermediate and is more often associated with systems of higher ring strain and weaker leaving group abilities than cellulose,[24] and thus is not often considered for cellulose hydrolysis theoretical studies.

1.3.3 Dissolution Studies: Medium Effect on Mechanism

To utilise cellulose as a fuel and chemical feedstock, the lignocellulosic biomass must be processed into a form suitable for hydrolysis.[19] These pre-treatment processes deconstruct the complex biomass structure, separating cellulose from the other lignocellulosic components and allowing access to the sugar polymer glycosidic linkages in a manner that facilitates hydrolytic

cleavage of the β-(1,4)-glycosidic linkages, resulting in the release of the glucose subunits from the fibrils for fermentation or further chemical conversions.[26]

However, the complexity of cellulose's supramolecular structure, detailed in the preceding sections, further complicates this process by hindering effective dissolution in most typical solvents,[48] limiting chemical modification to sugar polymers located at the surface.[49] As such, many experimental and theoretical studies have been conducted into the mechanism of cellulose dissolution with a consensus being reached that effective dissolution is a result of a solvent's ability to disrupt the inter- and intramolecular bonding networks.[29,50] Accordingly, proposed mechanisms of action for common cellulose solvents, illustrated in Figure 1.8,[28] are centralised around solvent interaction with the hydroxyl groups of the biopolymer.



Dimethylimidazolone/LiCl DMAc/LiCl NMMO CS_2/NaOH

Figure 1.8 Common cellulose solvents.[28]

Solvents that chemically modify the hydroxyl groups through covalent interactions forming soluble and labile ether, ester, or acetal intermediates, are classified as derivatizing solvents.[51] Most early cellulose solvent systems, such as sodium hydroxide/carbon disulfide, used in the viscose process, are derivatizing.[28,29,52] The alternative non-derivatizing solvents, such as NMMO, do not form intermediates.[28] Instead, the solvent interacts with the hydroxyl functionality through intermolecular hydrogen bonding, cleaving the intramolecular hydrogen bonds of cellulose to form stronger hydrogen bonds with the active solvent moiety, creating a soluble complex.[29]

While the first dissolution studies of cellulose in room temperature ILs was reported in 2002, by Rogers et al.,[13] cellulose solubility in molten salt systems had been suggested before in 1934, but was not considered feasible at the time.[13,28,53] Conversely, Rogers et al.'s investigation into the cellulose non-derivatizing dissolution properties of 1-butyl-3-methylimidazolium ILs launched numerous studies of the solvation abilities and mechanism of ILs, specifically with regards to their applicability to biorefinery processes.[13,54]

While these studies have explored the compositional versatility of ILs and discovering high solubility systems with other industrially attractive properties,[30] opinions on the action and significance of each of the ionic species still vary.[49]

Many experimental and theoretical solvation studies of common IL cellulose solvents have indicated the importance of the anion.[13,30,55,56] ^{13}C and $^{35/37}Cl$ NMR relaxation measurements of $[C_4C_1im][Cl]$ have shown strong stoichiometric hydrogen bonding interactions between cellobiose equatorial hydroxyl groups and the solvent chloride anions with little noticeable effects for the cation.[55] This was extended in a QM/MD study by Xu et al. of $[C_4C_1im][Cl]$ solvation of cellulose oligomers, which assessed the role of each ionic component in the disruption of the intramolecular hydrogen bonding network.[57] It was found that the chloride anions predominantly occupied the immediate space around the cellulose segment. This was a result of their stronger interactions with the cellulose hydroxyl groups, decreasing the distance between the proton acceptor anions and proton donor hydroxyl groups.[56] However, the cations were noted to interact with the hydroxyl groups to a lesser degree as proton donors, inducing a proton acceptor/donor duality in the cellulose hydroxyl groups.[57] A few other studies have also indicated that cations, specifically methylimidazolium and methylpyridinium cores, do contribute to dissolution, but to a far lesser extent than the anions.[23,28,30,58] This has been postulated to be due to weaker hydrogen bonding with cellulose hydroxyls and significant van der Waals interactions with the top and bottom surfaces of the cellulose chain, which disrupts the stacking of the sheets in crystalline cellulose.[30,49]

Studies comparing various anions used in IL solvents, have identified promising solvents to contain strong hydrogen-bond acceptor anions, such as chloride, carboxylates and dialkyl phosphates. This finding ties in with the general consensus that cellulose dissolution is largely governed by the hydrogen bonding abilities of solvent molecules.[13,28,30]

1.4 Aims and Objectives

The principle aim of this study is to assess the effect an ionic liquid has on the mechanism and thermodynamics of cellobiose acid hydrolysis, both as a solvent and a catalyst. To this end, the study is split into two parts, each focusing on a specific effect. The first aims to measure the catalytic activity of various Brønsted acid-functionalised ionic liquids through prediction of pK_a values in ionic liquids. This is accomplished through the following objectives:

- Establishment of a computational protocol (this includes the procedure by which solvation is modelled in the thermodynamic cycle as well as the level of theory associated with each step) for the prediction of pK_a values in ionic liquids.
- Benchmarking of a series of computational protocols using structurally related compounds with known pK_a values.
- Application of the most accurate, precise and efficient protocol(s) to a series of Brønsted acidic ionic liquid catalysts.
- Comparison of pK_a trends with known experimental trends in acidity, activation energy or yield.

The second aim focuses on the solvation effect of various imidazolium-based ionic liquids on free energy requirements of the Loerbroks et al. cellobiose hydrolysis mechanism,[24] through

- Replication of the original study.
- Evaluating the effect of calculation specifics (level of theory and solvent model).
- Determination of free energy changes in ionic liquid medium for each step in the mechanism.
- Comparison of ionic liquids to aqueous free energies, generated using equivalent calculation specifics (level of theory, solvent model), to assess any structural or free energy changes.

2.

Theoretical Background

2.1 A Brief Foray into Quantum Mechanics

The fundamental postulates of quantum mechanics (QM) states that systems at the atomic scale must be described by wave mechanics. With QM as a base, the observable chemical properties of these molecular systems can be predicted theoretically using wave functions.[59] For computational chemists this all starts with the *time-independent Schrödinger equation*[60]

$$\hat{H}\Psi = E\Psi \tag{2.1}$$

where Ψ represents the molecular wave function, $\hat{H}$ is the Hamiltonian operator and E is a scalar energy value. In QM, an isolated system state is mathematically described by a wave function, which characterises the properties of the system state. Operators act on wave functions to yield physical observables. Equation 2.1 is an example of an Eigen equation making Ψ an eigenfunction and E the eigenvalue for the Hamiltonian.

The *Hamiltonian* describes nuclear and electronic interactions and for a system with N electrons and M nuclei is

$$\hat{H} = -\sum_{i}^{N}\frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_{A}^{M}\frac{\hbar^2}{2M_A}\nabla_A^2 - \sum_{i}^{N}\sum_{A}^{M}\frac{e^2 Z_A}{4\pi\varepsilon_0 r_{iA}} + \sum_{i}^{N}\sum_{j>i}^{N}\frac{e^2}{4\pi\varepsilon_0 r_{ij}} + \sum_{A}^{M}\sum_{B>A}^{M}\frac{e^2 Z_A Z_B}{4\pi\varepsilon_0 r_{AB}} \tag{2.2}$$

$$\text{with } \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{2.3}$$

Where *i,j* and *A,B* correspond to electrons and nuclei respectively, and $\hbar$ is Planck's constant divided by 2π, m_e the mass of an electron, M_A the mass of nucleus A, ∇^2 the Laplacian operator which differentiates with respect to all coordinates of the considered particle (Equation 2.3), e the charge of an electron, Z charge of the nucleus, ε_0 is the permittivity of free space, and lastly r is the distance between two particles.

The first two terms describe the kinetic energies of the electrons and nuclei separately, expressed in a wave formulation. The remaining terms describe the Coulombic attraction or repulsion of the particles. Thus, the Schrödinger equation describes the motion and interaction

of electrons and nuclei in a molecular system, but no analytical solutions exist beyond simple cases like a hydrogen atom and applications to more complicated systems require simplifications to be made to some of the individual terms present within the Hamiltonian operator.

Molecular nuclei have far greater masses than electrons and hence move much slower. Assuming this mass and subsequent speed disparity to be great enough that nuclei can be considered stationary relative to the electrons, allows for particle motions to be decoupled. Called the *Born-Oppenheimer approximation (BO)*,[61] this simplifies the system to electrons moving in the electric field of fixed nuclei. In this system the nuclear kinetic energy ($\sum_A^M \frac{\hbar^2}{2M_A}\nabla_A^2$) is neglected and the nuclear-nuclear repulsion energy ($V_{nn} = \sum_A^M \sum_{B>A}^M \frac{e^2 Z_A Z_B}{4\pi\varepsilon_0 r_{AB}}$) is constant for a specific nuclear geometry.[62] The resultant terms form the electronic Hamiltonian ($\hat{H}_{\text{electronic}}$), which when substituted into Equation 2.1 forms the *electronic Schrödinger equation* (in atomic units)[i]

$$\hat{H}_{\text{electronic}}\Psi_{\text{electronic}} = E_{\text{electronic}}\Psi_{\text{electronic}} \qquad \textbf{(2.4)}$$

$$\hat{H}_{\text{electronic}} = -\sum_i^N \frac{1}{2}\nabla_i^2 - \sum_i^N \sum_A^M \frac{Z_A}{r_{iA}} + \sum_i^N \sum_{j>i}^N \frac{1}{r_{ij}} \qquad \textbf{(2.5)}$$

In Equation 2.5, the first term is the electronic kinetic energy, the second is the electron-nuclear attraction energy and the final the electron-electron repulsion energy.

The expectation value of $H_{\text{electronic}}$ applied to the normalized $\Psi_{\text{electronic}}$ is thus the electronic energy ($E_{\text{electronic}}$)

$$\langle E_{\text{electronic}} \rangle = \frac{\int \Psi^*_{\text{electronic}} \hat{H}_{\text{electronic}} \Psi_{\text{electronic}} d\tau}{\underbrace{\int \Psi^*_{\text{electronic}} \Psi_{\text{electronic}} d\tau}_{=1 \; for\ a\ normalised\ wavefunction}} \qquad \textbf{(2.6)}$$

which combined with V_{nn} yields the total energy.

Despite this simplification, solutions are still limited to one-electron systems due to the final term in Equation 2.5 (the Coulombic pairwise interaction), the electron-electron repulsion

[i] In a system of atomic units, physical quantities are represented as multiples of fundamental constants which have been assigned a value of 1. These constants are the mass of an electron (m_e), electron charge (e), the permittivity of free space ($4\pi\varepsilon_0$), and angular momentum ($\hbar = \frac{h}{2\pi}$).[232]

energy. Approximate solutions are divided into two approaches, wave function-based and density functional theory. Wave function-based approaches expands the electronic wave function as a linear sum of *Slater determinants*, simplest of which is the Hartree-Fock (HF) method. Alternatively, density functional theory (DFT) by-passes the use of wave functions by using *electron density*. Both approaches will be detailed in Sections 2.2-2.4. It should be noted that given that all further approximate solutions are only concerned with solving the electronic Schrödinger equation, the electronic subscript is dropped from all wave functions, eigenvalues and operators.

2.2 Hartree-Fock Method

While the BO approximation has simplified the system through separation and exclusion of nuclear correlation, the remaining electron interactions still prove problematic. Electrons affect each other's motion through repulsion because of their like charges. One can assume that for stabilised systems electrons move to minimise the repulsive forces between them, hence correlating their movements.

Finding a wave function to represent this system of highly correlated electrons is extremely challenging. Wave function-based methods attempt to solve the problem of electron-correlation by employing an *independent-particle model*, where each electron acts independently of the others motion. Thus, each has its own wave function and the total wave function can thus be constructed from these one-electron wave functions. This is based on the *variational principle* which states any trial wave function ($\widetilde{\Psi}$) produces an energy (E) greater than or equal to the exact ground state energy (E_0). Minimisation of this energy with regard to all eligible N-electron wave functions will give the true ground state energy E_0

$$E_0 = \min_{\Psi \to N} E\left[\widetilde{\Psi}\right] \tag{2.7}$$

$$\text{with } E\left[\widetilde{\Psi}\right] = \langle E \rangle = \int \widetilde{\Psi}^* \hat{H} \widetilde{\Psi} d\tau \tag{2.8}$$

Given the impracticality of this search, further application of the variational principle to a subset of all viable functions gives a best approximation of the E_0, without determination of the exact wave function. Wave function-based approaches apply this through construction of trial wave functions from Slater determinants (SDs) of the one-electron wave functions, also called molecular orbitals (MOs). The use of SDs ensures the trial wave function produced is

antisymmetric, as an exchange of any two rows or columns results in a sign change. Physically this represents the interchange of any two electrons resulting in a sign change, satisfying the fermion nature of electrons and *Pauli's exclusion principle*. Thus, ensuring the trial wave functions are physically viable.

The simplest of these approaches, Hartree-Fock, approximates the trial wave function as a single SD (Φ_{HF}),

$$\Psi_{SD} = \Phi_{HF} = \frac{1}{\sqrt{N!}}\begin{vmatrix} \chi_1(x_1) & \chi_2(x_1) & \cdots & \chi_N(x_1) \\ \chi_1(x_2) & \chi_2(x_2) & \cdots & \chi_N(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(x_N) & \chi_2(x_N) & \cdots & \chi_N(x_N) \end{vmatrix} \tag{2.9}$$

where N electrons occupy N one-electron spin-orbitals ($\chi_i(x_i)$) or MOs, consisting of a spatial function ($\Psi(r)$) and an orthonormal spin function ($\alpha(s)$ or $\beta(s)$).[63] Thus, x_i is a coordinate composed of both the three spatial and one spin coordinates of the electron i.

The variability of the SD trial wave function is limited to that of its spin-orbitals. It therefore follows that the lowest energy wave function is determined through optimisation of the one-electron spin-orbitals ($\chi_i(x_i)$). This is done by minimising the Hamiltonian expectation value of Ψ_{SD} with respect to the spin-orbitals, giving the HF energy as follows[64]

$$E_{HF} = \int \Psi_{SD}^{*} \hat{H} \Psi_{SD} dx_i = \sum_{i}^{N} \left(i\middle|\hat{h}\middle|i\right) + \frac{1}{2}\sum_{i}^{N}\sum_{j}^{N} (ii|jj) - (ij|ji) \tag{2.10}$$

$$\left(i\middle|\hat{h}\middle|i\right) = \int \chi_i^{*}(x_1)\hat{h}\chi_i(x_1)dx_1 \tag{2.11}$$

$$(ii|jj) = \iint \chi_i^{*}(x_1)\chi_i(x_1)\frac{1}{r_{12}}\chi_j^{*}(x_2)\chi_j(x_2)dx_1dx_2 \tag{2.12}$$

$$(ij|ji) = \iint \chi_i^{*}(x_1)\chi_j(x_1)\frac{1}{r_{12}}\chi_j^{*}(x_2)\chi_i(x_2)dx_1dx_2 \tag{2.13}$$

Equation 2.11 defines contributions from the electron kinetic energy and the nuclear-electron attraction energy, where $\hat{h}$ is the sum of the operators dependent on one electron coordinate

$$\hat{h}(x_1) = -\frac{1}{2}\nabla^2 - \sum_{A}^{M}\frac{Z_A}{r_{1A}} \tag{2.14}$$

Equations 2.12 and 2.13 represent contributions due to electron-electron interactions. The *Coulomb integral* (2.12) represents the classical electrostatic repulsion between electron densities. The *exchange integral* (2.13) is an artefact of the antisymmetry of the wave function. It describes the QM interaction between electrons of the same spin and eliminates self-interaction introduced by 2.12 when $i = j$.

However, Equation 2.10 only holds provided the spin-orbitals remain orthonormal throughout the minimisation. To this end, Lagrangian multipliers (ϵ_i) are introduced, resulting in the *integro-differential Hartree-Fock equation* (2.15)

$$\hat{f}(x_1)\chi_i(x_1) = \epsilon_i\chi_i(x_1) \tag{2.15}$$

$$\hat{f}(x_1) = \hat{h}(x_1) + v^{HF}(x_1) \tag{2.16}$$

which identifies the best spin-orbitals to minimise the E_{HF}, where ϵ_i is the orbital energy determined using the Fock operator ($\hat{f}$). A one-electron energy operator, $\hat{f}$ consists of the one-electron Hamiltonian ($\hat{h}$) and a one-electron potential operator called the HF potential (v^{HF}). Composed of the Coulomb ($\hat{J}_j$) and exchange ($\hat{K}_j$) operators, v^{HF} simplifies explicit electron-electron interactions to interaction with the averaged potential field of all other electrons.

$$v^{HF}(x_1) = \sum_j \left(\hat{J}_j(x_1) - \hat{K}_j(x_1)\right) \tag{2.17}$$

$$\hat{J}_j(x_1)\chi_i(x_1) = \left[\int \chi_j^*(x_2)\frac{1}{r_{12}}\chi_j(x_2)dx_2\right]\chi_i(x_1) \tag{2.18}$$

$$\hat{K}_j(x_1)\chi_i(x_1) = \left[\int \chi_j^*(x_2)\frac{1}{r_{12}}\chi_i(x_2)dx_2\right]\chi_j(x_1) \tag{2.19}$$

The Coulomb operator (2.18) defines the average local Coulomb potential at position x_1 due to the average charge distribution of an electron in orbital χ_j. Summed over all non-equivalent orbitals this yields the total averaged potential at a point (x_1) generated by the N-1 other electrons. Thus, it approximates repulsion between all electrons. $\hat{K}_j$ (2.19) incorporates the exchange potential arising from correlation same spin electrons, due to the antisymmetric nature of the SD wave function.

The final solutions found would correspond to the exact HF spin-orbitals, however the equation can only be solved exactly for atoms. However, when it is necessary to describe molecular

systems, through the use of molecular orbitals (MO), the spin-orbitals, for such systems, are expressed as a linear combination of known basis functions called atomic orbitals (LCAO)

$$\chi_i = \sum_{\mu} C_{\mu}^{i} \phi_{\mu} \tag{2.20}$$

where ϕ_{μ} are generally atom-centred functions and C_{μ}^{i} are MO expansion coefficients determined using the variational method. Inserting equation 2.20 into the integro-differential Hartree-Fock equation (2.15) and converting it into a matrix form results in the *Roothaan-Hall equation*

$$\mathrm{FC} = \mathrm{SC}\varepsilon \tag{2.21}$$

where F is the Fock matrix, C the matrix of MO coefficients C_{μ}^{i}, and S the overlap matrix. Exact descriptions of the orbitals are a result of using infinite basis sets, but this is not computationally viable. Finite basis sets are thus used, introducing basis set truncation errors accuracy. Regardless, even at the limit of a complete basis set, energies produced by HF only represent approximately 99% of the exact non-relativistic BO electronic energy.[65] The difference is attributed to the insufficient approximation of electron-electron correlation. Post-HF methods are divided into two categories, semiempirical methods and electron correlation methods. Semiempirical methods make use of experimental data and fitting parameters to solve the deficiencies of HF. Electron correlation methods remain *ab initio* and build upon HF by introducing additional determinants to more accurately represent electron-electron correlation and approximate the exact solution.

2.3 Electron Correlation Methods

Electron-electron correlation consists of two indivisible parts, a static and dynamic part. Dynamic correlation is the “instantaneous” interaction between electrons in the same spatial orbital. Physically, this represents how an electron moves to avoid areas close to the instantaneous positions of all the other electrons. On the other hand, static correlation relates to interactions between electrons in different spatial orbitals. Sometimes called a near-degeneracy effect, it is significant in systems with nearly degenerate states.[65] While HF includes correlation for parallel spin electrons both dynamic and static correlation are poorly addressed, leading to an overestimation of the exact energy.

Use of a single determinant insufficiently describes molecular systems with degeneracies, and approximation of electron-electron interactions using an averaged potential field neglects movement due to instantaneous electron repulsion. Therefore, both static and dynamic electron correlation are neglected, resulting in a variation from the exact non-relativistic BO electronic energy. This difference is called the electron correlation energy (E_{corr}).

$$E_{\text{corr}} = E - E_{HF} \tag{2.22}$$

where E is the exact non-relativistic electronic energy, within the BO approximation, and E_{HF} is the HF limit. While only contributing approximately 1% to the total energy, E_{corr} is significant in the accurate prediction of chemical quantities such as binding energy.[65] Furthermore, as the number of electron pairs increase so will the correlation effects, decreasing the accuracy of HF for larger more complicated systems. Electron correlation is therefore required for the accurate description of large chemical systems.

While HF is limited in application, it provides a basis from which better approximations can be built, which attempt to account for electron correlation. Most standard post-HF methods, detailed in the following sections, construct a trial wave function from multiple determinants, including the HF single determinant. These additional determinants account for the neglected electron correlation effects.

2.3.1 Configuration Interaction

Configuration interaction (CI) methods are the simplest version of this multideterminant expansion technique.[66] In the same manner that MOs were improved using a LCAOs, CI improves the trial wave function using a linear combination of determinants

$$\Psi_{CI} = a_0\Phi_{HF} + \sum_{ia} a_i^a \Phi_i^a + \sum_{\substack{i<j \\ a<b}} a_{ij}^{ab} \Phi_{ij}^{ab} + \sum_{\substack{i<j<k \\ a<b<c}} a_{ijk}^{abc} \Phi_{ijk}^{abc} + \sum_{\substack{i<j<k<l \\ a<b<c<d}} a_{ijkl}^{abcd} \Phi_{ijkl}^{abcd} + \cdots \tag{2.23}$$

where a_0 and $a_{i\ldots}^{a\cdots}$ are expansion coefficients determined using the variational method, ensuring the resultant energy is a minimum. Φ_{HF} is taken as the reference term, representing the upper bound of the exact lowest energy molecular state. The determinants that follow the HF SD are called excited determinants.

CI works through consideration of alternative electron arrangements/configurations. The reference or HF calculation generates the MOs, both occupied and unoccupied or virtual, used

to construct the subsequent determinants. This is done by replacing occupied orbitals in the reference (Φ_{HF}) with virtual orbitals. Φ_i^a is therefore, the substitution of occupied spin-orbital *i* with virtual orbital *a*, or the excitation of one electron. This makes $\sum_{ia} a_i^a \Phi_i^a$ the sum of all possible single excitations. Denoting all occupied orbitals with *i*, *j*, *k* … and all virtual orbitals with *a*, *b*, *c* …; it follows that the subsequent terms are the sums of all possible double, triple, quadruple and so on electron excitations. As the number of virtual orbitals generated for a system N electrons and M basis functions is $M - \frac{N}{2}$, the number of determinants in 2.23 is only limited by the basis set size. At this limit, called full CI, the correlation energy is recovered yielding a numerically exact wave function for the specific basis set. Consequently, under infinite or complete basis set conditions, full CI results in the exact solution to the electronic Schrödinger equation. A feat that is however not computationally viable for anything but the smallest of molecular systems, given the factorial scaling of computational cost with system and basis set size. Applications to larger systems thus require Equation 2.23 to be limited to a specific excitation level. Truncating both the determinant and basis set expansion results in approximate methods like CISD, limited to single and double excitations only.

As with the basis set limitation, truncating of the determinant expansion induces errors. The size-consistency and size-extensivity of full CI is not maintained for the truncated methods. Thus, correlation energy no longer scales in a linear manner with number of particles in the system (size-extensivity); and the energy of infinitely separated particles is no longer equivalent to the sum of each's individual energy (size-consistency). Chemically, the loss of these properties is significant, limiting the applicability of truncated methods like CISD.

2.3.2 Perturbation Theory

One of the simplest manners in which electron correlation is addressed makes use of perturbation theory, which approximates a solution to an insolvable or unanswered problem from the exact solution(s) of a simpler related problem. When applied in QM, it is termed *Rayleigh-Schrödinger perturbation theory*, where a complex operator is split into a more applicable form composed of a solvable related operator and a series of mapped perturbations.[67] In this manner, corrections can be added to independent electron energy approximations as a perturbation, with the Hamiltonian (complex operator) split into

$$\hat{H} = \hat{H}^{(0)} + \lambda \hat{V} \qquad \textbf{(2.24)}$$

where $\hat{H}^{(0)}$ is the zeroth-order Hamiltonian, with all eigenfunctions and eigenvalues known, $\hat{V}$ the perturbation operator and λ a parameter representing the strength of the perturbation, varying from 0 to 1. Increasing the perturbation would result in continuous change in the wave function and energy. The exact wave function and energy can therefore be expressed in terms of a Taylor expansion in powers of λ

$$\Psi = \lambda^0\Psi^{(0)} + \lambda^1\Psi^{(1)} + \lambda^2\Psi^{(2)} + \lambda^3\Psi^{(3)} + \cdots \quad \textbf{(2.25)}$$

$$E = \lambda^0 E^{(0)} + \lambda^1 E^{(1)} + \lambda^2 E^{(2)} + \lambda^3 E^{(3)} + \cdots \quad \textbf{(2.26)}$$

making the exact energy and wave function an infinite sum of the corrections from each order.

Applying Equations 2.25 and 2.26 to the electronic Schrödinger equation and grouping all terms of the same order of λ, leads to the derivation of the following set of equations for calculating the energies at different orders (using Bra-Ket notation)[ii]

$$E^{(0)} = \langle\Psi^{(0)}|\hat{H}^{(0)}|\Psi^{(0)}\rangle \quad \textbf{(2.27)}$$

$$E^{(1)} = \langle\Psi^{(0)}|\hat{V}|\Psi^{(0)}\rangle \quad \textbf{(2.28)}$$

$$E^{(2)} = \langle\Psi^{(0)}|\hat{V}|\Psi^{(1)}\rangle \quad \textbf{(2.29)}$$

$$\vdots$$

$$E^{(n)} = \langle\Psi^{(0)}|\hat{V}|\Psi^{(n-1)}\rangle \quad \textbf{(2.30)}$$

When applied in the context of electron correlation as the correction, the zeroth-order operator ($\hat{H}^{(0)}$) is defined as the sum of the one-electron Fock operators (f_i)

$$\hat{H}^{(0)} = \sum_i f_i \quad \textbf{(2.31)}$$

in a method called *Møller-Plesset (MP) Theory*.[68,69] This however results in electron-electron repulsion being counted twice, necessitating the inclusion of a correction term. This correction is the perturbation operator $\hat{V}$, the difference between the true electronic Hamiltonian (2.5) and the reference (2.31)

[ii]Also called Dirac notation, a standard notation often used in Quantum Mechanics. The right part, denoted the ket, is typically a column vector. The left is thus the bra, the Hermitian conjugate of the ket vector, often a row vector. When written with an operator, it often used as a representation of matrix multiplication.

$$\hat{V} = \hat{H} - \hat{H}^{(0)} = \sum_{i=1}\sum_{j>i}\frac{1}{r_{ij}} - \sum_{i=1}\sum_{j=1}\left(J_j(i) - K_j(i)\right) \quad \textbf{(2.32)}$$

Under these conditions (2.31 and 2.32), the first-order wave function is Φ_{HF} and thus the first-order energy, the sum of the zeroth- and first-order corrections, equal to the HF energy. This dictates that electron correlation correction effects are only present from second-order onwards. However, given that second-order (MP2) still recovers about 80-90% of the correlation energy with costs just above that of HF, perturbation is generally truncated at second-order. Furthermore, it remains size-consistent and size-extensive, increasing applicability beyond that of more accurate methods like CISD. Increasing beyond MP2 however results in costs comparable to more accurate methods like CI and coupled cluster. Furthermore, MP_n methods can show slow convergence, oscillation or complete divergence of the series expansion;[70] this is attributed to the non-variational nature of the method and choice of the HF Hamiltonian as the unperturbed operator.

2.3.3 Coupled Cluster

Known to be one of the most robust methods for the prediction of atomic and molecular electronic structures, coupled cluster maintains size-consistent and size-extensive despite truncation. Truncation errors are rectified through the use of an exponential form of the full CI wave function

$$\Psi_{CC} = e^{T}\Phi_{HF} \quad \textbf{(2.33)}$$

$$\text{with } e^{T} = 1 + \frac{T}{1!} + \frac{T^2}{2!} + \frac{T^3}{3!} + \cdots = \sum_{k=0}^{\infty}\frac{T^k}{k!} \quad \textbf{(2.34)}$$

Where e^{T} is a Taylor expansion of T, a series of excitation operators (T_i)

$$T = T_1 + T_2 + T_3 + \cdots \quad \textbf{(2.35)}$$

which when applied directly to the reference wave function generates all the excited SDs in Equation 2.23.

$$(1 + T_i)\Phi_{HF} = \Psi_{CI} \quad \textbf{(2.36)}$$

$$T_1\Phi_{HF} = \sum_{ia} t_i^a \Phi_i^a \quad \textbf{(2.37)}$$

$$T_2\Phi_{HF} = \sum_{\substack{i<j \\ a<b}} t_{ij}^{ab}\Phi_{ij}^{ab}$$

$t_{i\cdots}^{a\cdots}$ are amplitude terms determined variationally such that Equation 2.33 holds true. However, coupled cluster does not treat excitations in the additive manner like CI. Excitations are treated in a multiplicative manner, making the coupled cluster version of Equation 2.36

$$\left[\prod_i (1 + T_i)\right]\Phi_{HF} = \Psi_{CC} \qquad \textbf{(2.38)}$$

Where can be simplified to e^T. It is this Taylor series expansion of excitation operators that ensures size-consistency is maintained in coupled cluster through the generation of excitation operator products or disconnected states. For example, for a CISD method, the coupled cluster exponential term is

$$e^{T_1+T_2} = 1 + T_1 + \left(T_2 + \frac{T_1^2}{2}\right) + \left(T_2T_1 + \frac{T_1^3}{6}\right) + \left(\frac{T_2^2}{2} + \frac{T_2T_1^2}{2} + \frac{T_1^4}{24}\right) + \cdots \qquad \textbf{(2.39)}$$

CISD would be only the first three terms of 2.39, the reference, singly (T_1) and doubly (T_2) excited states. The additional terms are the products of excitation operators or disconnected excited states. Disconnected terms correspond to the movement of non-interacting electrons or electron pairs. For example, both T_2 and $\frac{T_1^2}{2}$ represent double excitations, the first representing the movement of two interacting electrons (an electron pair) and the second the excitation of two non-interacting electrons. These product excitations facilitate inclusion of higher order excitations than dictated by the connected (T_i) operator, as in 2.39 where triple, quadruple and so on excitations are present for a method limited to a double excitation operator. It is this property that allows CC to remain size-consistent regardless of truncation, expanding beyond the operator limit. This is beneficial because like CI, CC scales factorially with system size, so truncation is once again necessary for practical application. Once again, improvement from HF is only notable from double excitations (CCSD) but impractical beyond that point due to scaling. Alternative methods have been used to expand to triple excitations using perturbative approximations as in the CCSD(T) method, increasing accuracy while maintaining applicability in terms of computational cost. Considered the "gold standard" for QM methods, CCSD(T) using a large basis set can predict chemical properties with near experimental accuracy.

2.3.3.1 Domain-based Local Pair Natural Orbital Coupled Cluster Method

As stated above coupled cluster remains one of the most accurate *ab initio* methods used in the study and evaluation of chemical systems however its associated extreme costs limit its applicability in terms of system size and complexity. The system sizes that coupled cluster calculations are limited to often fall outside of the sphere of chemical interest, making the technique redundant in most computational chemistry research. However, given the possible accuracy that can be achieved by incorporating this level of electron correlation many avenues of improving these computational costs while minimally impacting accuracy have been explored.

In terms of hardware, parallel implementations on supercomputers have been investigated and shown impressive results. However, the high cost of the required hardware (and often also software) leads to limited use by the broader computational chemistry community.[71]

Another avenue looks at modifications to the method itself to reduce calculation complexity. One of the most successful of these is based on the short-range nature of the electron correlation covered in CC methods. These approaches apply approximations of locality to the electron correlation, allowing for the use of localised molecular orbitals limiting the space of interest to a concentrated region of the molecule or atom. One such method of interest in this study is the DLPNO or domain based local pair natural orbital approach. Applied to CCSD the method shows near linear-scaling with regard to system size in terms of computational costs, like time and disk and memory requirements, while maintaining an accuracy of about 4 kJ mol^{-1}.[72]

Developed by Neese et al. the method is based on the combined use of pair natural orbitals (PNOs) with projected atomic orbitals (PAOs).[71] This is based on the popularity of pair natural orbital use in local correlation methods, which assigns a separate set of natural orbitals to each electron pair. When applied to an electron pair, PNOs are localised to the same region of space as the electron pair. Thus, PNOs can broadly represent the locality of actual bonding environments of electron pairs under study while efficiently compacting the space needed to recover correlation. Incorporating these PNOs with approximations based on locality lead to the formation of accurate and efficient correlation treatments like LPNO-CCSD.[73] However, these correlation treatments (LPNO-CCSD etc) are still limited to systems of 100 atoms or less, due to the use of canonical virtual orbitals in the expansion of the PNOs. To improve upon this, expansion is done instead using a set of PAOs that span the external space. A comprehensive theoretical background of the general local pair natural orbital and domain based local pair

natural orbital approaches can be found in the base literature by Neese et al.[71,73,74] Within the context of this study only the general steps involved in the application of this process to CCSD(T) calculations will be addressed.

In accordance with the CCSD(T) version of this approach, localised orbitals are created from the occupied orbitals of a single determinant wave function, usually HF of nature, using the Foster-Boys localisation method.[75] The correlation energy is then determined by summing over all the electron pair correlation energies of the localised orbitals. These energies decrease with distance between the orbital centres which allows for weaker pairs with negligible contributions to be excluded. This exclusion gives rise to a linear scaling in the number of remaining electrons pairs.

Exclusions are performed based on whether pair correlation energies, approximated using a multipole estimate followed by a semicanonical local MP2 (SC-LMP2) calculation, are above a certain threshold ($T_{CutPairs}$).[74,76] The SC-LMP2 densities of the remaining pairs are used to construct the PNOs. These are once again trimmed based on an occupation number threshold (T_{CutPNO}).[73,77] The final resulting PNOs with occupation numbers larger than the cut off are expanded using a set of PAOs of a domain whose initial size is determined according to a final truncation parameter (T_{CutMKN}).[77] This cut off is based on atom Mulliken population size of the localised MO of interest. This overall allows the wave function to be compacted while incurring minimal costs due to the introduction of the electron pair specific sets of correlating orbitals.

2.4 Composite Methods

As has been illustrated throughout the preceding sections, the accuracy of *ab initio* methods is dependent upon the level of correlation included and the basis set size, with an exact solution the result of taking both to their limit. However, there is an inverse relationship between the accuracy of these calculational methods and the number of electrons or complexity of the studied system.[78,79] Unfortunately, time and computational costs scale drastically as each increase, forcing both to be truncated. Truncation introduces significant errors, decreasing accuracy. Thus, only small systems can be tested at a high level of accuracy or correlation and basis size must be limited according to system size (Figure 2.1).

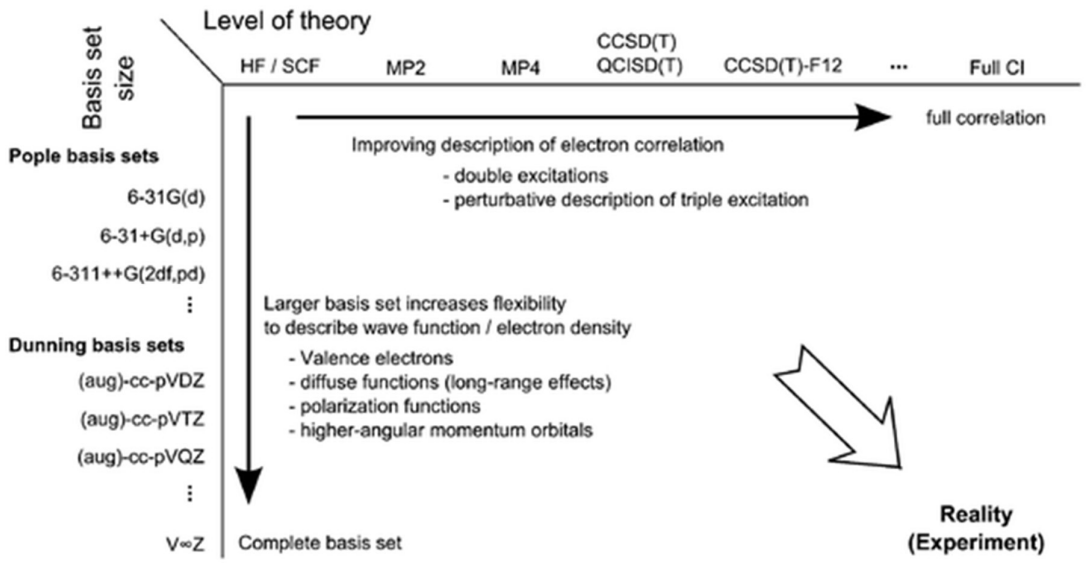


Figure 2.1 Variation of a Pople diagram illustrating the relationship between the level of theory/basis set combination and accuracy for *ab initio* methods.[79,80]

By assuming that different properties converge at different rates, and that basis set extension effects are additive, various procedures have been proposed to approximate the correlation energy and correct for a finite basis set size. A high-level calculation can be split into many lower cost and thus lower level calculation steps which can be optimised according to the required accuracy. For realistic and chemical significance, accuracy is generally defined as ~ 4 kJ mol^{-1}, termed "chemical accuracy".[81] Thus, *high-accuracy multilevel methods* or *QM composite methods* were developed, defining a series of steps which cumulatively resulted in a "chemically accurate" final energy, approximating the infinite correlation/basis set energy. The two most commonly used are *Complete Basis Set* (CBS) and *Gaussian-n* (G*n*) methods. Both methods follow similar steps of optimisation and frequency calculations at low levels followed by a series of single-point calculations at increasing correlation level and/or basis set with appropriate corrections for any inherent approximations. They differ however in how they extrapolate electron correlation energy.

2.4.1 Gaussian-n Theories

Originally developed by the Pople group,[82] the Gaussian-*n* (G*n*) theories have three main objectives. The first is to theoretically determine experimentally reproducible energies to a prescribed accuracy. The second is to be broadly applicable in terms of molecular system nature and size. And the third, in accordance with the main objective of all composite methods, is to

maintain computational efficiency regardless of system complexity or calculation level. This is done by approximating the energies of expensive calculations by combining a high-level correlation method (CCSD(T) or CISD(T)) at a moderate basis set (6-31G(d)) with a series of energies calculated at lower levels (MP2-MP4) with larger basis sets.

There are many iterations of the method, called families of methods (G1-G4). Each build on the previous correcting for past errors, widening applicability and increasing accuracy while maintaining or decreasing computational cost. First developed to achieve an accuracy of ~ 8.4 kJ mol^{-1} for G1,[83] constant improvements have led to accuracies of ~ 3.5 - 5.8 kJ mol^{-1} for currently used methods (G4 – G3(MP2)).[82] This accuracy is determined in terms of a mean absolute deviation (MAD) from experiment, with experimental references collected in a test set. The test sets contain series of experimental enthalpies of formation, ionisation potentials, electron and proton affinities. Values in the set are chosen according to the main G*n* objectives. Thus, test set values are for molecular species containing various elements and bonding environments to ensure wide applicability, determined experimentally to ensure no theoretical bias and at a specific experimental accuracy. The first formal set, G2, contained a total of 125 energies for small molecules at an experimental accuracy limit of ~ 4 kJ mol^{-1} or better.[84,85] As the families of methods were developed, additional species were added to the test set. These additions aim to correct applicability deficiencies in the methods through inclusion of larger more complex species containing heavier atoms and various bond types. The latest test set, G3/05, consists of 454 energies, adding energies from nonhydrogen species, molecules containing third row elements and hydrogen-bonded complexes to the existing test.[86]

Later iterations like G3 and G4 have mostly rendered G1 and G2 series obsolete (Table 2.1) in terms of accuracy, applicability and cost.

Table 2.1 Mean absolute deviations (MAD, in kJ mol^{-1}) from experiment for various base G*n* methods according to the G2/97 test set.[82]

Energy (kJ mol^{-1})	**G*n* Method**		
	G2	*G3*	*G4*
Ionisation energies	5.90	4.69	3.89
Electron affinities	5.90	4.06	4.35
Proton affinities	4.52	5.61	4.23
Enthalpies of formation	6.57	6.57	2.93
Overall	6.19	4.14	3.35

Regardless of iteration, the same general procedure is followed, which consists of the following steps (representation adapted from references 81 and 87),

1) A geometry optimisation and frequency calculation at a low level of theory (HF/6-31G(d))
2) A geometry optimisation at a MP2 (higher correlation level) and same basis set as previous calculations (6-31G(d)), to be used as the reference geometry for the following calculations
3) A single-point calculation at target correlation level (QCISD(T)) with the same basis set used in steps 1-3
4) A MP4 basis set expansion consisting of
    a) A single-point calculation at MP4 with a diffuse expansion of the basis set (6-31+G(d))
    b) A single-point calculation at MP4 with a higher polarisation version of the basis set (6-31G(2df,p))
5) A correction for the larger basis set effects and assumption that basis extensions could be treated separately, consisting of
    a) A single-point calculation at MP2(FU) with a targeted large basis set, defined by G*n* method (G3Large)
    b) Single point of diffuse extension at MP2
    c) Single point of polarisation extension at MP2
6) Addition of atomic spin-orbital corrections and higher level corrections (HLC), both of which can sometimes have empirical origins.[82]

Through combination of these steps the energy at high correlation level with a large basis set (QCISD(T,Full)/G3Large) can be determined for a MP2 optimised geometry. Represented mathematically, using the correlation levels and basis set examples for G3 given in steps 1-6,[88]

$$E[\mathrm{G}n] = E[\mathrm{QCISD(T)/6\text{-}31G(d)}] + E[\mathrm{plus}] + E[\mathrm{2df,p}] + E[\Delta\mathrm{G3L}] + E[\mathrm{SO}] + E[\mathrm{HLC}] + E[\mathrm{ZPE}] \quad \textbf{(2.40)}$$

$$E[\mathrm{plus}] = E[\mathrm{MP4/6\text{-}31{+}G(d)}] - E[\mathrm{MP4/6\text{-}31G(d)}] \quad \textbf{(2.41)}$$

$$E[\mathrm{2df,p}] = E[\mathrm{MP4/6\text{-}31G(2df,p)}] - E[\mathrm{MP4/6\text{-}31G(d)}] \quad \textbf{(2.42)}$$

$$E[\Delta\mathrm{G3L}] = E[\mathrm{MP2(Full)/G3Large}] - E[\mathrm{MP2/6\text{-}31{+}G(d)}] - E[\mathrm{MP2/6\text{-}31G(2df,p)}] + E[\mathrm{MP2/6\text{-}31G(d)}] \quad \textbf{(2.43)}$$

$$E[\mathrm{HLC}] = \begin{cases} -6.386n_\beta - 2.977(n_\alpha - n_\beta) \cdots for\ molecules \\ -6.219n_\beta - 1.185(n_\alpha - n_\beta) \cdots for\ atoms \end{cases} \quad \textbf{(2.44)}$$

where $E[\mathrm{QCISD(T)/6\text{-}31G(d)}]$ is the energy calculated at QCISD(T)/6-31G(d) using the MP2/6-31G(d) optimised geometry (Step 3). The other terms, expanded in Equations 2.41 – 2.44, correspond to the steps listed above as follows. Both $E[\mathrm{plus}]$ and $E[\mathrm{2df,p}]$ are energies associated with basis set expansions for diffusion (Step 4 (a)) and polarisation (Step 4 (b)) respectively. $E[\Delta\mathrm{G3L}]$, corresponding to Step 5, corrects for large basis set effects and the non-additivity associated with treating the basis set expansions separately. It contains the overall

calculation targeted large basis set, G3Large, which expands on the G2 target large basis set of 6-311+ G(3df,2p). It is done at a MP2(FU) level to include core-related contributions. The final terms relate to the spin-orbital ($E[\mathrm{SO}]$) and higher level ($E[\mathrm{HLC}]$) empirical corrections applied, calculated in relation to the species of interest (Step 6). Spin-orbital corrections are only applied in the case of atomic species. Lastly $E[\mathrm{ZPE}]$ is the scaled zero-point energy, calculated at the specified low level of theory (Step 1).

Families are the results of each iteration (G2, G3, G4) being altered to improve efficiency with minimal accuracy loss. G*n* theory with reduced order perturbation theory methods or G*n*(MP2) alter the basis set expansion to reduce computational cost, either performing the basis set expansions and corrections at MP2 level only or contracting the $E[\mathrm{plus}]$, $E[2\mathrm{df,p}]$ and $E[\Delta\mathrm{G3L}]$ calculations to the difference between the large basis set and the low level basis set. This generally results in a reduction in accuracy of about 1.31 kJ mol^{-1} (G2-G4 average of G2/97 test set evaluations).[89–91] Furthermore, taking advantage of the cost efficiency and accuracy of DFT methods, steps 1 and 2 can be performed using B3LYP with negligible loss in accuracy, in methods denoted by G*n*B3. These modifications can be combined, using reduced perturbation to improve efficiency and DFT to reduce accuracy loss. Table 2.2 below compares these modifications for G3 and G4, detailing the key procedure changes and the resultant accuracy.

Table 2.2 Procedural specifications and mean absolute deviations (MAD, in kJ mol^{-1}) from experiment for specific variations of G3 and G4 methods according to the G2/97 and G3/99[a] test sets.[86,91–94]

G*n* Method	***(1)*** Zero-point energy [scaled by]	***(2)*** Geometry optimisation	***(3)*** Highest correlation energy	***(4) & (5)*** Basis set expansion	**MAD (kJ mol^{-1})**
G3	HF/6-31G(d) [0.8929]	MP2(FU)/ 6-31G(d)	QCISD(T)/ 6-31G(d)	6-31+G(d) 6-31G(2df,p) G3Large	**4.14 (4.44)**
G3(MP2)	HF/6-31G(d) [0.8929]	MP2(FU)/ 6-31G(d)	QCISD(T)/ 6-31G(d)	G3MP2Large	**5.36 (5.48)**
G3B3	B3LYP/6-31G(d) [0.96]	B3LYP/6-31G(d)	QCISD(T)/ 6-31G(d)	6-31+G(d) 6-31G(2df,p) G3Large	**4.06**
G3B3(MP2)	B3LYP/6-31G(d) [0.96]	B3LYP/6-31G(d)	QCISD(T)/ 6-31G(d)	G3MP2Large	**5.19**
G4	B3LYP/ 6-31G(2df,p) [0.9854]	B3LYP/ 6-31G(2df,p)	CCSD(T)/ 6-31G(d)	6-31+G(d) 6-31G(2df,p) G3LargeXP	**(3.35)**
G4(MP2)	B3LYP/ 6-31G(2df,p) [0.9854]	B3LYP/ 6-31G(2df,p)	CCSD(T)/ 6-31G(d)	G3MP2LargeXP	**(4.31)**

[a]Shown in brackets

2.4.2 Complete Basis Set Methods

Complete basis set (CBS) methods vary from G*n* methods in that they make no assumption of basis set additivity. Instead they aim to explicitly extrapolate the basis set to an infinite limit or the complete basis set limit (mentioned in Section 2.2). This is done through extrapolation of the N^{-1} asymptotic convergence of MP2 pair energies calculated using pair natural orbital expansions (of size N) to the CBS limit.[95,96]

As a multilevel or composite method, like G*n* methods, CBS methods are procedures which make use of a series of calculation steps to build up the desired correlation/basis set limit energy. These steps are as follows (adapted from references 81 and 87):

1) A geometry optimisation and frequency calculation at a low level of theory (HF/3-21G(d) or HF/6-31G(d†)), yielding an unscaled ZPE.[iii]
2) A geometry reoptimisation at a higher correlation level, MP2/6-31G(d†), establishing the reference geometry for the following calculations.
3) Estimation of finite basis set error through
    a) A MP2 single-point calculation at a basis set larger than step 2 (6-31+G(d†) or 6-311+G(3d2f,2df,2p))
    b) Pair natural orbital extrapolation of this MP2 result to the CBS limit.
4) Estimation of higher-order electron correlation level effects, consisting of one or both of the following steps depending on CBS method
    a) A MPx order expansion consisting of a
        i. A single-point calculation at MP4(SDQ) using a 6-31G, 6-31G(d†) or 6-31+G(d,p) basis set
        ii. A single-point calculation at MP2 using the same basis set
    b) A QCI expansion consisting of a
        i. A single-point calculation at MP4(SDQ) using a 6-31G or 6-31+G(d†) basis set
        ii. A single-point calculation at QCISD(T) using the same basis set
5) Corrections for remaining electron correlation effects and possible spin contamination, which are calculated empirically.

The resulting total energy is calculated as follows (using the CBS-Q method as an example),[97]

$$E[\text{CBS-x}] = E[\text{MP2}] + E[\text{CBS}] + E[\text{MP4}] + E[\text{QCI}] + E[emp] + E[spin] + E[\text{ZPE}] \quad \textbf{(2.45)}$$

$$E[\text{MP4}] = E[\text{MP4(SDQ)/6-31+G(d,p)}] - E[\text{MP2/6-31+G(d,p)}] \quad \textbf{(2.46)}$$

$$E[\text{QCI}] = E\left[\text{QCISD(T)/6-31+G(d}^{\dagger}\text{)}\right] - E\left[\text{MP4(SDQ)/6-31+G}\left(\text{d}^{\dagger}\right)\right] \quad \textbf{(2.47)}$$

[iii] For d† 6-311G(d,p) polarisation exponents are used for 6-31G sp functions to give a modified version of 6-31G(d).

$$E[emp] = -0.00533 \sum_i \left(\sum_\mu C_{\mu_{ii}} \right)^2 |S|_{ii}^2 \quad \textbf{(2.48)}$$

$$E[spin] = -0.0092\Delta\langle S^2 \rangle \quad \textbf{(2.49)}$$

where $E[\mathrm{MP2}]$ is the energy calculated at MP2/6-311+G(3d2f,2df,2p) using B3LYP/6-31G(d†) optimised geometry (Step 3), which is extrapolated to the complete basis set limit accounted for in $E[\mathrm{CBS}]$. Finally, like for the Gn methods, $E[\mathrm{ZPE}]$ is the scaled zero-point energy, calculated at the specified HF or B3LYP level of theory (Step 1). The other terms, expanded in Equations 2.46 – 2.49, correspond to the estimation and correction steps listed above. Both $E[\mathrm{MP4}]$ and $E[\mathrm{QCI}]$ are energies associated with correlation level expansions at the Møller-Plesset (Step 4 (a)) and CI (Step 4 (b)) levels respectively. $E[emp]$ accounts for any other correlation corrections needed using a size extensive empirical expression, which involves the trace of the first-order wave function coefficients for the natural orbital pair specified ($C_{\mu_{ii}}$), the absolute spatial overlap between spin components of the specified molecular orbital ($|S|_{ii}$), and a fitting constant determined using a reference data set.[98] $E[spin]$ on the other hand, is an empirical correction for spin contamination, added to select CBS methods where the UHF method is used for open-shell species.[81]

Variation of these steps gives rise to the several core iterations, CBS-4, CBS-Q and CBS-APNO; arranged here in order of increasing accuracy and computational cost. A summary of these variations and their effects on accuracy is shown below (Table 2.3).

Table 2.3 Procedural specifications and mean absolute deviations (MAD, in kJ mol^{-1}) from experiment for various CBS methods according to the G2 and G2/97[a] test sets.[81,97,99,100]

CBS Method	***(1)*** Zero-point energy [scaled by]	***(2)*** Geometry optimisation	***(3)*** PNO extrapolation	***(4)*** Correlation level expansion	**MAD (kJ mol^{-1})**
CBS-4	UHF/3-21G(d) [0.91671]	UHF/3-21G(d)	MP2/6-31+G(d†)	MP4(SDQ)/6-31G	**8.3**
CBS-Q	UHF/6-31G(d†) [0.91844]	MP2(FC)/ 6-31G(d†)	UMP2/ 6-311+G(3d2f,2df,2p)	(a) MP4(SDQ)/ 6-31+G(d(f),p) (b) QCISD(T)/ 6-31+G(d†)	**4.1**
CBS-QB3	B3LYP/ 6-311G(2d,d,p) [0.99]	B3LYP/ 6-311G(2d,d,p)	UMP2/ 6-311+G(3d2f,2df,2p)	(a) MP4(SDQ)/ 6-31+G(d(f),p) (b) CCSD(T)/ 6-31+G(d†)	**3.6 (4.6)**
CBS/APNO	UHF/ 6-311G(d,p) [0.9251]	QCISD/ 6-311G(d,p)	UMP2/6-311+G(6s6p3d2f,4s2p1d)	(a) n/a (b) QCISD(T)/6-311+G(2df,p)	**2.2**

[a]Shown in brackets

The 4th order extrapolation method, CBS-4, limits the correlation expansion to MP4 only, excluding Step 4 (b). This makes it the least computationally expensive, increasing applicability in terms of systems size. Generally, the minimum population localisation expansion, CBS-4M, is more commonly used, given its applicability to systems of ~ 20 heavy atoms.[99] However, the efficiency does come at an accuracy cost, with both methods giving mean absolute deviations (MAD) from experiment (according to the G2 test set) of 8.3 kJ mol^{-1} and 9.0 kJ mol^{-1} respectively.[97,99]

At the other end of the scale CBS-APNO, where APNO stands for asymptotic pair natural orbitals, is extremely accurate, yielding a G2 MAD of 2.2 kJ mol^{-1}.[97,101] This accuracy is achieved by using higher correlation levels and larger basis sets for the majority of the steps listed above. The geometry is optimised at QCISD/6-311G(d,p) (step 2) and the correlation correction limited to only a QCISD(T) single point (step 4 (b) only) at a larger basis set of 6-311+G(2df,p). This however does impose severe limits on system of interest size and complexity, limiting the number of heavy atoms possible to ~ 5.

A happy medium is found in quadratic CI CBS (CBS-Q),[97] which generally follows the procedure described in the aforementioned steps exactly. This expands the correlation correction to include a QCISD(T) calculation while also maintaining lower level geometry, ZPE and basis limit extrapolation steps. The MAD is greatly improved compared to CBS-4 methods (4.1 kJ mol^{-1}) while maintaining a greater system size applicability than CBS-APNO.[100] This performance is further improved through the incorporation of DFT. DFT functionals like B3LYP have a good accuracy to computational cost ratio, decreasing electron correlation errors while remaining widely applicable, making their use in the optimisation and frequency steps of composite methods very attractive. To this end the CBS-Q method was modified to use B3LYP/6-311G(2d,d,p) for steps 1 and 2, as it gave more accurate geometries than those found previously using MP2. Furthermore, the QCI expansion from MP4 was changed to a CCSD(T) expansion, as it was shown to perform better than QCISD(T) when combined with B3LYP geometries and frequencies. This expanded applicable system size to ~ 10 heavy atoms while also lowering the G2 MAD to 3.6 kJ mol^{-1}.[100] This generally results in CBS-QB3 being the most widely used CBS method, performing on par with G3 methods in terms of accuracy and speed. It still has a greater limit on system size than the equivalent G*n* methods, resulting in a general preference for G*n* methods overall given their extensive method families and consistently good accuracy to cost ratio.

2.5 Density Functional Theory

As is clearly illustrated above, the implementation of electron correlation and exchange are both key to improving calculation or model accuracy but also detrimental to computational costs and thus system applicability. An alternative to these electron correlation methods is presented in the form density functional theory (DFT), which has been shown to replicate the accuracy of high correlation level methods like MP2 while maintaining computational costs similar to HF.

The key to this technique is the abandonment of the many-electron wave function in favour of a function of electron density, a function itself dependent on the three spatial coordinates. This function of a function, or functional, reduces the number of variables to 3 from the 3N needed for wave function-based methods and allow the implicit incorporation many-body correlation and exchange effects. This increases the accuracy with which the method describes many body systems like molecular systems, while lowering computational costs.

The method maintains the same QM background as the previously detailed *ab initio* methods, with the fundamental difference in the descriptor of the system. Thus, the independent-particle and BO approximations remain but implementation requires the derivation of the Hamiltonian in terms of electron density ($\rho(r)$). This involves the derivation of the kinetic and potential energy components according to electron density. The *Thomas-Fermi equations*, shown below, were the first iterations of this, done in terms of a classical system.[102]

$$V_{ne}[\rho(r)] = -\int \frac{Z_A \rho(r)}{|r - r_A|} dr \qquad \textbf{(2.50)}$$

$$J_{ee}[\rho(r)] = \frac{1}{2}\int\int \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1 dr_2 \qquad \textbf{(2.51)}$$

$$T_{TF}[\rho(r)] = \frac{3}{10}(3\pi^2)^{2/3}\int \rho^{5/3}(r) dr \qquad \textbf{(2.52)}$$

The first two equations correspond to the nuclear attractive (V_{ne}) and electron-electron repulsion (J_{ee}) potential energies. The final equation corresponds to the kinetic energy contributions (T_{TF}) in relation to a continuous charge distribution. This necessitated the establishment of a fictious substance called the uniform electron gas, originally jellium. It is characterised by uniformly distributed positive charge and composed of infinite electrons moving in an infinite volume.

However, given their classical basis these models proved highly inaccurate when applied to molecular systems, with the exchange and correlation effects completely ignored, necessitating the development of a more rigorous foundation and method. This led to the *Hohenberg-Kohn theorems* and Kohn-Sham DFT.

2.5.1 The Hohenberg-Kohn Theorems

To show that the energy is dependent on the density, Hohenberg and Kohn, had to show that the Hamiltonian could be uniquely determined by electron density. They did this by employing two theorems. The first, called the *existence theorem*, defines the external potential (V_{ext}) as a functional of the ground-state electron density ($\rho(r)$).[103] Taken with the definition of total number of electrons (N) as the integration of the electron density over all space,

$$N = \int \rho(r) dr \tag{2.53}$$

this fully defines the Hamiltonian, dependent only on the nuclear positions and atomic numbers (encapsulated in the external potential) and total number of electrons. By fixing the Hamiltonian, the ground state energy can also be assumed a unique functional of electron density. This is proven by showing that all observations to the contrary are impossible, called proof by *reductio ad absurdum*, the details of which are touched on in more comprehensive texts like Cramer's Essentials of Computational Chemistry.[104]

The second theorem applies the variational principle to ensure the density considered truly describes the ground state system, and hence is called the *variational theorem*. This theorem states that a well-behaved trial ground-state density ($\rho_t(r)$), that is one which obeys Equation 2.53, will minimise energies such that the following relationship occurs,

$$E_t[\rho_t(r)] \geq E_0[\rho_0(r)] \tag{2.54}$$

This relationship states that said trial density produce expectation electronic energies that are upper bounds to the true ground-state energy, until the trial density is equal to the true ground-state density ($\rho_0(r)$).[103] At this point the trial expectation energy is equal to the real ground-state energy of the system.

These theorems thus surmise the existence of an energy functional dependent on density ($F_{HK}[\rho(r)]$) that can accurately predict the exact ground-state energy but not how to determine it. This is solved in the *Kohn-Sham methodology*, which puts these theorems into practice.

2.5.2 The Kohn-Sham Method

Fundamental to this methodology is the assumed existence of a reference system of non-interacting electrons. This is to account for problems in defining the kinetic energy operator and electron exchange-correlation effects.[105] This reference system has a ground-state electron density equal to that of a real system, with energy functional defined as follows,

$$E[\rho(r)] = T_{ni}[\rho(r)] + V_{ne}[\rho(r)] + J_{ee}[\rho(r)] + E_{XC}[\rho(r)] \quad \textbf{(2.55)}$$

$$V_{ne}[\rho(r)] = \int \rho_0 V_{ext} dr \quad \textbf{(2.56)}$$

$$E_{XC}[\rho(r)] = \Delta T[\rho(r)] + \Delta V_{ee}[\rho(r)] \quad \textbf{(2.57)}$$

T_{ni} is the kinetic energy operator for a non-interacting or uncorrelated system. The introduction of the non-interacting system helps to solve this issue in terms of kinetic energy, allowing it to be treated as a sum of one-electron functions or orbitals (χ_i).

$$T_{ni}[\rho(r)] = -\frac{1}{2}\sum_{i}^{N}\langle\chi_i|\nabla^2|\chi_i\rangle \quad \textbf{(2.58)}$$

The true ground state density can then be defined in terms of $\chi_i$as,

$$\rho_0(r) = \sum_{i}^{N}|\chi_i|^2 \quad \textbf{(2.59)}$$

The remaining operator terms (V_{ne} and J_{ee}) correspond to potential energy contributions. V_{ne} and J_{ee} are the classic nucleus-electron attraction and electron-electron repulsion operators. Equation 2.56 shows classic Coulomb attraction potential energy operator (Equation 2.50) rewritten in terms of electron density and the external potential (V_{ext}) or potential due to interactions between electrons and nuclei at a distance r ($\frac{Z_A}{r}$).

The last term, called the exchange-correlation functional, is the sum of the corrections needed to account for deviations of the reference from the real system, caused by the lack of electron-electron interactions in the reference system.

$$\Delta T[\rho(r)] = T_{real}[\rho(r)] - T_{ni}[\rho(r)] \quad \textbf{(2.60)}$$

$$\Delta V_{ee}[\rho(r)] = V_{ee}[\rho(r)] - J_{ee}[\rho(r)] \quad \textbf{(2.61)}$$

$\Delta T[\rho(r)]$ is the deviation of T_{ni} from the real kinetic energy ($T_{real}[\rho(r)]$), while $\Delta V_{ee}[\rho(r)]$ corrects for the presence of non-Coulombic electron-electron interactions like exchange and Coulomb correlation and self-interactions.

Substituting Equation 2.59 into Equation 2.55 allows it to be rewritten in terms of these KS molecular orbitals. Through application of the variational principle the orbitals that minimise this energy expression can be determined. This is provided that the orbitals are orthonormal and obey the following equations,[106]

$$\underbrace{\left[-\frac{1}{2}\nabla_i^2 - \sum_A^{nuclei}\frac{Z_A}{|r_i - r_A|} + \int \frac{\rho(r_2)}{r_{i2}}dr_2 + V_{XC}(r_i)\right]}_{\hat{h}^{KS}}\chi_i(r_i) = \epsilon_i^{KS}\chi_i(r_i) \tag{2.62}$$

$$V_{XC}(r_i) = \frac{\partial E_{XC}[\rho(r_i)]}{\partial \rho(r_i)} \tag{2.63}$$

These are the KS equations, which can be solved iteratively to give both the electronic energy and corresponding molecular orbitals. $\hat{h}^{KS}$ is defined as the KS one-electron operator. It contains the derivative of the exchange-correlation functional according to density or exchange-correlation potential (V_{XC}). The exact form of this term is not known and is thus approximated using known functions of density. And thus, the exact equations of DFT are solved in an approximate manner, with the accuracy dependent on how well E_{XC} is approximated.

2.5.3 Exchange-Correlation Functionals

For approximation of the exchange correlation energy, $E_{XC}[\rho(\mathrm{r})]$ is usually split into its individual correlation and exchange contributions as shown below allowing each to be designed individually.

$$E_{XC}[\rho(\mathrm{r})] = \int \rho(\mathrm{r})\varepsilon_X[\rho(\mathrm{r})]\mathrm{dr} + \int \rho(r)\varepsilon_C[\rho(\mathrm{r})]\mathrm{dr} \tag{2.64}$$

Where each contribution is defined in terms of an energy density (ε).[107]

These functionals are arranged into a hierarchy based on sophistication according to a classification method proposed by Perdew et al.[108,109] The hierarchy is conceptualised as a ladder of five rungs, with each proceeding rung building upon the previous through new or modified approximations. The lowest rung, simplest approximation, is the local density

approximation (LDA), which assumes that the local density changes slowly or is a uniform electron gas.[107] The exchange energy is then be determined as,

$$E_X^{LDA} = -C_X \int \rho^{4/3}(\mathrm{r})\,\mathrm{dr} \tag{2.65}$$

Which is expressed in terms of individual densities when the spin densities are unequal, called the local spin density approximation (LSDA). Correlation cannot be expressed explicitly in the same manner is instead determined analytically using various methods depending on the density-limit or neglected completely. Overall these functionals have been found inappropriate for many chemical problems and thus have been displaced by more accurate approximations like the generalized gradient approximation (GGA).[107] GGA functionals improve on LDA by also incorporating the first derivative ($\nabla\rho(r)$) of the density as a variable, making the functional depend on density and its gradient (Equation 2.66).

$$E_{XC}^{GGA} = \int \rho(r)\varepsilon_{XC}^{GGA}\big(\rho(r), \nabla\rho(r)\big)\,dr \tag{2.66}$$

This allows for better description of inhomogeneous electron density systems, which is more in accordance with real molecular systems. These functionals can be expanded and improved through addition of the second derivative ($\nabla^2\rho(r)$) to give meta-GGA, the third rung of functionals.

$$E_{XC}^{m-GGA} = \int \rho(r)\varepsilon_{XC}^{m-GGA}\big(\rho(r), \nabla\rho(r), \nabla^2\rho(r)\big)dr \tag{2.67}$$

This is generally altered to be dependent on the kinetic energy density ($\tau(r)$) instead, which is the kinetic energy summed over all occupied orbitals, as it effectively contains the same information as the second derivative but is more numerically stable.[107]

$$E_{XC}^{m-GGA} = \int \rho(r)\varepsilon_{XC}^{m-GGA}\big(\rho(r), \nabla\rho(r), \tau(r)\big)\,dr \tag{2.68}$$

$$\tau(r) = \frac{1}{2}\sum_{i}^{occ} |\nabla^2\chi_i(r)|^2 \tag{2.69}$$

An example of this functional class would be the Tao-Perdew-Staroverov-Scuseria (TPSS) exchange-correlation functional.[110]

The next rung of the ladder, hybrid functionals, calculate portions of the exchange energy using HF theory with KS molecular orbitals. This is due to the ability of the HF equations to offer

the true exchange energy if the KS orbitals were equivalent to that of the real wave function. Given that the exchange is the far larger contribution this would greatly improve the accuracy of the method. Furthermore, adding the HF 'exact energy' is a reduction in the self-interaction error present in all DFT methods which results from expressing the Coulomb interaction as a double integration over independent electron densities. The energy is thus calculated as a weighted sum of HF exchange energy and DFT exchange-correlation energy as shown below,

$$E_{XC}^{Hybrid} = (1 - c_X)E_X^{DFT} + c_X E_X^{HF} + E_C^{DFT} \quad \textbf{(2.70)}$$

The B3LYP functional is a prominent example which combines GGA (B88[111], LYP[112]) and LSDA (VWN[113]) exchange and correlation functionals with HF exchange using various scaling factors.[114,115]

For the last rung, hybrids are improved through the full usage of KS orbitals, including both occupied and virtual. Called double hybrid methods, these functionals include real' electron correlation contributions through usage of post-HF, like MP2, method equations with KS orbitals. However, these methods have very high computational costs, associated with use of perturbation expressions, and thus are limited in application.

2.6 Continuum Solvation Models

It has long since been established, both thermodynamically and experimentally, that the environment or surroundings effect the state of a system. An example of this would be the effect of solvation.

A solvated system, called a solution, consists of the solute (the molecule or molecules of interest) surrounded by numerous solvent molecules. The solvent molecules can be divided into shells with the effects diminishing the further they are from the solute. The closest shell, where molecules directly interacting with the solute, is called the first solvation shell. The effects of this solvent shell on the solute are often larger and more specific than the outer solvent shells that encapsulate it. Each layer of solvent molecules shields the solute, thus reducing the effects of additional solvent shells on the solute. The effects of the solvent can thus be divided into two types, non-specific or long-range effects and specific or short-range effects.[116]

Specific effects are related to the interaction of the first solvent shell with the solute and include intermolecular interactions, structural effects, charge transfer effects and solvent-solute

dynamics. These effects are dependent on the nature of the specific solvent molecules, affected by molecular structure and bonding abilities. Likewise, non-specific effects are related to the effect of the bulk solvent and include polarisation and dipole orientation. Computationally, the way to model the solvent such that all of these effects would be accounted for would be to represent the solvent as box of multiple individual solvent molecules which is then statistically sampled. Called explicit solvation, these techniques are extremely accurate, provided all interactions are realistically described, but prove to have a high computational cost, leading to the development of simpler more cost-effective techniques called implicit models. Under implicit modelling conditions the solvent is reduced to a homogenous continuum defined primarily by a dielectric constant. This takes into account non-specific effects while negating the need to sample a many-body explicit phase space.

Implicit models are far more compatible with high accuracy QM techniques, like those mentioned above, which already incur high computational costs at large system sizes if any accuracy is to be expected. To this end a brief background is given below, followed by descriptions of two model types relevant to this study.

2.6.1 Background

As was stated above, the continuum or implicit solvent model simplifies the modelling of solvation by capturing the solvent effect using continuum and not multiple explicit solvent shells. The solvent is reduced to a uniform, continuous polarisable field. Vectors along this field represent the gradient of the electrostatic potential.[117] As an initial step, a specifically shaped cavity is created in the field into which the solute is placed. The field reacts to the charge distribution of the solute, leading to the formation of a so-called reaction field that needs to be solved self-consistently. The creation of the cavity costs energy as the solvent medium is destabilised. This, along with contributions from solvent-solute interactions, forms part of the overall solvation free energy,

$$\Delta G_{\mathrm{solv}} = \Delta G_{\mathrm{cav}} + \Delta G_{\mathrm{vdw}} + \Delta G_{\mathrm{elec}} \qquad \textbf{(2.71)}$$

ΔG_{cav} is the cavity formation energy or destabilisation cost of creating the solute cavity in the dielectric medium, accounting for the reorganisation of the solvent molecules around the cavity. ΔG_{vdw} is the energy attributed to van der Waals interactions between the solute and the solvent. The final term is the electrostatic stabilisation or interaction of the solute charge distribution with the solvent medium. This is due to the polarisation of the medium by the

solute charges and its subsequent polarisation due to the polarised medium. How these effects are modelled and defined can vary, leading to a variety of *reaction field models* which in turn define various types and families of implicit models.[116]

The dominant free energy contribution for implicit models is due to electrostatics, given that the solvent medium is defined according to its dielectric constant and how it interacts with a solute charge distribution. This can be described using the Poisson equation, which describes the response of a dielectric medium to the embedding of charge. For a position-independent charge distribution or density, the equation is as follows,

$$\nabla^2\phi(r) = -\frac{4\pi\rho(r)}{\varepsilon} \tag{2.72}$$

where ε is the dielectric constant of the solvent continuum, $\rho(r)$ the charge density of the solute, shown here as an electron density, and $\phi(r)$ the electrostatic potential. However, for a continuum solvent model Equation 2.72 is written as

$$\nabla\varepsilon(r)\cdot\nabla\phi(r) = -4\pi\rho(r) \tag{2.73}$$

Due to the presence of the solute cavity in the continuum creating two regions, within and outside the cavity, to which Equation 2.72 can be applied.

These equations are only applicable for neutral solvents. For ionic solvents the equation becomes,

$$\nabla\varepsilon(r)\cdot\nabla\phi(r) - \varepsilon(r)\lambda(r)\kappa^2\frac{k_BT}{q}\sinh\left[\frac{q\phi(r)}{k_BT}\right] = -4\pi\rho(r) \tag{2.74}$$

$$\kappa^2 = \frac{8\pi q^2 I}{\varepsilon k_B T} \tag{2.75}$$

Called the *Poisson-Boltzmann (PB) equation*, this considers how ions will accumulate based on the potential. q is the charge magnitude of the solvent ions and λ is a switching function that is zero in solvent inaccessible regions and one everywhere else. The κ^2 term is the Debye-Hückel parameter, which measures the extent to which electrostatic effects reach in the solution, with I corresponding to the solution's ionic strength.

How a cavity is created can vary based on cavity size and shape. The simplest models make use of spherical and ellipsoid shapes, called ideal cavities, as they allow the solute-medium electrostatic interactions (Equations 2.73 and 2.74) to be calculated in an analytical manner.

However, cavity shape is important, with more realistic shapes leading to better accuracy when compared to experimental data. Thus, other models like van der Waals surfaces and solvent accessible surfaces (SAS),[118,119] employ more realistic shapes based on the solute's molecular structure and are more widely utilised.

A van der Waals surface creates a surface using interlocking spheres localised on each nucleus with sphere size decided by a factor of atomic and van der Waals radii. However, these surfaces can allow for the creation of gaps in the surface where solvent molecules can enter. The alternative SAS approach which traces the surface of the solute using a spherical particle of a given radius often based on water. This method provides marginal improvements in terms of accuracy but greatly increases the computational costs, leading to van der Waals surfaces being more commonly used. Cavities can also be sourced directly from the wave function. For these surfaces solving the Poisson/PB equation is more complicated. The surface is divided into i tesserae of a fixed area s_i, over which solute charge density is spread evenly. The electrostatic potential is then calculated numerically for each of these areas in an iterative manner and all non-zero areas summed to solve the PB equation.

The actual energy needed to create this cavity or hole in the dielectric medium along with the energy introduced by solute-solvent dispersion interactions is generally calculated in relation to the surface area. This can be done by either taking the energy as directly proportional to the total SAS, using a single proportionality constant, or using empirical solvation data to apply fitting parameters specific to each atom.

2.6.2 Conductor-like Polarizable Continuum Model

The conductor-like polarization continuum model (C-PCM) is an iteration of the conductor-like screening model (COSMO) that uses the framework of the polarizable continuum or PCM solvent model.[120,121] As such it is a conductor-like apparent surface charge method. Within these models the cavity is molecular in shape, created using as a van der Waals surface with additional non-atom centred spheres added to form a solvent-excluding surface. Dispersion and cavity formation contributions are parameterised according to surface area according to expressions derived by Pierotti, and Floris and Tomasi respectively.[122–125]

According to the C-PCM and COSMO documentation the solvation free energy is defined as,

$$\Delta G_{\mathrm{solv}} = \Delta G_{\mathrm{cav}} + \Delta G_{\mathrm{dis}} + \Delta G_{\mathrm{elec}} + \Delta G_{\mathrm{rep}} \qquad \textbf{(2.76)}$$

with ΔG_{elec} defined as,

$$\Delta G_{\text{elec}} = \langle \Psi | \hat{H}_{solute} | \Psi \rangle + \frac{1}{2} \sum_{i}^{\text{tesserae}} q_i V_i \quad \textbf{(2.77)}$$

where V_i is the electrostatic potential at tessera i. This evolves as product of treating the solvation as a perturbation of the Hamiltonian of the isolated solute by operator $\hat{V}$,[126]

$$\hat{H} = \hat{H}_{solute} + \hat{V} \quad \textbf{(2.78)}$$

Operator $\hat{V}$ encapsulating all electrostatic solute-solvent interactions. Polarisation charges, for each tessera are then determined according to the conductor-like boundary condition. This condition ensures the total electrostatic potential cancels out on the cavity surface by treating the surrounding dielectric medium as a conductor, through assumption of an infinite dielectric constant. Defining $\hat{V}$ in terms of apparent polarisation charges (q_i) present in each tessera (i), the electrostatic potential (V) is then determined as,[126]

$$V(\vec{r}) + \sum_{i}^{tesserae} V_{q_i}(\vec{r}) = 0 \quad \textbf{(2.79)}$$

where V is due to the solute and V_{q_i} the polarisation charges. Using the matrix A with elements[120,126]

$$A_{ii} = 1.07 \sqrt{\frac{4\pi}{S_i}} \quad \textbf{(2.80)}$$

$$A_{ij} = \frac{1}{|\vec{r}_i - \vec{r}_j|} \quad \textbf{(2.81)}$$

where S_i is the tessera area, the vector (V) of the polarisation charges (Q), which includes electrostatic potential due to solute tesserae, can be defines as[126]

$$\mathrm{AQ} = -\mathrm{V} \quad \textbf{(2.82)}$$

If the solvent dielectric is considered then Q is scaled as follows to represent the real charges (q),[121]

$$q = \frac{\varepsilon - 1}{\varepsilon} \mathrm{Q} \quad \textbf{(2.83)}$$

2.6.3 Solvation Model based on Density

Based on the SM*x* models of Cramer and Truhlar, the solvent model based on density (SMD) is a universal solvent model,[127] which varies from other self-consistent reaction field models by using solute charge density or electron density instead of partial atomic charges. Electrostatics are treated using a generalised Born type approach, while dispersion terms are parameterised based on the solvent exposed surface area.[127,128]

Its universality is due to its ability to be applied to any solute-solvent system if the key solvent descriptors for the model are available. These key descriptors are as follows,

- Dielectric constant
- Refractive index
- Bulk surface tension
- Acidity and basicity parameters – defined as Kamlet-Taft hydrogen bond donor and acceptor parameters

The SMD model uses these descriptors to split the solvation free energy into two main contributions based on electrostatics or cavity creation, solvent structure and solute-solvent dispersion interactions,[127]

$$\Delta G_{\text{solv}} = \Delta G_{\text{ENP}} + \Delta G_{\text{CDS}} \tag{2.84}$$

The first term encapsulates electrostatic contributions, accounting for all electronic, nuclear and polarization interactions. In situations where the geometry remains the same through the phase transition, this term reduces to represent electronic polarisation contributions only (ΔG_{EP}). This is determined through a molecular self-consistent reaction field calculation of the Poisson equation (Equation 2.73). The boundary or cavity shape is a SAS made up of interlocking nuclear-centred spheres with radii, called intrinsic Coulomb radii, dependent on each atom's atomic number.[127] As such the Poisson equation is solved as a sum over tesserae of the surface.

The ΔG_{CDS} term accounts for free energies associated with cavity creation and the resultant changes in dispersion energy and local solvent molecule structure. It is intended to describe all specific solute-solvent interactions and is calculated according to the following equation,[127]

$$\Delta G_{\text{CDS}} = \sum_{k}^{\text{atoms}} \sigma_k A_k\left(\text{R}, \{R_{Z_k} + r_{\text{S}}\}\right) + \sigma^{[\text{M}]} \sum_{k}^{\text{atoms}} A_k\left(\text{R}, \{R_{Z_k} + r_{\text{S}}\}\right) \tag{2.85}$$

$$\sigma^{[\mathrm{M}]} = \tilde{\sigma}^{[\gamma]} \frac{\gamma}{1\ \mathrm{cal\ mol^{-1}\AA^2}} + \tilde{\sigma}^{[\phi^2]}\phi^2 + \tilde{\sigma}^{[\psi^2]}\psi^2 + \tilde{\sigma}^{[\beta^2]}\beta^2 \quad \textbf{(2.86)}$$

Where A_k is the solvent accessible surface area (SASA) of the atom *k*, which is reliant on the solute geometry R which is set $\{R_{Z_k}\}$ of all atomic van der Waals radii plus the solvent radius (r_S). $\sigma^{[\mathrm{M}]}$ is the molecular surface tension, calculated as a function of solvent macroscopic surface tension (γ), carbon aromaticity (ϕ), electronegative halogenicity (ψ) and Abraham hydrogen bond basicity (β) all multiplied by empirical parameters. Carbon aromaticity and electronegative halogenicity are the fraction of solvent molecule atoms that are aromatic carbons atoms or halogen atoms respectively. σ_k is the atomic surface tension of atom k, calculated as shown below,[127]

$$\sigma_k = \tilde{\sigma}_{Z_k} + \sum_{z}^{\text{atoms}} \tilde{\sigma}_{Z_k Z_z} T_k(\{Z_k, R_{kz}\}) \quad \textbf{(2.87)}$$

T_k is a switching function that is dependent on geometry, and $\tilde{\sigma}_{Z_k}$ and $\tilde{\sigma}_{Z_k Z_z}$ are parameters that depend on the atomic number of a specific atom or atoms. These parameters are defined by a set of solvent descriptors (Equation 2.88), making both surface tensions dependent on the solvent.[18,127]

$$\tilde{\sigma}_{Z_k}\ or\ \tilde{\sigma}_{Z_k Z_z} = \tilde{\sigma}^{[n]} n + \tilde{\sigma}^{[\alpha]}\alpha + \tilde{\sigma}^{[\beta]}\beta \quad \textbf{(2.88)}$$

These solvent descriptors are the room temperature refractive index (n) and Abraham hydrogen bond acidity (α) and basicity parameters, which all multiplied by various specific empirical parameters ($\tilde{\sigma}^{[n]}$, $\tilde{\sigma}^{[\alpha]}$, $\tilde{\sigma}^{[\beta]}$).

3.

Computing Ionic Liquid Acidity using Thermodynamic Cycles

This chapter details investigations of acidity calculations in and of ionic liquids, specifically of Brønsted functionalised ionic liquids, which can be used as cellulose hydrolysis catalysts in another ionic liquid solvent. This includes a brief background on the use of various thermodynamic cycles to calculate a theoretical pK_a, establishment of a preferred cycle and calculation of the relative acidity of acid functionalised ionic liquids in comparison to a known highly effective molecular acid catalyst.

Initial testing indicated a proton transfer thermodynamic cycle is significantly more accurate (compared to experimental values) and precise (consistency across levels of theory) than a direct cycle, whereby the accuracy and precision of pK_a prediction is highly dependent on the gas phase level of theory. Benchmarking indicated that DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP performance fell within the range of commonly used high level methods and thus was acceptable for use in larger systems. Application to the acidic ionic liquid systems of interest produced high pK_as relative to the molecular acids with the sulfonyl IL far more acidic than the other functionalities. This matched relative acidity trends seen experimentally, correlating well with available Hammett acidity functions. Thus, the sulfonic acid functionalised ionic liquid was shown to have the highest relative acidity in an ionic liquid solvent of the cationic acids tested.

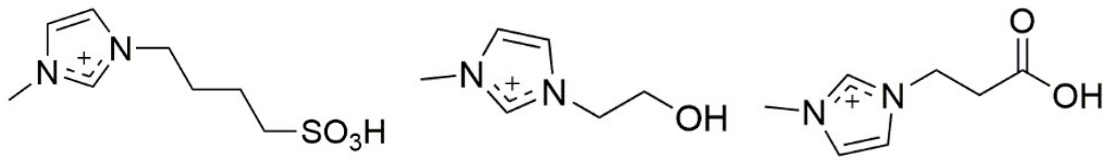
3.1 Background

The chemical transformation of cellulose is a serious challenge to its use in biorefinery processes,[30] specifically due to the structural protections present that necessitate harsh conditions.[8,50] These problems have been addressed from both a solvation and catalytic point of view, focusing on low-cost acid hydrolysis techniques over costlier enzymatic routes. Past experimental studies of cellobiose hydrolysis in traditional solvents, such as water, have indicated a stepwise mechanism, with the protonation of the glycosidic linkage oxygen atom

as a prerequisite step.[8,24] However, due to the rigid crystalline structure of cellulose and protective nature of the ring hydroxyl positioning, this step is the most energy intensive. Thus, the acidity of the catalyst or rather its free energy of protonation in solution can be hypothesised to correlate to the efficacy of cellulose hydrolysis or glucose production,[16] given the high energy requirement for the protonation of the glycosidic linkage.[15,24]

This step can be performed by solid catalysts, avoiding issues of erosion and by-product formation, but issues occur in later processing in terms of separation and effective access to reaction sites inside the cellulose crystal.[38] Dissolution, discussed further in the following chapter, breaks up the supramolecular structure of the cellulose crystals. This increases exposure of the glycosidic linkages for hydrolysis to take place. Using traditional strong mineral acids, the linkage can be protonated but necessitates the use of harsh reaction conditions making the process less "environmentally" advantageous.[38] Furthermore, separation would remain an issue. A one-pot synthesis process would be more attractive overall for industrial application due to reduced synthetic steps and processing.[10] Given the preferred use of ionic liquids in cellulose solvation and their structural versatility, the viability of acidic ionic liquids as catalysts has been investigated experimentally.[11,16,17,38,39,129–131] These experiments only indicate efficacy in terms of glucose or total reducing sugar (TRS) production but give no direct indication of the acidity of these acids, which is of significance in terms of effective application and development of the biorefinery process.

Thus, to investigate the viability of an ionic liquid based one-pot synthetic processing of cellulose, the acidity of three Brønsted-functionalised imidazolium based ionic liquid cations was assessed (Figure 3.1). Cations were chosen based on experimental cellulose hydrolysis findings to allow for comparison to total reducing sugar yields.[38] However, given the lack of experimental data on the acidity in ionic liquids and acidity of functionalised ionic liquids in general,[132–134] a protocol needed to be established, which would allow for the accurate computation of acidity in non-traditional solvents.



(i) **(ii)** **(iii)**

Figure 3.1 Structures of the acid-functionalised cations of the IL catalysts of interest **(i)** 3-methyl-1-(4-sulfobutyl)-1-imidazol-3-ium ($[IL\text{-}SO_3H]^+$), **(ii)** 1-(2-hydroxyethyl)-3-methyl-1-imidazol-3-ium ($[IL\text{-}OH]^+$), and **(iii)** 1-(2-carboxyethyl)-3-methyl-1-imidazol-3-ium ($[IL\text{-}COOH]^+$).

Two types of thermodynamic cycles, predominantly found in literature, were used to assess the relative acidity of three acid-functionalised imidazolium cations, 3-methyl-1-(4-sulfobutyl)-1-imidazol-3-ium ($[IL\text{-}SO_3H]^+$), 1-(2-hydroxyethyl)-3-methyl-1-imidazol-3-ium ($[IL\text{-}OH]^+$), and 1-(2-carboxyethyl)-3-methyl-1-imidazol-3-ium ($[IL\text{-}COOH]^+$) (Figure 3.1). Counterions were excluded from calculations. Given that they are generally similar in nature to the solvent IL anion (halides and/or basic), their effect cannot be distinguished from the solvent anion stabilisation effect assessed in Chapter 4.

3.1.1 pK_a Determination and Thermodynamic Cycles

Gas phase acidity can be defined in terms of the dissociation reaction of an acid HA, as shown below:

$$\mathrm{HA\ (g)} \rightarrow \mathrm{A^-(g)} + \mathrm{H^+(g)} \tag{3.1}$$

Solution phase acidities are often represented by the negative base-10 logarithm of the acid dissociation equilibrium constant (K_a), or simply pK_a,[135]

$$\mathrm{p}K_\mathrm{a} = \log_{10}(K_\mathrm{a}) \tag{3.2}$$

$$\text{with } K_\mathrm{a} = \frac{[\mathrm{A^-}][\mathrm{H^+}]}{[\mathrm{HA}]} \tag{3.3}$$

which is related to the reaction Gibbs free energy in solution ($\Delta G^*_{\mathrm{soln}}$) of the acid dissociation reaction as follows (See Appendix A for the derivation):

$$\mathrm{pK_a} = \frac{\Delta G^*_{\mathrm{soln}}}{RT\ln(10)} \tag{3.4}$$

$$\text{where } \Delta G^*_{\mathrm{soln}} = G^*_{\mathrm{soln}}(\mathrm{H^+}) + G^*_{\mathrm{soln}}(\mathrm{A^-}) - G^*_{\mathrm{soln}}(\mathrm{HA}) \tag{3.5}$$

The superscript * in Equations 3.4 and 3.5 indicates that the terms are calculated at a standard state concentration of 1 mol L^{-1}. This contrasts with a reference state of 1 atm (°), commonly used for calculating gas phase energies and can be converted to a 1 mol L^{-1} reference state using the following correction term (see Appendix B for the derivation)

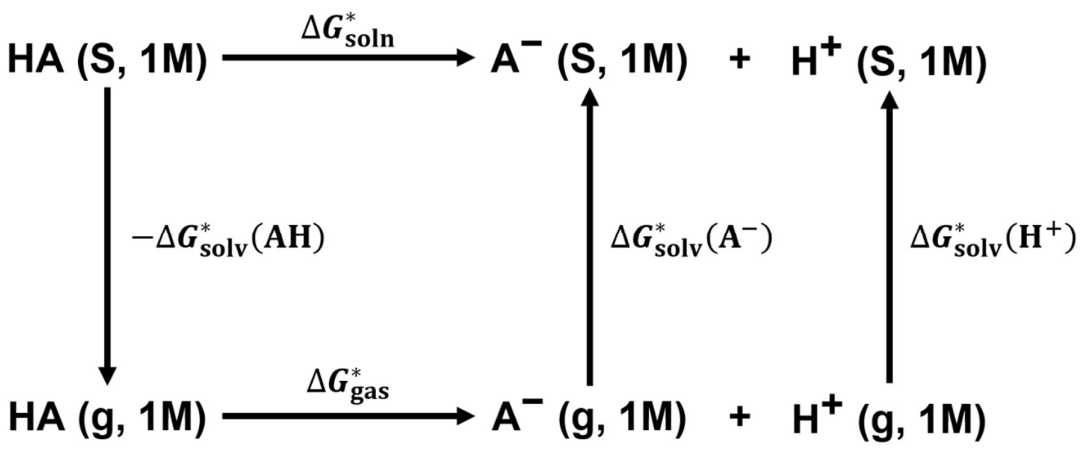
$$\Delta G^{\circ\rightarrow *} = \Delta n RT \ln\left(\tilde{R}T\right) \quad \textbf{(3.6)}$$

where Δn is the change in moles across the reaction and R and $\tilde{R}$ are the universal gas constant in units of J mol^{-1} K^{-1} and L atm K^{-1} mol^{-1}, respectively. This is of great significance in thermodynamic cycles where correct application demands all components to be in the same standard state.[136]

While many experimental techniques exist to determine the acidity of a compound in either gas or solvent phase, they are not easily applied in an accurate manner and limited in application to compounds at the extremities of acidity (super acids or very weak acids).[137] These limitations have spurred the development of in silico procedures for accurately predicting acidities of novel compounds in any solvent.

In general, computational methods require calculation of the reaction Gibbs free energy change in solution ($\Delta G^*_{\mathrm{soln}}$), which can be directly determined using continuum solvation models, such as the conductor-like polarizable continuum model (C-PCM) and the Solvation Model based on Density (SMD). These models have been shown to accurately predict solvation energies ($\Delta G^*_{\mathrm{solv}}$) but are limited in terms of total free energies in solution ($\Delta G^*_{\mathrm{soln}}$) prediction accuracy by the original design, which used less accurate low levels of theory (basis set size and correlation level) and specific compound classes.[135,138] Furthermore, given that an error of 5.7 kJ mol^{-1} in $\Delta G^*_{\mathrm{soln}}$ results in an error of 1 pK_a unit,[139] accurate pK_a predictions require highly accurate energy calculations. However, unlike gas phase protocols, condensed phase protocols are typically dependent on empirical parameterisation, and have no way of *systematically* improving the computation based on a well-defined and convergent theoretical framework.[137]

Thermodynamic cycles provide an alternative that seeks to improve the accuracy of the condensed phase reaction by coupling it to the equivalent gas phase reaction, which can be experimentally determined or calculated using high accuracy computational methods.[137] The solvation free energy, which is the change in free energy associated with the solute moving from gas to solvent phase,[140] links the reactions, completing the cycle as illustrated below in Scheme 3.1.



Scheme 3.1 Absolute or direct thermodynamic cycle for generic acid HA.

The free energy change in solution ($\Delta G^*_{\mathrm{soln}}$) can then be calculated as follows:

$$\Delta G^*_{\mathrm{soln}} = \Delta G^*_{\mathrm{gas}} + \Delta\Delta G^*_{\mathrm{solv}} \quad \textbf{(3.7)}$$

where for the cycle shown in Scheme 3.1

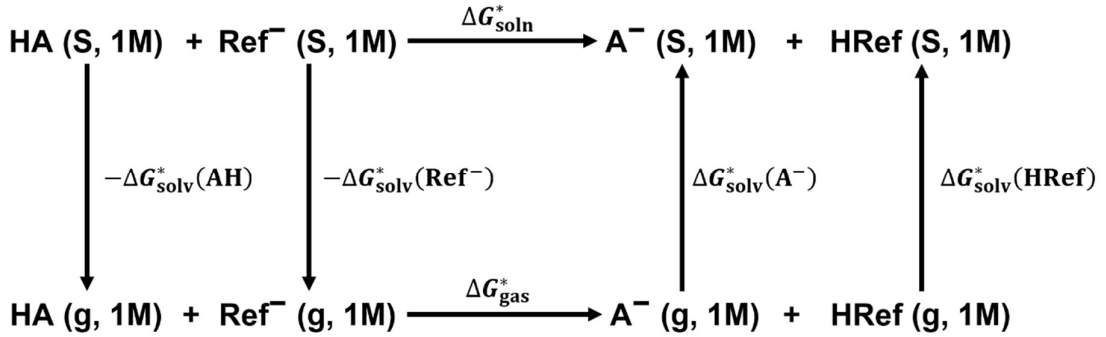
$$\Delta\Delta G^*_{\mathrm{solv}} = \Delta G^*_{\mathrm{solv}}(\mathrm{H}^+) + \Delta G^*_{\mathrm{solv}}(\mathrm{A}^-) - \Delta G^*_{\mathrm{solv}}(\mathrm{HA}) \quad \textbf{(3.8)}$$

$$\Delta G^*_{\mathrm{gas}} = G^*_{\mathrm{gas}}(\mathrm{H}^+) + G^*_{\mathrm{gas}}(\mathrm{A}^-) - G^*_{\mathrm{gas}}(\mathrm{HA}) \quad \textbf{(3.9)}$$

Using this process, each component of $\Delta G^*_{\mathrm{soln}}$ can be calculated at their optimal levels of theory for greatest accuracy.[137] Thus, the solvation terms can be calculated using continuum models at their original parameterisation levels of theory, producing the most accurate solvation energies; and the gas phase energies can be calculated at higher levels of theory that are more suitable for the calculation of reaction energies.[135]

Aside from allowing independent improvement in accuracy for each reaction phase, various thermodynamic cycle types can be used to allow for the use of available empirical data and cancellation of method (solvent model or level of theory) errors through structural or charge conservation across the reaction or phases.[139] Most however, tend to be extensions of two key cycles.[141] The first deals with the dissociation reaction in its simplest form and thus attempts a direct calculation of the acidity.[135,137,139] It is aptly named the *direct* or *absolute* cycle, an example of which is shown in Scheme 3.1. The second determines the acidity indirectly by using a reference acid/base system and calculating the reaction Gibbs free energy of a proton transfer reaction. Essentially, the acid of interest protonates the reference base (Ref^-), forming

the respective conjugate base (A^-) and reference acid (HRef) in what is called a *proton exchange cycle* (Scheme 3.2).[135,137]



Scheme 3.2 The proton exchange cycle for generic acid, HA, and reference acid, HRef.

The pK_a is calculated relative to the known pK_a of the reference acid (pK_a(HRef)),[142] such that

$$\mathrm{p}K_\mathrm{a} = \frac{\Delta G^*_\mathrm{soln}}{RT\ln(10)} + \mathrm{p}K_\mathrm{a}(\mathrm{HRef}) \qquad \textbf{(3.10)}$$

ΔG^*_soln is once again calculated using Equation 3.7, but the solvation and gas phase terms are defined in accordance with the proton exchange cycle as:

$$\Delta\Delta G^*_\mathrm{solv} = \Delta G^*_\mathrm{solv}(\mathrm{HRef}) + \Delta G^*_\mathrm{solv}(\mathrm{A}^-) - \Delta G^*_\mathrm{solv}(\mathrm{HA}) - \Delta G^*_\mathrm{solv}(\mathrm{Ref}^-) \qquad \textbf{(3.11)}$$

$$\Delta G^*_\mathrm{gas} = G^*_\mathrm{gas}(\mathrm{HRef}) + G^*_\mathrm{gas}(\mathrm{A}^-) - G^*_\mathrm{gas}(\mathrm{HA}) - G^*_\mathrm{gas}(\mathrm{Ref}^-) \qquad \textbf{(3.12)}$$

While both cycles have been shown to produce accurate results, each has its advantages and disadvantages, making the choice often dependant on the availability of experimental quantities for the systems of interest. The direct cycle has been shown to be very accurate for systems solvated in water or other commonly used solvent systems,[142–144] due to the availability of experimental values for the solvation free energy of the proton. Since the bare proton lacks electron density, this latter quantity cannot be computed by implicit solvation methods that rely on a description of the electronic structure of the system.[139] However, the accuracy of the cycle is clearly dependent on this value, which is often limited by experimentally available data and within said data subject to a fair amount of variation.[135,139,145] Moreover, charge is not conserved across the reaction, with a neutral reactant producing two charged products (Scheme 3.1). This may result in significant errors in the solvation terms as implicit solvation treatment poorly represents intermolecular short-range interactions,[146] thought to be crucial to ion/charged solute solvation.[147,148]

The proton exchange cycle is also reliant on experimentally available quantities, specifically with regards to the choice of reference acid, which is also limited by the need for a structurally similar compound to the acid of interest, for greatest accuracy. Two reasons can be given for this. First, the implicitly calculated solvation free energy depends on the overall electron density and empirical atomic descriptors; thus, if these are similar, the errors in calculated free energy should be similar.[138,147,149] Second, there is in inherent error in correlation energy in the electron structure calculation, and compounds with a similar number of bonds and bond types can be assumed to have errors of the same order (this concept is the core of so-called "isodesmic reactions").[150–152] Thus, the same problem as seen for the direct cycle is encountered, whereby effective use is limited by availability of external data. Furthermore, the cycle generates equal numbers of charged components and thus allows for a cancellation of errors introduced by deficiencies in the treatment of charged solutes.[146] This cancellation would not occur in most direct cycle cases as both products are charged while the reactant typically has an overall charge of zero.

In summary thus, the choice of a thermodynamic cycle is often dependant on the availability of experimental data for the systems of interest, limiting solvent system choice to more commonly used systems; or accuracy of theoretical means used to circumvent these shortcomings, one of which is discussed below.

3.1.2 Cluster Solvation Using Thermodynamic Cycles

The use of a direct thermodynamic cycle requires a value for the proton solvation free energy. These values are often not readily available in non-traditional solvents and if available highly contested due to large variation seen in experimental values available in literature.[135,143,145,153–157] Proton solvation values are therefore often estimated using computational techniques; however, besides the inability of electronic structure methods to compute the bare H^+ nucleus,[139] the protocol is further complicated by inherent inadequacies in continuum charged solute (ion) solvation.[158] Thus, proton solvation free energies are often calculated using purely explicit methods or hybrid explicit-implicit methods.[137,139,156] Explicit solvation can be time consuming and has a high computational cost; thus, hybrid explicit-implicit methods such as the cluster-continuum approach are often used.[148,158]

Explicit-implicit techniques work on the assumption that intermolecular solute-solvent interactions decrease in significance with distance from the solute, with the most significant interactions occurring in the first solvation shell.[148,158,159] Thus, the deficiencies of continuum

models can be addressed by the introduction of explicit solvent molecules around the solute, which also better approximates the specific solute-solvent intermolecular interactions that might be absent in the implicit model (Figure 3.2).[158]

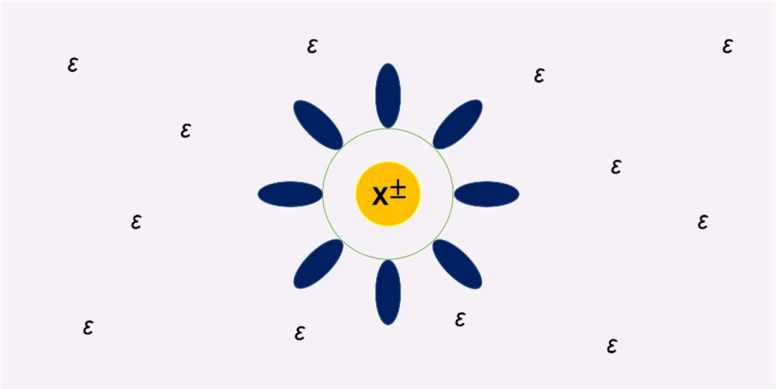


Figure 3.2 Simplified visualisation of hybrid explicit-implicit solvation, where an ion ($X^{\pm}$) is clustered by explicit solvent molecules (blue ovals) encapsulated in a dielectric continuum (ε, pale purple). Picture adapted from Xue et al. (2015).[160]

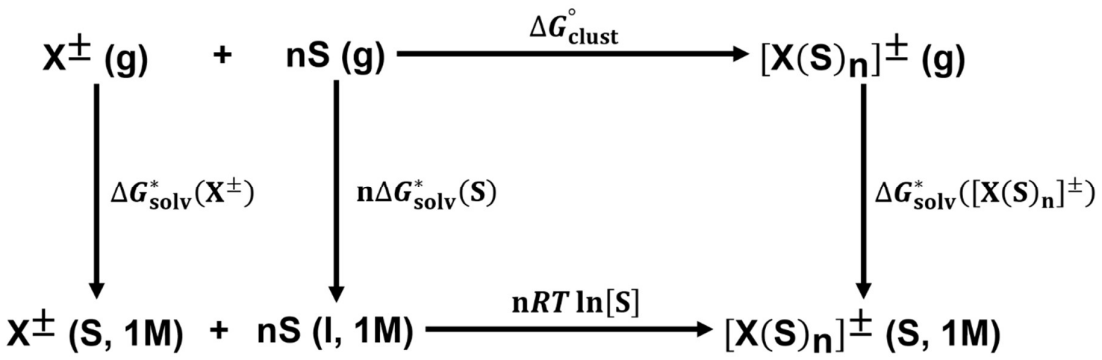
The *cluster-continuum model*, as proposed by Pliego and Riveros, surrounds the system of interest with an explicit cluster of solvent molecules, which is then embedded in a continuum.[148] It was specifically designed for the calculation of the solvation thermodynamic properties of ionic species. The explicit cluster represents the first solvation shell, integrating strong solute-solvent interactions that affect the solvent behaviour at small distances into a continuum model.[135,137] The technique is based on the notion that the cluster ($[X(S)_n]^{\pm}$) is a rigid species whose formation can be treated as a separate chemical reaction

$$X^{\pm}\,(g) + nS\,(g) \rightarrow [X(S)_n]^{\pm}\,(g) \qquad \textbf{(3.13)}$$

where $X^{\pm}$ is the ion of interest and S an explicit solvent molecule, which can then be modelled using *ab initio* methods.[148] The cluster formation reaction (3.13) is then solvated by the bulk solvent, represented by the continuum model, resulting in the following overall reaction[148,160]

$$X^{\pm}\,(g) + nS\,(l) \rightarrow [X(S)_n]^{\pm}\,(S) \qquad \textbf{(3.14)}$$

Approximating the cluster solvation process as a reaction allows for the application of thermodynamic cycles in cluster-continuum calculations. Using the above reaction, the thermodynamic cycle shown below can be constructed (Scheme 3.3).



Scheme 3.3 Cluster-continuum *monomer* cycle, with the solvent treated as individual non-interacting monomers (nS).[136,148]

The solvation free energy, according to the above cycle, is calculated as follows:

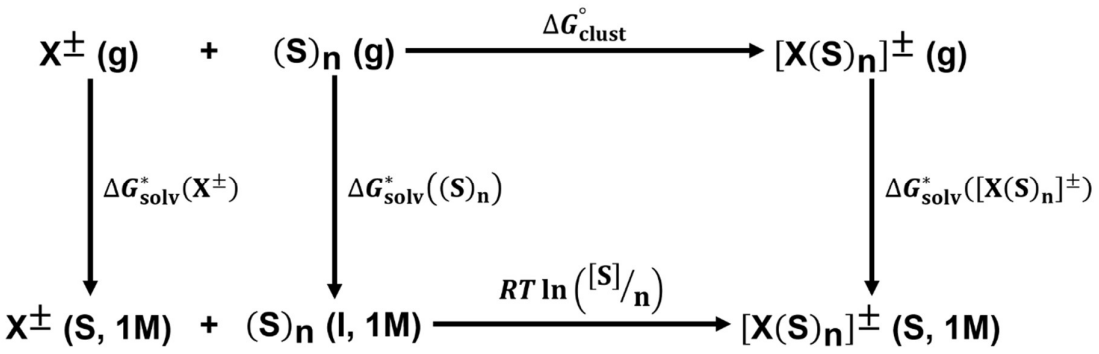
$$\Delta G^*_{\mathrm{solv}}(\mathrm{X}^{\pm}) = \Delta G^{\circ}_{\mathrm{clust}}([\mathrm{X(S)_n}]^{\pm}) - \mathrm{n}\Delta G^{\circ\rightarrow *} + \Delta G^*_{\mathrm{solv}}([\mathrm{X(S)_n}]^{\pm}) - \mathrm{n}\Delta G^*_{\mathrm{solv}}(\mathrm{S}) - \mathrm{n}RT\ln[\mathrm{S}] \quad \textbf{(3.15)}$$

The solvation free energy of an ion $\mathrm{X}^{\pm}$ can thus be determined from the solvation free energy of the solvent ($\Delta G^*_{\mathrm{solv}}(\mathrm{S})$), the concentration of the pure solvent ($RT\ln[\mathrm{S}]$), the solvation free energy of the cluster ($\Delta G^*_{\mathrm{solv}}([\mathrm{X(S)_n}]^{\pm})$) and the cluster formation free energy at a standard state concentration of 1 mol L^{-1} ($\Delta G^*_{\mathrm{clust}}([\mathrm{X(S)_n}]^{\pm}) = \Delta G^{\circ}_{\mathrm{clust}}([\mathrm{X(S)_n}]^{\pm}) - \mathrm{n}\Delta G^{\circ\rightarrow *}$), which is equal to the reaction energy of reaction 3.13.

$$\Delta G^*_{\mathrm{gas}}(3.13) = \Delta G^*_{\mathrm{clust}}([\mathrm{X(S)_n}]^{\pm}) = G^*_{\mathrm{gas}}([\mathrm{X(S)_n}]^{\pm}) - G^*_{\mathrm{gas}}(\mathrm{X}^{\pm}) - \mathrm{n}G^*_{\mathrm{gas}}(\mathrm{S}) \quad \textbf{(3.16)}$$

This approach employs a monomer treatment of the solvent cluster, whereby the clustering reaction is treated as a reaction between the solute and n number of non-interacting solvent monomers (S).[136] Aptly named the *monomer cycle* (Scheme 3.3), this represents the simplest cluster cycle. However, it disregards changes in solvent-solvent interactions that occur when the solute is introduced.

Thus, an alternative method treats the reactant form of the solvent as an interacting cluster ($(\mathrm{S})_\mathrm{n}$), reacting the solute with the entire solvent cluster and not individual monomers (Scheme 3.4).[136] This approach is intended to include the solvent-solvent interactions and changes in solvent-solvent interactions with the introduction of a solute, neglected by the monomer cycle.



Scheme 3.4 Cluster-continuum *cluster* cycle, with the solvent treated as an interacting cluster ($(S)_n$).[136,148]

Termed the *cluster cycle*, this technique allows the solute solvation free energy to be calculated according to

$$\Delta G^*_{\mathrm{solv}}(X^{\pm}) = \Delta G^{\circ}_{\mathrm{clust}}([X(S)_n]^{\pm}) - \Delta G^{\circ\rightarrow *} + \Delta G^*_{\mathrm{solv}}([X(S)_n]^{\pm}) - \Delta G^*_{\mathrm{solv}}((S)_n) - nRT\ln\left({}^{[S]}/_{n}\right) \quad \textbf{(3.17)}$$

where the equivalent of reaction 3.13 now becomes

$$X^{\pm}\,(g) + (S)_n\,(g) \rightarrow [X(S)_n]^{\pm}\,(g) \quad \textbf{(3.18)}$$

with the cluster formation free energy calculated as:

$$\Delta G^*_{\mathrm{gas}}(3.18) = \Delta G^*_{\mathrm{clust}}([X(S)_n]^{\pm}) = G^*_{\mathrm{gas}}([X(S)_n]^{\pm}) - G^*_{\mathrm{gas}}(X^{\pm}) - G^*_{\mathrm{gas}}((S)_n) \quad \textbf{(3.19)}$$

Bryantsev et al. found this solvent cluster representation to be far more accurate for aqueous solvation of selected cations.[136] However, literature still largely favours the use of the monomer cycle, which has been shown to be reasonably accurate, without comparison to other cycle types.[160–164]

The cluster-continuum model has since been expanded upon by Xue et al. and used for the solvation of ions by ionic liquids, referred to as the *ion-biased cluster-continuum model*. This technique is based on the cluster-continuum method but assumes that for ion-based solvents (like ionic liquids), solvent clusters consist of only the ionic liquid component (cation or anion) of opposite charge to that of the solute.[160] This is based on various studies which indicate that for some ionic liquids, solvent ion pairs are ephemeral due to weak association between the charged ions.[165–167] The solvation of positively charged ion, X^+, in a generic ionic liquid (IL), can then be represented by the chemical reaction,

$$\mathrm{X^+\ (g) + n[A]^-\ (IL) \rightarrow [X(A)_n]^{1-n}\ (IL)} \quad \textbf{(3.20)}$$

Where IL represents the full ion pair, composed of a generic cation ($[\mathrm{C}]^+$) and anion ($[\mathrm{A}]^-$). The solvation free energy, according to a monomer cycle, is calculated as follows:

$$\Delta G^*_{\mathrm{solv}}(\mathrm{X^+}) = \Delta G^\circ_{\mathrm{clust}}([\mathrm{X(A)_n}]^{1-\mathrm{n}}) - \mathrm{n}\Delta G^{\circ\rightarrow *} + \Delta G^*_{\mathrm{solv}}([\mathrm{X(A)_n}]^{1-\mathrm{n}}) - \mathrm{n}\Delta G^*_{\mathrm{solv}}([\mathrm{A}]^-) - \mathrm{n}RT\ln[\mathrm{IL}] \quad \textbf{(3.21)}$$

As observed above, methodology to theoretically predict solvation free energies of ions using thermodynamic relationships and cycles is highly variable. Aside from the choice of gas phase level of theory and implicit solvent model, there are now extra factors to consider related to the explicit solvent cluster. As illustrated above, how the cluster is represented is significant and one should consider whether the solvent molecules interact, which components of a solvent are present and the size of the solvent cluster. Key to all solvent and solute systems considered is *n*, the number of explicit solvent molecules used or size of the solvent cluster. Original cluster-continuum literature and subsequent studies usually determine this number variationally, selecting the *n* corresponding to the lowest solvation free energy. However, the value can also be chosen as a convergence point of increasing values of *n* (seen for cluster cycles) or to match experimental first shell coordination numbers.[148,157,162,168] Furthermore, Xue et al. raised the possibility of steric issues and Coulombic interactions, whereby solvent molecule size and charge dictated where solvent ions would be placed and how many could be fitted around the central ion of interest.[168]

3.2 Computational Details

Thermodynamic cycle calculations were done at different levels of theory to benchmark the most accurate combination of gas phase and solvent phase methods. Accordingly, $\Delta G^*_{\mathrm{gas}}$ and $\Delta\Delta G^*_{\mathrm{solv}}$ were calculated using various levels of theory.

3.2.1 Gas Phase Calculations

Highly accurate gas phase calculations are usually performed using correlated methods such as configuration interaction (CI) or coupled cluster (CC), ideally with corrections including single, double and triple excitations to the reference determinant. In addition, extrapolation to the basis set limit can be done to correct for basis set deficiencies. However, these methods are computationally expensive and tend to have high calculation run times, more so for large molecules, such as cellobiose. Thus, based on evidence of high accuracy with reasonable time

efficiency for larger more complex chemical systems,[71,76,82,97,100,139,142,143,169,170] four high level (HL) methods were chosen, namely CBS-QB3, G3B3(MP2), G4(MP2) and DLPNO-CCSD(T).[71,90,91,94,100] The first three are composite methods, described in Chapter 2, and the last is a coupled cluster method shown to have near-linear scaling for large chemical systems.[71] Given the computational cost of composite methods, a lower level DFT method was used for all geometry optimisations. All structures were initially optimised using TPSS/def2-TZVP[171,172] and the resultant conformations used as initial geometries for all subsequent CBS-QB3, G3B3(MP2), G4(MP2) and DLPNO-CCSD(T) energy calculations. Note that composite methods have an optimisation and frequency calculation built into their standard protocol, in order to include appropriately scaled zero point energies, and the use of TPSS/def2-TZVP[171,172] initial geometries was only to assure consistent conformations in each case. More details of the specific levels of theory (LoTs) and scaling factors used are given in Chapter 2: Section 2.4. At the time of study, DLPNO-CCSD(T) was only implemented for electronic energy calculations or single point calculations. DLPNO-CCSD(T) electronic energies were thus converted to free energies using the TPSS/def2-TZVP thermal corrections. The scaling factor for TPSS/def2-TZVP is approximately 1 (1.002) and thus the thermal corrections were taken as calculated.[173]

All optimisations and the composite method free energy calculations were run using Gaussian09.[174] Tight convergence criteria (for both the SCF and geometry convergence criteria) and an ultrafine grid for numerical integration of the exchange-correlation energy were used. DLPNO-CCSD(T) has only been implemented in the ORCA package and thus all DLPNO-CCSD(T) single-point calculations were run with ORCA 4.0.1,[175] in conjunction with the cc-pVTZ[176–180] basis set, and aug-cc-pVTZ[181–183] fitting basis set for the calculation of the correlation energy and tight SCF convergence criteria.

3.2.2 Solvent Phase Calculations

The calculation of $\Delta\Delta G^*_{\mathrm{solv}}$ involves the determination of each reactant and product's solvation free energy ($\Delta G^*_{\mathrm{solv}}$) according to the following equation for a generic acid HA:

$$\Delta G^*_{\mathrm{solv}}(\mathrm{HA}) = G^*_{\mathrm{soln}}(\mathrm{HA}) - G^*_{\mathrm{gas}}(\mathrm{HA}) \qquad \textbf{(3.22)}$$

The conformations resulting from initial optimisation using TPSS/def2-TZVP[171,172] were used as input geometries for gas phase optimisation and frequency calculations (to determine $G^*_{\mathrm{gas}}(\mathrm{HA})$). The level of theory was changed to B3LYP/6-31G* and M05-2X/6-31+G**,[112–

[115,184] both chosen out of the list of levels of theory used to originally parameterise the SMD solvent model to remain consistent with the original parameterisation.[127] The resultant B3LYP/6-31G* and M05-2X/6-31+G**[112–115,184] gas phase optimised geometries were used as initial geometries for subsequent solvent phase optimisation and frequency calculations using the SMD model at the same levels of theory.[18,136] All calculations were performed using Gaussian09[174] with the same convergence criteria used as for the gas phase leg calculations.

3.2.3 Cluster Calculations

All quantities required for cluster calculations, see Section 3.1.2, were determined using the B3LYP/6-31G* and M05-2X/6-31+G** levels of theory. Structures were optimised in both gas and solvent phase with tight convergence criteria applied.

While cluster-continuum theory provides the theoretical framework for the solvation of ions with ionic liquids, the practicalities of efficiently generating viable clusters require the use of an automated procedure that can globally optimise cluster geometries. Consequently, the software package ABCluster,[185] which implements an artificial bee colony algorithm, was used to generate ionic liquid clusters of various sizes (determined by *n*, the number of solvent molecules). For each solvent molecule type (full ion pair or anion) a series of 250 different optimised proton-solvent arrangements were generated. Clusters were generated using an initial structure guess of sphere with a radius of 10 Å.

To reduce computational expense, the 250 solute-solvent cluster arrangements were initially optimised using GFN-xTB.[186] The five lowest energy arrangements were then reoptimized at the selected levels of theory. Some clusters were altered before optimisation to more closely resemble arrangements found in literature.[187] For example, when examining the clusters generated for a proton surrounded by a two $[NTf_2]^-$ anions, it was noted that all the lowest energy arrangements had the proton covalently bonded to one of the sulfonyl oxygen atoms and the bonding configuration of the anion altered (e.g. Figure 3.3 shows changes in the configuration around the nitrogen atom). This is inconsistent with what would be expected from the structure of the conjugate acid.[160]

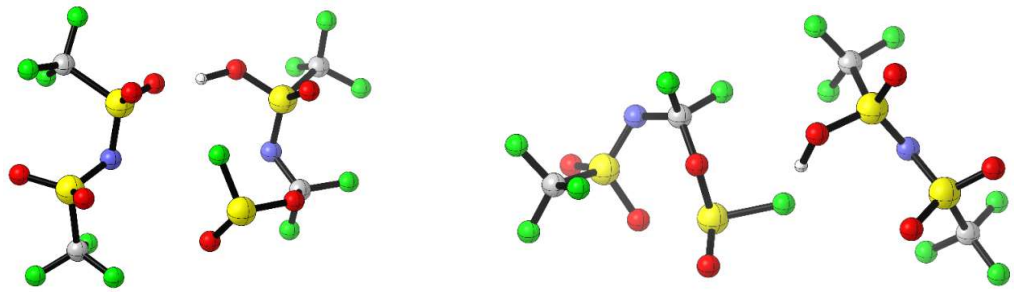


Figure 3.3 ABCluster GFN-xTB lowest electronic energy $[H(NTf_2)_2]^-$ clusters.

Analysis of a $[H(NTf_2)]$ complex was used to assess the viability of GFN-xTB to describe the protonation. A series of structures were created, each with a different site protonated, and optimised using GFN-xTB, B3LYP/6-31G* (through a G3B3(MP2) protocol) and TPSS/def2-TZVP to assess which produced the lowest energy structure. Anion conformation and orientation to each other based on NMR observations and the findings of Xue et al.[160,188]

The protonation of the nitrogen atom was indeed most favourable in all cases, specifically also when using GFN-xTB. The anomalous configurations produced in the cluster generation were thus an artefact of the procedure and not due to an inherit shortcoming of the semi-empirical method.

3.2.4 SMD Extended for Ionic Liquids

The SMD solvation model, explained in more detail in Chapter 2, is termed universal in that it can be applied to any solvent so long as certain macroscopic solvent parameters are available.[127] However, given the limited experimental data available for ionic liquids, Bernales et al. formulated a procedure that converts available experimental data to allow SMD simulation of more ionic liquids.[18] This involves the determination of a correlation between two sets of scaled hydrogen bonding parameters and the generation of generic solvent parameters from a series of ionic liquids to supplement known values or create a general ionic liquid solvation environment.

The hydrogen bonding parameters needed for SMD input are the Abraham parameters ($\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$), determined in terms of free energy data of solute behaviour, and are often not available for ionic liquids. Rather, Kamlet-Taft hydrogen bonding parameters (α and β), determined in terms of solvent behaviour data, are more commonly available. Bernales et al. thus determined a linear correlation between the parameters using a series of organic solvents

for which both sets of hydrogen bonding parameters were available. Their predictive relationships, with corresponding coefficients of determination (R^2), are

$$\Sigma\alpha_2^{\mathrm{H}} = 0.4098\alpha + 0.0064 \qquad R^2 = 0.94 \qquad \textbf{(3.23)}$$

$$\Sigma\beta_2^{\mathrm{H}} = 0.6138\beta + 0.0890 \qquad R^2 = 0.61 \qquad \textbf{(3.24)}$$

To address cases where the remaining solvent parameters are also not known (or known partially), two further protocols were created. The first uses averages of all solvent descriptors across a series of ionic liquids, for which all measurements are available in literature, to create a generic ionic liquid solvation model (SMD-GIL). The other model serves as a compromise, where generic values are used to supplement existing descriptors, resulting in a partial generic parameters SMD model (SMD-PGP).

The ionic liquid solvents used throughout the study with their SMD solvent descriptors are displayed in Table 3.1 below. They were chosen based on experimental studies that have been done on acidity in ionic liquid mediums and cellulose solubility.

Table 3.1 Ionic liquid solvents and their respective experimental solvent parameters used for all implicit calculations performed throughout this work.

Gaussian input	[C_4C_1im][BF_4]	[C_4C_1im][Cl]	[C_4C_1im][$MeSO_4$]	[C_4C_1im][NTf_2]
`eps` = ε (dielectric constant)	11.70[a]	15.00[b]	14.80[b]	11.52[c]
`epsinf` = n^2 (refractive index)^2	2.0207[d]	2.0449[*]	2.1818[e]	2.0366[f]
`HBondBasicity` = $\Sigma\beta_2^{\mathrm{H}}$ (Abraham's hydrogen bond basicity)	0.320[g]	0.605[h]	0.501[i]	0.238[g]
`HBondAcidity` = $\Sigma\alpha_2^{\mathrm{H}}$ (Abraham's hydrogen bond acidity)	0.263[g]	0.187[h]	0.230[i]	0.259[g]
`SurfaceTensionAtInterface` = γ (surface tension) [cal mol^{-1} $Å^{-2}$ (298 K)]	67.07[f]	69.37[j]	62.32[k]	53.97[f]
`CarbonAromaticity` = ϕ(fraction of non-hydrogen atoms that are aromatic carbon atoms)	0.200	0.273	0.188	0.120
`ElectronegativeHalogenicity` = ψ (fraction of non-hydrogen atoms that are electronegative halogen atoms)	0.267	0.091	0.000	0.240

[*]Solvent-independent generic refractive index average[18]
[a]Ref 189
[b]Ref 190
[c]Ref 191
[d]Ref 192
[e]Ref 193
[f]Ref 194
[g]Kamlet-Taft parameters taken from Ref 195
[h]Kamlet-Taft parameters taken from Ref 196
[i]Kamlet-Taft parameters taken from Ref 197
[j]Ref 198
[k]Ref 199

3.2.5 RMSD Calculation and Hydrogen Bond Analysis

The changes in structure and intermolecular bonding were assessed using RMSDs and comparing differences in the number and geometry of intramolecular hydrogen bonds. RMSDs were generated using the `Match` functionality for all heteroatoms and carbon atoms present in the structure, using UCSF Chimera 1.13.1.[200] Possible hydrogen bonding present was determined using the Chimera `FindHbond` functionality with the bond angle restrictions widened by 60°.[201]

3.2.6 Graphics Generation

Line structures graphics were generated using PerkinElmer ChemDraw Professional 15.1.0.144. UCSF Chimera 1.13.1[200] was used to generate images showing overlaid conformers as well as some individual conformer images displaying the predicted intramolecular hydrogen bonding. Other images were created using CYLview v1.0.565 BETA.[202]

3.3 Results and Discussion

3.3.1 Establishing a Thermodynamic Cycle Protocol

The AIL cations were chosen to represent common Brønsted acid-functionalities observed in experimental cellulose hydrolysis studies (Figure 3.1).[38] However, their experimental pK_as in ionic liquids are not known. Thermodynamic cycles can be employed to predict these values; however, the quality of the results would depend on the cycle used. In such cases, known systems can be used to test the accuracy and precision of a protocol/technique indicating the error associated with values predicted for related (but unknown) systems.[203] Thus, by using a series of N-base amine conjugate acids (Figure 3.4) with known pK_a values in [C_4C_1im][NTf_2] and [C_4C_1im][BF_4],[204] total error and probable sources of error associated with each protocol can be determined.

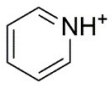
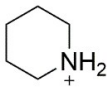
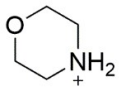
			
	(i)	**(ii)**	**(iii)**
$[C_4C_1im][NTf_2]$	*10.5*	*18.4*	*16.6*
$[C_4C_1im][BF_4]$	*8.6*	*15.6*	*13.9*

Figure 3.4 Structures and experimental pK_as, in $[C_4C_1im][NTf_2]$ and $[C_4C_1im][BF_4]$, of the N-base amine conjugate acid model systems **(i)** pyridin-1-ium, **(ii)** piperidin-1-ium and **(iii)** morpholin-4-ium, used to test thermodynamic cycle protocol viability in ionic liquid pK_a prediction.[204]

As detailed in Section 3.1.1, each of the cycles of interest has inherent sources of error. However, the accuracy of a thermodynamic cycle is dependent on the level of theory and thus, a protocol, in the context of this study, is defined as a gas/solvent phase level of theory combination applied to a specific thermodynamic cycle.

Solvent phase accuracy is dominated by the solvation model, in this case limited to an implicit SMD model. Thus, to optimise solvent model performance, solvent phase level of theory options were chosen to match those used initially to test and parameterise the solvent model.[127]

Gas phase levels of theory are chosen to deliver high accuracy. However, as was discussed in Chapter 2, increases in accuracy are typically a result of larger basis sets and an improved treatment of electron correlation, which places constraints on the size of the system due to the computational costs (Figure 2.1).[78,79,203] Given the size of the system of interest, cellobiose, calculation costs are thus a significant factor in the choice of level of theory. Domain-based local pair natural orbital coupled cluster theory with single-, double-, and perturbative triple excitations (DLPNO-CCSD(T)) allows for coupled-cluster treatment of large molecular systems with near linear scaling, resulting in relatively low calculation times for large and complex systems.[76] Given its computational efficiency, application to cellulose ionic liquid processing calculations seems prudent. Considering this, the comparative accuracy of DLPNO-CCSD(T) to established efficient high accuracy methods is of interest. This method, along with three composite methods, were therefore combined with two DFT-based solvent levels of theory to determine the best overall level of theory combination and the compare the performance of DLPNO-CCSD(T). The solvents $[C_4C_1im][NTf_2]$ and $[C_4C_1im][BF_4]$ were

represented implicitly using the SMD model. These solvents were chosen based on availability of experimental data.[38]

The viability of the protocol was assessed in terms of its accuracy and precision, with the error defined as the variation of the calculated pK_a values ($\hat{x}_i$) from their experimentally determined equivalents found in literature (x_i).

Accuracy is defined as the extent a result varies from a standardised value,[205] is measured using a mean absolute error (MAE), determined according to

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |\hat{x}_i - x_i|}{n} \tag{3.25}$$

where n is the number of results and $|\hat{x}_i - x_i|$ is the absolute error.

Precision, on the other hand, is defined as the closeness between a series of measurements or results and used to assess the consistency of a protocol's accuracy across various conditions (model compound, solvent system, level of theory, etc).[205] It is represented by the standard deviation (SD) of absolute errors (e), as shown below

$$\mathrm{SD} = \sqrt{\frac{\sum_{i=1}^{n} (e_i - \mathrm{MAE})^2}{n-1}} \tag{3.26}$$

$$\text{with } e_i = |\hat{x}_i - x_i| \tag{3.27}$$

3.3.1.1 Proton Exchange Cycle

As was detailed in the background, the accuracy of this cycle is highly dependent on the reference acid (HRef) chosen, specifically how structurally similar the reference acid is to the examined acid (HA).[135] Thus, for the N-base amine conjugate acids, 2-piperazin-4-ium-1-ylethanol and piperazin-4-ium-1-carbaldehyde (Figure 3.5) were chosen as reference acids given that they are N-base amine conjugate acids of similar molecular structure with known acidities in the ionic liquids of interest.[38]

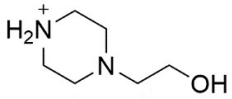
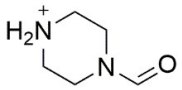
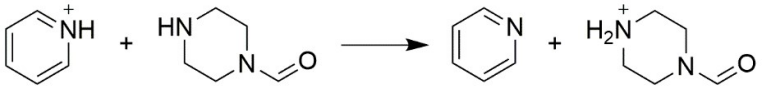
		
	(i)	**(ii)**
$[C_4C_1im][NTf_2]$	*14.3*	*13.2*
$[C_4C_1im][BF_4]$	*17.5*	*13.4*

Figure 3.5 Structures and experimental pK_as, in $[C_4C_1im][NTf_2]$ and $[C_4C_1im][BF_4]$, of N-base amine proton exchange reference acids (HRef) **(i)** 2-piperazin-4-ium-1-ylethanol and **(ii)** piperazin-4-ium-1-carbaldehyde.[38]

Each cycle is based upon a proton transfer reaction, as seen in Scheme 3.5, where the reference acid conjugate base (Ref^-) acts as a proton acceptor.

(i)



(ii)

Scheme 3.5 Proton exchange reaction for pyridin-1-ium with reference base (Ref^-)/acid (HRef) systems **(i)** 2-piperazin-1-ylethanol/2-piperazin-4-ium-1-ylethanol and **(ii)** piperazine-1-carbaldehyde/piperazin-4-ium-1-carbaldehyde.

Table 3.2 summarises the impact of LoT on the results. These values are calculated from the absolute errors across all acids, reference acids and solvents combined. Thus, ignoring the structural similarity between acid of interest and reference acid and focusing only on the computational methods.

Values down a column show the effects of the gas phase LoT, with the final total column (light shading) displaying the results averaged across both solvent phase levels of theory. Likewise, results along a row show the effect of the solvent phase LoT, with the bottom total row (light shading) indicating each averaged over all gas phase levels of theory. The intersection of both

totals (in bold and dark shading) represents the performance of the proton exchange (PX) cycle as a whole for these N-base amine conjugate acids in implicitly represented ionic liquid solvents, combining the effects of the level of theory combination used and reference acid system (HRef) compatibility and experimental accuracy.

Table 3.2 Mean absolute error (MAE), standard deviation (SD) and minimum/maximum errors (MIN/MAX) for pK_a prediction of N-base amine conjugate acids using proton exchange (PX) cycles at the various level of theory (LoT) combinations.

LoT	*B3LYP/6-31G**				*M05-2X/6-31+G***				***Total***			
	MAE	**SD**	**MIN/MAX**		**MAE**	**SD**	**MIN/MAX**		**MAE**	**SD**	**MIN/MAX**	
CBS-QB3	1.06	0.69	0.02	2.16	1.04	0.55	0.19	2.08	1.05	0.61	0.02	2.16
G3B3(MP2)	0.92	0.56	0.13	1.79	0.88	0.52	0.04	1.77	0.90	0.53	0.04	1.79
G4(MP2)	0.93	0.57	0.13	1.75	0.89	0.52	0.04	1.76	0.91	0.54	0.04	1.76
DLPNO[a]	0.98	0.59	0.13	2.08	0.96	0.59	0.29	1.99	0.97	0.58	0.13	2.08
Total	0.97	0.59	0.02	2.16	0.94	0.53	0.04	2.08	**0.96**	**0.56**	**0.02**	**2.16**

[a]DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP

The most accurate procedures are given by the G*n* methods, G3B3(MP2) and G4(MP2), in combination with M05-2X/6-31+G**, which result in a MAE of 0.88 and 0.89 pK_a units respectively. M05-2X/6-31+G** consistently gives the best results when used for the solvent leg. Moreover, considering the level of theory used for the gas phase leg of the TC, the accuracy remains consistent as well, carrying across both solvent phase LoT. From highest to lowest accuracy (lowest MAE to highest MAE), the trend is G3B3(MP2) < G4(MP2) < DLPNO-CCSD(T)/cc-pVTZ < CBS-QB3.

The influence of the gas phase levels of theory appears to fall into two separate groupings, the G*n* methods on the one hand and DLPNO-CCSD(T) and CBS-QB3 on the other, which vary in the same manner for each solvent LoT. The G*n* methods differ from each other by 0.01 pK_a units independent of solvent LoT and increase by the same amount (0.04 pK_a units) between solvent LoT (from M05-2X/6-31+G** to B3LYP/6-31G**). The same trend is noted for DLPNO-CCSD(T)/cc-pVTZ and CBS-QB3. Between the solvent phase levels of theory, they both increase by 0.02 pK_a units from M05-2X/6-31+G** to B3LYP/6-31G*, and DLPNO-CCSD(T) is consistently 0.08 pK_a units lower than CBS-QB3.

In terms of accuracy, the greatest effect is seen for the gas phase LoT. MAE values differ by only 0.03 pK_a units between solvent LoT, compared to the maximum gas phase LoT difference of 0.15 pK_a units between CBS-QB3 and G3B3(MP2).

The precision results are consistent with the accuracy findings. This can be seen in the mimicry of the MAE LoT trends in the SD results. For B3LYP/6-31G* the most precise gas phase LoT is G3B3(MP2) followed by the other G*n* method (G4(MP2), then DLNPO-CCSD(T)/cc-pVTZ and lastly CBS-QB3. The only variation is seen for M05-2X/6-31+G** where CBS-QB3 has a lower SD (0.55 pK_a units) than DLPNO-CCSD(T) (0.59 pK_a units) and the G*n* methods are equivalent. Thus, for M05-2X the trend in order of increasing SD (decreasing precision) for the gas phase LoT order becomes G3B3(MP2) = G4(MP2) < CBS-QB3 < DLPNO-CCSD(T)/cc-pVTZ. Averaged over both solvent phase LoT, G3B3(MP2) appears to be the most precise method, with the remaining ordering following that seen before for the accuracy results, i.e.: G3B3(MP2) < G4(MP2) < DLPNO-CCSD(T)/cc-pVTZ < CBS-QB3.

Deviation between solvent phase LoT precision, averaged over all gas phase LoT, is rather large, with the difference between B3LYP/6-31G* and M05-2X/6-31+G** total SDs, double the MAE difference (0.06 pK_a units). Furthermore, the gas phase results are once again split into the same two groups as seen for accuracy. The first is comprised of the G*n* methods, and the second of DLPNO-CCSD(T)/cc-pVTZ and CBS-QB3. The first grouping once again showed the lowest values, implying highest precision. Variation amongst the gas phase LoT SD values was not significantly larger (0.08 pK_a units) than that seen between solvent LoT, unlike what was observed for the accuracy results.

These total results imply that individual phase LoT choices make little difference in terms of precision. However, when analysing the combined results, the averages are somewhat misleading. While the effect of solvent LoT is quite small, its effect on the overall performance in combination with composite methods is significant. Specifically, in terms of CBS-QB3's performance, which varied significantly between solvent levels of theory, with the B3LYP SD 0.14 pK_a units larger. DLPNO-CCSD(T) was the only LoT where solvent LoT choice made no notable difference on overall results, having the same overall SD for both solvent phase LoT.

Combined with the accuracy results, this implies that the gas phase LoT choice has the greater effect on PX cycle performance. Furthermore, accuracy performances correlate with precision performances, such that the most accurate method, G3B3(MP2) [M05-2X/6-31+G** (SMD)], is also the most precise. However, these observations do not consider the effect of the structural similarity of the reference acid to the model acid, a key determinate of performance for this cycle type. These effects are seen in the results when specific reactant system ($AH + Ref^-$) performances are assessed, as shown in Table 3.3.

Table 3.3 Mean absolute error (MAE), standard deviation (SD) and minimum/maximum errors (MIN/MAX) N-base amine conjugate acids calculated pK_a using 2-piperazin-4-ium-1-ylethanol and piperazin-4-ium-1-carbaldehyde proton exchange (PX) cycles.

Acid	*2-piperazin-4-ium-1-ylethanol*				*piperazin-4-ium-1-carbaldehyde*				***Total***			
	MAE	**SD**	**MIN/MAX**		**MAE**	**SD**	**MIN/MAX**		**MAE**	**SD**	**MIN/MAX**	
Pyridin-1-ium	1.16	0.74	0.12	2.16	0.85	0.50	0.05	1.74	1.00	0.64	0.05	2.16
Piperidin-1-ium	0.85	0.29	0.32	1.36	1.30	0.38	0.72	1.87	1.07	0.40	0.32	1.87
Morpholin-4-ium	0.35	0.26	0.02	0.80	1.25	0.46	0.70	1.79	0.80	0.59	0.02	1.79
Total	0.79	0.58	0.02	2.16	1.13	0.48	0.05	1.87	**0.96**	**0.56**	**0.02**	**2.16**

Indeed, clear variation in accuracy and precision is noted between reference systems. Using the piperazin-4-ium-1-carbaldehyde reference acid is far less accurate on average, with a MAE of 1.13 pK_a units to 2-piperazin-4-ium-1-ylethanol's 0.79 pK_a units. However, within each reference system there is also a noticeable difference in performance ordering. While all the MAE values are far lower in general with 2-piperazin-4-ium-1-ylethanol, morpholin-4-ium far outperforms the other N-base amine conjugate acids, with a MAE less than half that of the other acids (0.35 pK_a units). However, with piperazin-4-ium-1-carbaldehyde, pyridine-1-ium gives the lowest MAE, followed by morpholin-4-ium. Piperidin-1-ium performs the worst overall (1.07 pK_a units), possibly indicating lowest structural similarity or compatibility to the chosen reference acids, specifically piperazin-4-ium-1-carbaldehyde. Overall, morpholin-4-ium had the best accuracy (0.59 pK_a units), due to its good performance with 2-piperazin-4-ium-1-ylethanol far outperforming all the other combinations.

Precision values indicate a contrary trend, with piperidin-1-ium performing the best overall (total SD of 0.40 pK_a units). Morpholin-4-ium still produces the lowest overall SD for a reactant system (0.26 pK_a units) but performs very poorly in combination with piperazin-4-ium-1-carbaldehyde (0.46 pK_a units), skewing the overall result. Reference acid preference is completely inverted, with 2-piperazin-4-ium-1-ylethanol 0.10 pK_a units higher than piperazin-4-ium-1-carbaldehyde (total row). This change is due to a very poor performance in combination with pyridine-1-ium. Piperazin-4-ium-1-carbaldehyde's SD values for both piperidin-1-ium and morpholin-4-ium were at least 0.09 pK_a units higher. However, piperazin-4-ium-1-carbaldehyde's SD values vary far less between the different N-base conjugate acids, leading to an overall smaller total variation; whereas pyridin-1-ium's uncharacteristically large deviation with 2-piperazin-4-ium-1-ylethanol (0.74 pK_a units), greatly skewed the generally consistently superior performance shown for the other N-base conjugate acids.

The general trends noted above are reflected in the individual results displayed in Tables 3.4 and 3.5; however, the use of absolute values negated directionality in the errors. Most errors produced were negative, indicating that the PX cycle generally overestimates, with a few exceptions; namely, when the reference acid piperazin-4-ium-1-carbaldehyde was used in conjunction with a $[C_4C_1im][NTf_2]$ solvent system, all predicted pK_as, aside from one pyridine-1-ium value, were underestimated.

Table 3.4 Calculated gas phase reaction free energy and pK_a of N-base amine conjugate acids with reference base/acid 2-piperazin-1-ylethanol/2-piperazin-4-ium-1-ylethanol in $[C_4C_1im][BF_4]$ and $[C_4C_1im][NTf_2]$.

			$[C_4C_1im][BF_4]$				**$[C_4C_1im][NTf_2]$**			
			*B3LYP/6-31G**		*M05-2X/6-31+G***		*B3LYP/6-31G**		*M05-2X/6-31+G***	
HA	**Level of Theory**	**ΔG^*_{gas} (kJ mol^{-1})**	**Calcd pK_a**	**Error[a]**	**Calcd pK_a**	**Error[a]**	**Calcd pK_a**	**Error[a]**	**Calcd pK_a**	**Error[a]**
Pyridin-1-ium	*CBS-QB3*	-4.49	9.4	-0.83	9.4	-0.76	12.7	-2.16	12.6	-2.08
	G3B3(MP2)	-8.15	8.8	-0.19	8.7	-0.12	12.0	-1.52	11.9	-1.44
	G4(MP2)	-7.26	8.9	-0.35	8.9	-0.28	12.2	-1.68	12.1	-1.60
	DLPNO-CCSD(T)/cc-pVTZ[b]	-5.00	9.3	-0.74	9.3	-0.67	12.6	-2.08	12.5	-1.99
Piperidin-1-ium	*CBS-QB3*	20.24	16.5	-0.94	16.4	-0.76	19.8	-1.36	19.6	-1.17
	G3B3(MP2)	19.29	16.4	-0.78	16.2	-0.59	19.6	-1.20	19.4	-1.01
	G4(MP2)	19.00	16.3	-0.72	16.1	-0.54	19.5	-1.15	19.4	-0.95
	DLPNO-CCSD(T)/cc-pVTZ[b]	17.70	16.1	-0.50	15.9	-0.32	19.3	-0.92	19.1	-0.73
Morpholin-4-ium	*CBS-QB3*	-5.52	13.9	-0.02	14.1	-0.19	17.1	-0.55	17.3	-0.70
	G3B3(MP2)	-6.38	13.8	0.13	13.9	-0.04	17.0	-0.40	17.1	-0.55
	G4(MP2)	-6.38	13.8	0.13	13.9	-0.04	17.0	-0.40	17.1	-0.55
	DLPNO-CCSD(T)/cc-pVTZ[b]	-4.94	14.0	-0.13	14.2	-0.29	17.2	-0.65	17.4	-0.80

[a]Difference between experimental pK_a and calculated pK_a (Error = expl pK_a – calcd pK_a)

[b]DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP

Table 3.5 Calculated gas phase reaction free energy and pK_a of N-base amine conjugate acids with reference base/acid piperazine-1-carbaldehyde/piperazin-4-ium-1-carbaldehyde in $[C_4C_1im][BF_4]$ and $[C_4C_1im][NTf_2]$.

			$[C_4C_1im][BF_4]$				**$[C_4C_1im][NTf_2]$**			
			*B3LYP/6-31G**		*M05-2X/6-31+G***		*B3LYP/6-31G**		*M05-2X/6-31+G***	
HA	**Level of Theory**	**ΔG^*_{gas} (kJ mol^{-1})**	**Calcd pK_a**	**Error[a]**	**Calcd pK_a**	**Error[a]**	**Calcd pK_a**	**Error[a]**	**Calcd pK_a**	**Error[a]**
Pyridin-1-ium	*CBS-QB3*	19.83	10.3	-1.74	9.9	-1.34	10.6	-0.05	10.2	0.35
	G3B3(MP2)	16.78	9.8	-1.20	9.4	-0.80	10.0	0.48	9.6	0.88
	G4(MP2)	17.76	10.0	-1.38	9.6	-0.97	10.2	0.31	9.8	0.71
	DLPNO-CCSD(T)/cc-pVTZ[b]	18.37	10.1	-1.48	9.7	-1.08	10.3	0.20	9.9	0.60
Piperidin-1-ium	*CBS-QB3*	44.56	17.4	-1.85	16.9	-1.33	17.7	0.75	17.1	1.26
	G3B3(MP2)	44.23	17.4	-1.79	16.9	-1.27	17.6	0.81	17.1	1.32
	G4(MP2)	44.01	17.4	-1.75	16.8	-1.24	17.6	0.85	17.0	1.35
	DLPNO-CCSD(T)/cc-pVTZ[b]	41.08	16.8	-1.24	16.3	-0.72	17.0	1.36	16.5	1.87
Morpholin-4-ium	*CBS-QB3*	18.80	14.8	-0.93	14.7	-0.76	15.0	1.56	14.9	1.73
	G3B3(MP2)	18.55	14.8	-0.89	14.6	-0.72	15.0	1.61	14.8	1.77
	G4(MP2)	18.63	14.8	-0.90	14.6	-0.73	15.0	1.59	14.8	1.76
	DLPNO-CCSD(T)/cc-pVTZ[b]	18.43	14.8	-0.87	14.6	-0.70	15.0	1.63	14.8	1.79

[a]Difference between experimental pK_a and calculated pK_a (Error = expl pK_a – calcd pK_a)

[b]DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP

Solvent type is seen to also affect performance. The $[C_4C_1im][BF_4]$ solvent resulted in lower errors for any combination of the gas and solvent phase levels of theory for the 2-piperazin-4-ium-1-ylethanol reference system. However, the inverse is generally true for the other reference acid, except for morpholin-4-ium. Given that the same solvation model was used for both, the effects must be due to the specific parameters provided for each solvent. Variance in performance would most likely be a result of increased stabilisation. For the systems of interest, this would involve charge stabilisation and hydrogen-bonding ability.

Finally, the PX cycle performs relatively well with an overall error of 1.0 ± 0.6 pK_a units, indicating that predicted values are on average approximately a pK_a unit off the literature values.

Specific attention should be given to the performance of linearly scaling CCSD, as it enables extension of the methodology to much larger systems than are possible with composite methods. DLPNO-CCSD(T)/cc-pVTZ performed reasonably well in terms of both precision and accuracy, being one of the most precise methods and differing from the highly accurate G*n* methods by about 0.02-0.07 pK_a units. Furthermore, accuracy performance in terms of the relative ordering remained consistent across both solvent levels of theory, differing by 0.02 pK_a units from B3LYP/6-31G* to M05-2X/6-31+G**. While precision results did not show the same consistency in ordering, the change between solvent levels was negligible. In terms of benchmarking, this indicates DLPNO-CCSD(T) performs within the range of accuracy and precision of other commonly used and well-established high accuracy methods. Furthermore, DLPNO-CCSD(T) drastically reduces computational time, compared to the other methods (Figure 3.6). The increase in the runtime differences between the composite methods and DLPNO-CCSD(T) with increased system size is significant for larger, more complicated systems such as the acid-functionalised ionic liquids of interest or biopolymer units such as cellobiose. Therefore, although G3B3(MP2) [M05-2X/6-31+G** (SMD)] is the most accurate and precise method overall, the effects of system size and complexity on execution time severely diminishes its applicability. DLPNO-CCSD(T)/cc-pVTZ provides a viable alternative with little decrease in accuracy, specifically in combination with M05-2X/6-31+G**. This validates its use for ionic liquid acidity predictions of more complex systems such as the proceeding acid-functionalised ionic liquid cations.

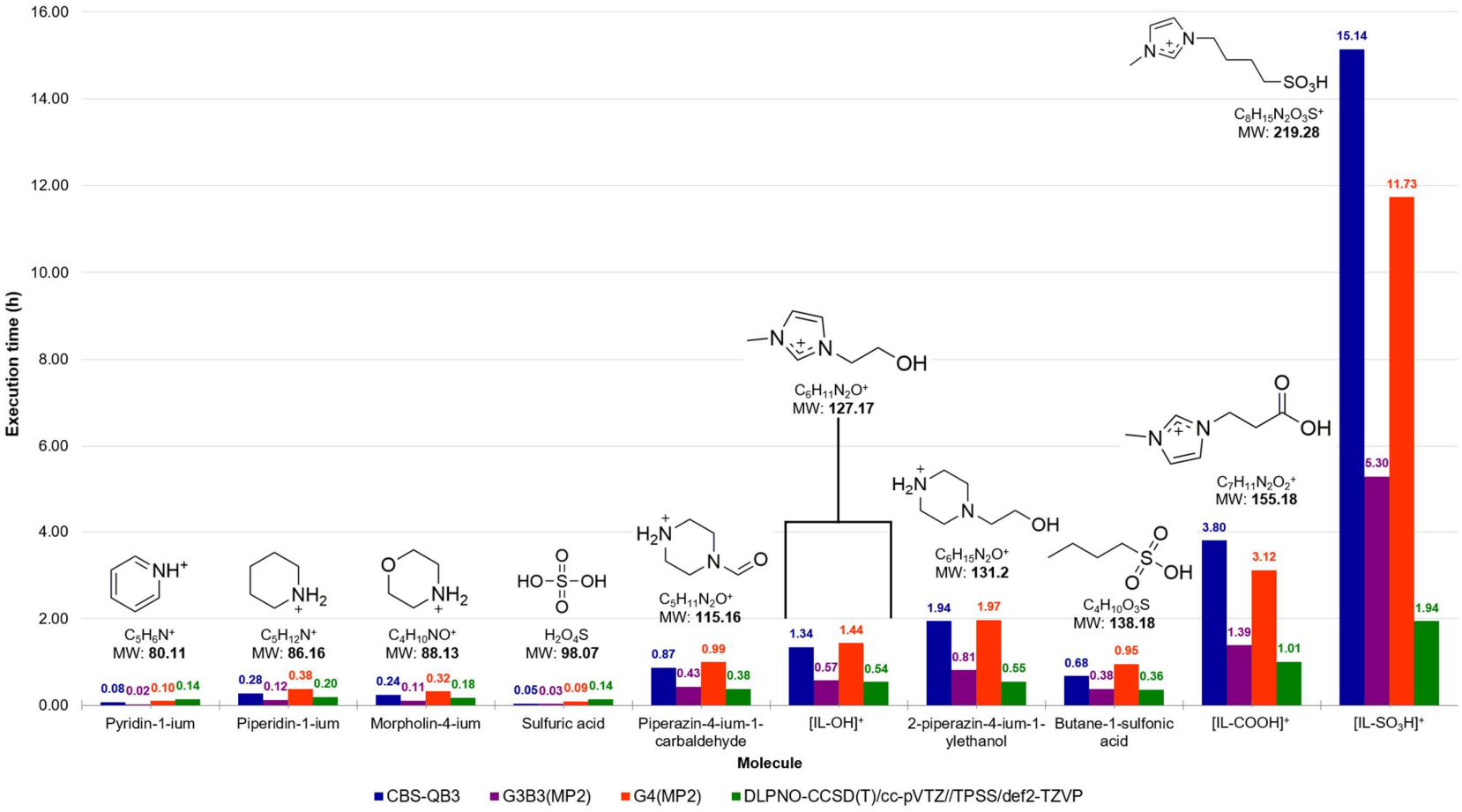


Figure 3.6 Bar chart illustrating the relationship between system size and execution time for a gas phase free energy calculation for each HL gas phase method. *Computations were run across 24 threads on a 12 core Intel Xeon E5-2690 v3 2.6 GHz processor.*

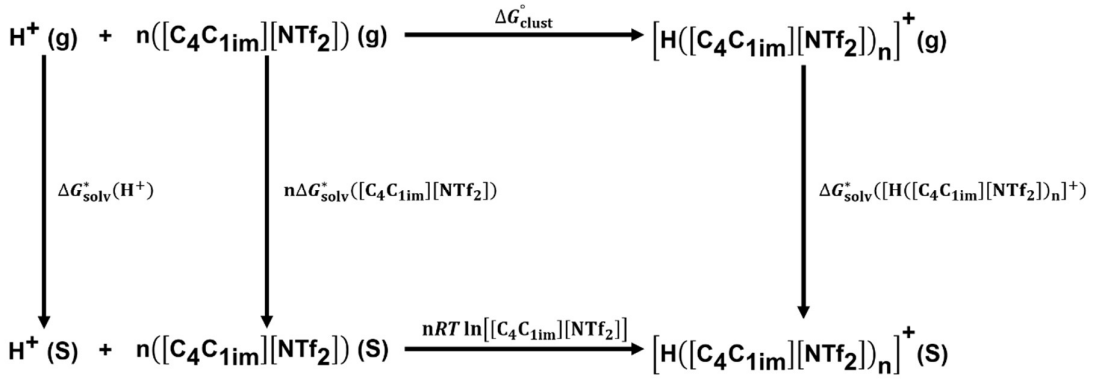
3.3.1.2 Direct Cycle

The direct cycle, as detailed in Section 3.1.1, requires the determination of the solvation free energy of a proton ($\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+)$)and the accuracy of this value is a significant determinant in the overall performance of the cycle.[135] $\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+)$ cannot be calculated directly using implicit solvents, as quantum mechanical calculations require the presence of at least one electron in the molecular system and continuum treatment of charged solutes is prone to errors given the specific interactions often associated with charged species.[147,206] Instead, an intermediary solvent model can be used, where a limited number of explicit molecules are placed in a continuum, generating a cluster-continuum model.[148,158] Thus, the proton is surrounded by explicit solvent molecules, representing the first solvation shell and forming a solute-solvent cluster. This cluster can then be solvated using implicit solvation. This splits the solvation of the proton into two steps, namely the formation of a solute-solvent cluster and the implicit solvation of this solute-solvent cluster. The solute-solvent cluster formation is treated as a chemical reaction and a thermodynamic cycle is used to determine the overall proton solvation free energy.[136] The resultant value can then be combined with a direct cycle that focuses on the protonation reaction, from which the pK_a value is obtained.

As illustrated by Schemes 3.3 and 3.4, this solvation cycle can be further varied based on how the reactant form of the solvent is represented. The reactant form of the solvent cluster can either be a cluster formed of independent non-interacting solvent monomers, i.e. *n*S (monomer cycle, Scheme 3.3), or an interacting cluster of solvent molecules, i.e., S_n (cluster cycle, Scheme 3.4).[136]

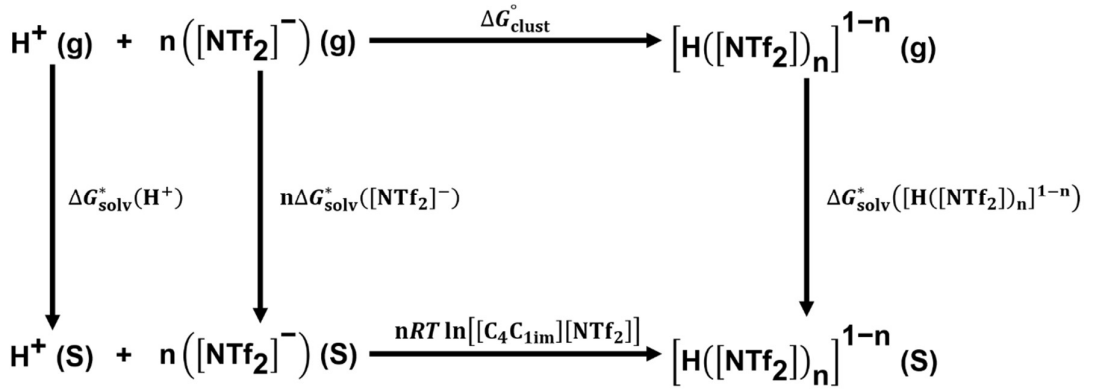
The problem is further complicated by the ionic nature of the solvents of interest. Schemes 3.3 and 3.4 are based on solvation of an ion solute by neutral molecular solvents, such as water. When solvated by an ionic species, ionic interactions between the species can affect the composition of the solvent cluster. Solvent ions of like charge will be repelled, and opposite charge attracted, so the first solvation shell would most likely be composed predominantly of solvent ions of opposite charge to the solute ion. Consequently, for ionic liquids, cluster representation is also varied such that *S* is either only anions/cations or the full anion-cation pairs.[160] In this study, the monomer and cluster cycles are accordingly each split into two more cycles, where the solvent, *S*, is the full ion pair (Scheme 3.6 (i) and (iii)) or only the anion (Scheme 3.6 (ii) and (iv)).

Scheme 3.6 Exemplars of the monomer and cluster thermodynamic cycles used to determine proton solvation, using $[C_4C_1im][NTf_2]$ as the solvent system, where **(i)** and **(ii)** are monomer cycles and **(iii)** and **(iv)** cluster cycles, with S = $[C_4C_1im][NTf_2]$ and S = $[NTf_2]^-$, respectively.



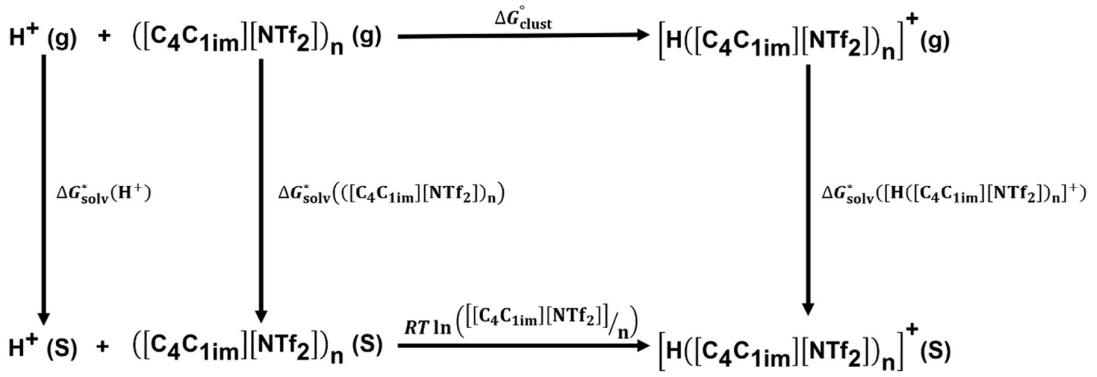
$$\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+) = \Delta G^\circ_{\mathrm{clust}} + \Delta G^*_{\mathrm{solv}}([\mathrm{H}([\mathrm{C_4C_1im}][\mathrm{NTf_2}])_\mathrm{n}]^+) - \mathrm{n}RT\ln\left[[\mathrm{C_4C_1im}][\mathrm{NTf_2}]\right] - \mathrm{n}\Delta G^*_{\mathrm{solv}}([\mathrm{C_4C_1im}][\mathrm{NTf_2}]) - \mathrm{n}\Delta G^{\circ\rightarrow *} \quad \textbf{(3.28)}$$

(i)



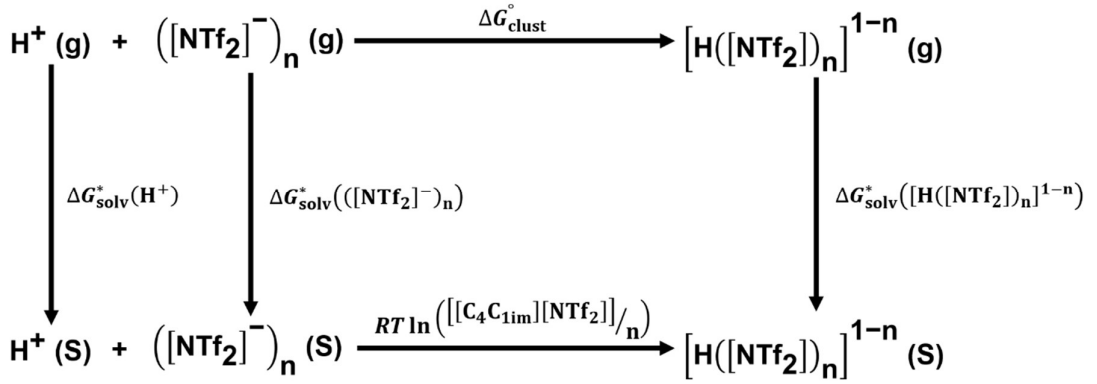
$$\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+) = \Delta G^\circ_{\mathrm{clust}} + \Delta G^*_{\mathrm{solv}}([\mathrm{H}([\mathrm{NTf_2}])_\mathrm{n}]^{1-\mathrm{n}}) - \mathrm{n}RT\ln\left[[\mathrm{C_4C_1im}][\mathrm{NTf_2}]\right] - \mathrm{n}\Delta G^*_{\mathrm{solv}}([\mathrm{NTf_2}]^-) - \mathrm{n}\Delta G^{\circ\rightarrow *} \quad \textbf{(3.29)}$$

(ii)



$$\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+) = \Delta G^\circ_{\mathrm{clust}} + \Delta G^*_{\mathrm{solv}}([\mathrm{H}([\mathrm{C_4C_1im}][\mathrm{NTf_2}])_\mathrm{n}]^+) - RT\ln\left(\left[[\mathrm{C_4C_1im}][\mathrm{NTf_2}]\right]/_\mathrm{n}\right) - \Delta G^*_{\mathrm{solv}}(([\mathrm{C_4C_1im}][\mathrm{NTf_2}])_\mathrm{n}) - \Delta G^{\circ\rightarrow *} \quad \textbf{(3.30)}$$

(iii)



$$\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+) = \Delta G^\circ_{\mathrm{clust}} + \Delta G^*_{\mathrm{solv}}([\mathrm{H}([\mathrm{NTf_2}])_\mathrm{n}]^{1-\mathrm{n}}) - RT\ln\left(\left[[\mathrm{C_4C_1im}][\mathrm{NTf_2}]\right]/_\mathrm{n}\right) - \Delta G^*_{\mathrm{solv}}(([\mathrm{NTf_2}]^-)_\mathrm{n}) - \Delta G^{\circ\rightarrow *} \quad \textbf{(3.31)}$$

(iv)

The total assessment of the direct cycle can thus be divided into two parts: the first is an assessment of proton solvation determined using thermodynamic cycles (Scheme 3.6) and the second of the level of theory effects, like that seen for the PX cycles.

Unlike the larger direct and PX cycle, the cluster cycles were tested at the solvent levels of theory (B3LYP/6-31G* and M05-2X/6-31+G**) for both gas and solvent phase legs of the cycle. While these are lower levels of theory and thus generally assumed to be of lower accuracy, they are more applicable to expanding system sizes. As was detailed in Chapter 2, composite methods consist of a combination of high correlation energy calculations.[82,97] Thus,

while these methods are designed to lower computational cost and increase applicability, this is relative to direct application of high correlation methods (CCSD(T), MP4, CISD(T)). They still exhibit a significant scaling cost with system size and as such are not applicable to large QM systems. In the context of cluster calculations this limits the expansion of the cluster, hindering convergence to an accurate value. Furthermore, maintenance of level of theory between solvent and gas phase legs allows for possible cancellation of errors inherent to the level of theory.

Given that the same level of theory, implicit model and solvent parameters are used for each of the solvation TCs, the accuracy of the proton solvation energy depends solely on the description of the cluster. This is determined by what *S* is (full cation-anion ion pair or anion only), the size of the cluster (*n* = number of solvent molecules) and whether the cluster is treated as a single entity (S_n, cluster cycle) or as individual non-interacting molecules (*n*S, monomer cycle). The results of these procedures, using *n* = 1 or 2, are shown in Table 3.6 below.

Table 3.6 $\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+)$ (kJ mol^{-1}) in [C_4C_1im][NTf_2] computed using monomer and cluster cycles with different numbers of solvent molecules present in the cluster.

		*B3LYP/6-31G**		*M05-2X/6-31+G***	
		1	**2**	**1**	**2**
Monomer Cycle	*Anion only*	-1101	-1097	-1060	-1072
	Full Ion Pair	-1064	-1062	-1017	-1037
Cluster Cycle	*Anion only*	-1101	-	-1060	-
	Full Ion Pair	-1064	-1120	-1017	-1046

Values were calculated using the equations seen in Section 3.1.2 for each cycle. The gas phase free energy (G°) of the proton can be calculated using statistical thermodynamics and the ideal gas (i.e. translations alone) partition function as follows

$$G^\circ = -\mathrm{n}kT\ln\left(\frac{\mathrm{V}}{\mathrm{N}\Lambda^3}\right) \qquad \textbf{(3.32)}$$

$$\text{with } \Lambda^3 = \left(\frac{h^2}{2\pi kmT}\right)^{3/2} \qquad \textbf{(3.33)}$$

where k is Boltzmann's constant, V the volume of an ideal gas, N the number of particles, m the mass of the particle, T the temperature in Kelvin, n the number of moles, and h Plank's constant. Under standard state temperature and pressure conditions, this yields a value of -26.28 kJ mol^{-1} for one mole of H^+ treated as an ideal gas.

Due to the nature of the solvent as a combination of charged ions, complications were encountered that limited application. Cluster cycle anion solvent clusters would not

convergence past n=1. This is due to the use of n=2 clusters necessitating the gas phase optimisation of a two anion only solvent cluster (Figure 3.7 (iii) and (vii)), which is not favourable due to the Coulombic interaction of like charges. The anions repel each other, with the lowest energy conformation existing when they are infinitely far away from each other. Consequently, cluster cycles where S = anion were limited to only n=1 clusters (Figure 3.7 (i) and (v)).

This problem is not encountered for the solute-solvent clusters, which place the proton between the two anions (Figure 3.8 (iii) and (vii)). Furthermore, the proton position is not equidistant between the anions, as it forms a covalent bond with one anion and is hydrogen-bonded to the other (Figure 3.8 (iii) and (vii)). This covalent bonding is seen in all solute-solvent clusters regardless of solvent LoT or cation presence (Figure 3.8).

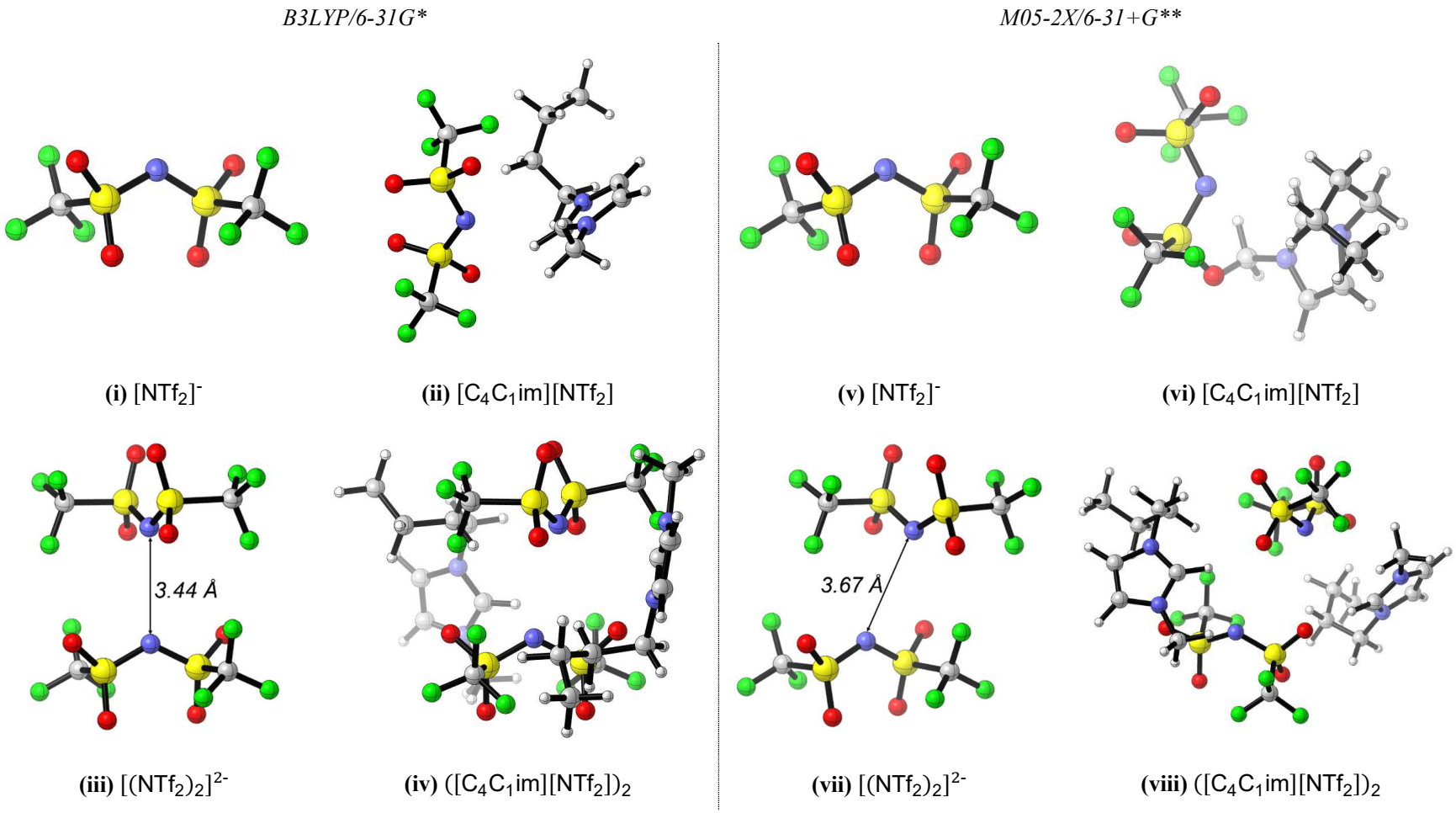
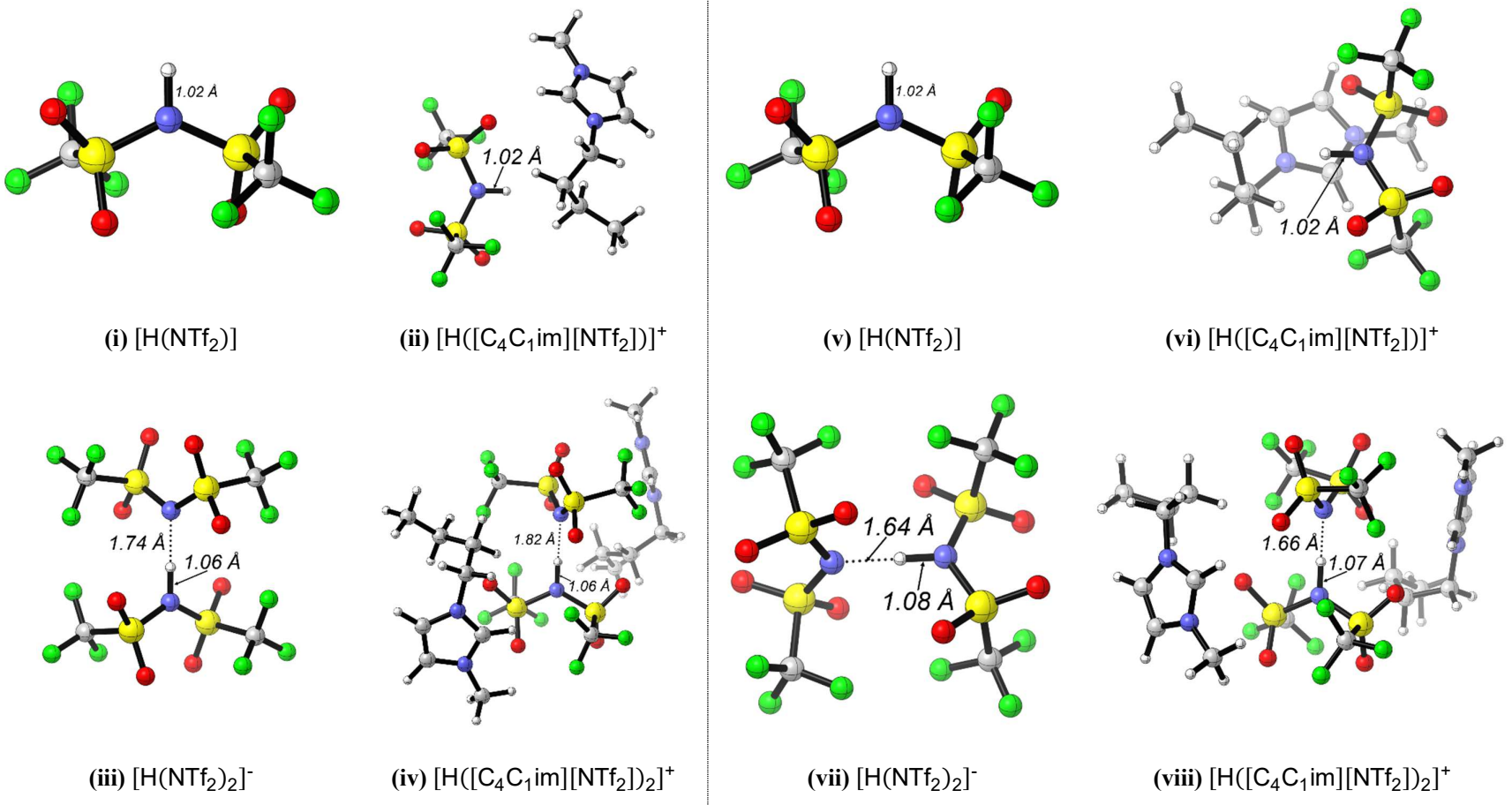


Figure 3.7 Solvent clusters for $[C_4C_1im][NTf_2]$ and $[NTf_2]^-$, optimised using implicit SMD $[C_4C_1im][NTf_2]$ conditions.

*B3LYP/6-31G**

*M05-2X/6-31+G***



(i) $[H(NTf_2)]$ **(ii)** $[H([C_4C_1im][NTf_2])]^+$ **(v)** $[H(NTf_2)]$ **(vi)** $[H([C_4C_1im][NTf_2])]^+$

(iii) $[H(NTf_2)_2]^-$ **(iv)** $[H([C_4C_1im][NTf_2])_2]^+$ **(vii)** $[H(NTf_2)_2]^-$ **(viii)** $[H([C_4C_1im][NTf_2])_2]^+$

Figure 3.8 H^+-solvent clusters for $[C_4C_1im][NTf_2]$ and $[NTf_2]^-$, using implicit SMD $[C_4C_1im][NTf_2]$ conditions.

The full ion pair n=2 B3LYP/6-31G* cluster shows the proton positioning as seen for the anion only clusters, simply surrounded by the cations (Figure 3.8 (iv)). The cations surround the anions and the overall structure does not change significantly when the solute (proton) is included. This links with many explicit ionic liquid solvation studies and NMR observations, which show the anion to be more strongly associated with effective solvation.[36,56,194] IL anions often have stronger hydrogen-bonding capabilities than cations and implying the formation of stronger interactions.[207] Comparison of the solvent and solute-solvent clusters show a root-mean-square deviation (RMSD) between carbon and all heteroatoms (N, O, F, S) of 0.630 Å (Figure 3.9 (i)). This occurs for many of the clusters, such as the n=1 anion clusters (Figure 3.7 (i) and (v) & Figure 3.8 (i) and (v)) and the M05-2X/6-31+G** n=1 full ion pair cluster (Figure 3.7 (vi) & Figure 3.8 (vi)). However, this observation does not hold for all clusters. While n=2 B3LYP obeys the trend, M05-2X shows far higher variation of 1.151 Å (Figure 3.9 (ii)). The inverse is true for the n=1 clusters, with B3LYP varying significantly compared to M05-2X (Figure 3.9 (iii) and (iv)). The N-H^+ covalent bond distance is maintained through the addition of a single cation to the cluster for both levels of theory. However, with a further increase in the number of explicit solvent molecules there is an increase in the N-H^+ bond distance due to the additional anion forming an intermolecular hydrogen bond to the protonated anion. For the n=2 clusters the bond angle is almost linear (Figure 3.8 (iii), (iv), (vii) and (viii)), implying stronger interaction. However, despite the similar angle variation is seen between the LoT, with M05-2X yielding longer N-H covalent bonds. This is probably due to the hydrogen bond distance, which is shorter in the clusters obtained using the M05-2X functional. Indeed, there is a large amount of structural variation between the levels of theory for the n=2 clusters.

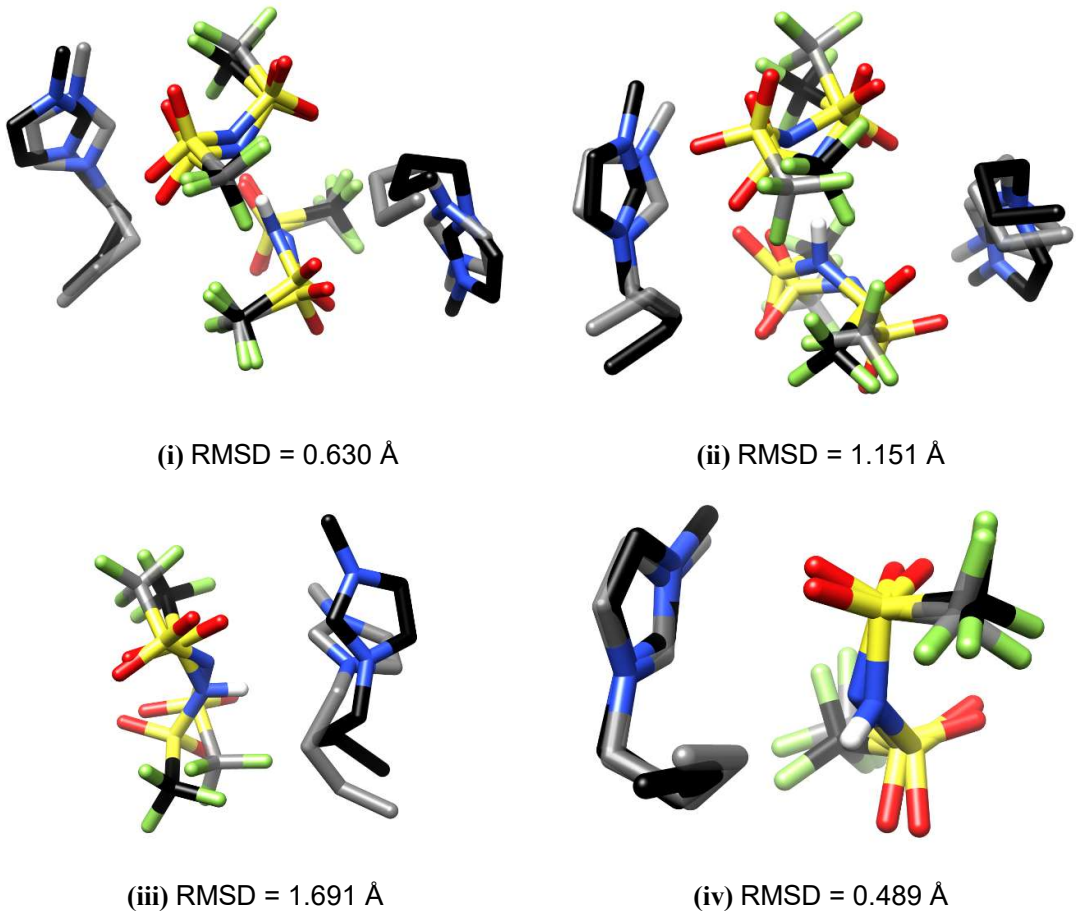


Figure 3.9 $[C_4C_1im][NTf_2]$ clusters superimposed according to carbon and heteroatom variation for *n=2* **(i)** B3LYP/6-31G* and **(ii)** M05-2X/6-31+G**, and *n=1* **(iii)** B3LYP/6-31G* and **(iv)** M05-2X/6-31+G** clusters. *(Structures rendered semi-transparent with carbon atoms coloured grey are solvent clusters, with opaque black carbon structures corresponding to H^+-solvent clusters.)*

$[C_4C_1im][BF_4]$, also used in the PX cycle tests, had to be excluded due to complications encountered with optimisations. All attempts to optimize this system led to a saddle point (single imaginary frequency) in which the anion dissociated into neutral HF and BF_3 molecules (Figure 3.10). This bond dissociation will include an anomalous contribution to the solvation free energy of the proton, making the cluster approach difficult to apply in this case and not directly comparable to the $[C_4C_1im][NTf_2]$ solvent.

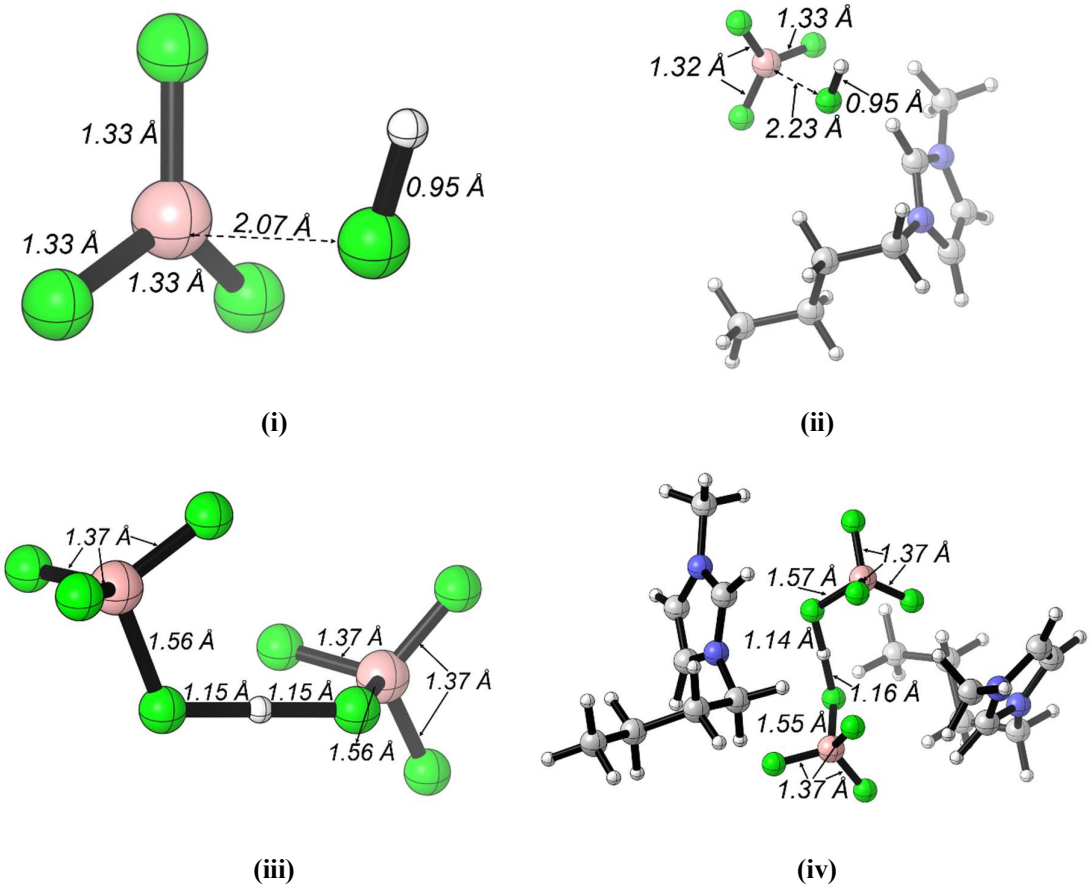


Figure 3.10 B3LYP/6-31G* $[C_4C_1im][BF_4]$ solute-solvent clusters with *n=1* **(i)** anion and **(ii)** full ion pair solvent molecules, and *n =2* **(iii)** anion and **(iv)** full ion pair solvent molecules.

Trends can be observed in the solvation results that fit simple charge arguments. Anion only solvent clusters produce lower (more negative) $\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+)$ values and, inclusion of cations in the first solvation shell consistently produces higher energy values. The proton is positively charged favouring Coulombic interactions with negatively charged anions, and strongly repels like charged ions. The existence of a cation close to the proton would therefore destabilise the system, leading to higher $\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+)$ values. This mimics observations seen in explicit IL solvation studies whereby the first solvation shell is predominantly made up of anions.[160]

Increasing the number of ion pairs causes the solvation free energy to become more negative. The additional anion would more effectively shield the proton from the cations and the stabilisation noted links to a broader trend where solvation free energy lowers as the number of anions increases. All n = 2 $\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+)$ are lower than their respective n = 1 counterparts, except for the B3LYP monomer cycle where the inverse trend is noted. The difference between these values is, however, very small (2-4 kJ mol^{-1}) compared to the large differences seen for

all the other clusters (30 kJ mol^{-1}). In the pioneering works by Bryantsev et al. and Pliego and Riveros, the optimal cluster size was determined by increasing the number of solvent molecules (n) until the solvation free energy converges.[136,148] This was not attempted in this case due to a variety of reasons. First regards the computational costs associated with the larger clusters, specifically regarding the full ion pair, with each additional pair adding 40 atoms to the system. Second, given the reasons stated above, comparisons would only be possible for cluster and monomer cycles where full (neutral) ion pairs are used. Increasing the cluster size when only anions are present, ensures no solvent cluster can be optimised to convergence. Furthermore, given the relative size of the anion compared to the proton, steric hinderance may become problematic when increasing the number of solvent molecules. Illustrated in the Figure 3.8 (iii) and (vii), two $[NTf_2]^-$ ions almost completely surround the proton, leading to the conclusion that more anions would at best interact very weakly with the system if not destabilise it.

Additionally, other studies have shown the solvent cluster for monovalent ions to generally consist of between two to four (neutral) solvent molecules.[162] Specifically, a study by Xue et al. determined a minimum proton solvation free energy in $[C_4C_1im][NTf_2]$ of -1079.5 kJ mol^{-1} with a two anion system.[168] Used in conjunction with a direct cycle, extremely accurate pK_a predictions were obtained (mean unsigned error of 0.3-0.5 pK_a units) for numerous benzoic acids and benzenethiols.[168] Methodology details indicate the use of a monomer cycle and M06-2X/6-31+G** for all steps. Many of the values predicted in this study are higher in energy by ~10 kJ mol^{-1}, with the closest result given by the two anion monomer obtained using the M05-2X/6-31+G** calculation method. This lends credence to the methodology used to calculate $\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+)$, as values generated match literature values calculated in a similar manner, shown to be highly accurate.

Another way to estimate the solvation free energy is by making use of the experimental pK_a values. If the gas and solvent phase free energy values calculated at the various levels of theory are assumed to be within experimental error of the true values, a "true/experimental" $\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+)$ can be approximated using a thermodynamic cycle in which $\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+)$ is now the only missing quantity. Given the accuracy and precision observed in the PX cycle results, this assumption seems founded. Equation 3.4 (Section 3.1.1) can therefore manipulated as follows to allow for the calculation of these "true/experimental" $\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+)$ values:

$$\mathrm{p}K_a = \frac{\Delta G^*_{\mathrm{soln}}}{RT\ln(10)} \qquad \textbf{(3.4)}$$

$$\Delta G^*_{\mathrm{gas}} + \Delta\Delta G^*_{\mathrm{solv}} = \mathrm{p}K_{\mathrm{a}} RT \ln(10)$$

$$G^*_{\mathrm{gas}}(\mathrm{N}) + G^*_{\mathrm{gas}}(\mathrm{H}^+) - G^*_{\mathrm{gas}}(\mathrm{HN}^+) + \Delta G^*_{\mathrm{solv}}(\mathrm{N}) + \Delta G^*_{\mathrm{solv}}(\mathrm{H}^+) - \Delta G^*_{\mathrm{solv}}(\mathrm{HN}^+) = \mathrm{p}K_{\mathrm{a}} RT \ln(10)$$

$$\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+) = \mathrm{p}K_{\mathrm{a}} RT \ln(10) + G^*_{\mathrm{gas}}(\mathrm{HN}^+) - G^*_{\mathrm{gas}}(\mathrm{N}) - G^*_{\mathrm{gas}}(\mathrm{H}^+) + \Delta G^*_{\mathrm{solv}}(\mathrm{HN}^+) - \Delta G^*_{\mathrm{solv}}(\mathrm{N}) \quad \textbf{(3.34)}$$

These "experimental" $\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+)$ values, shown in Table 3.7 below, are averages across all the N-base amine conjugate acids (HA, Figure 3.4) and reference acids/bases (HRef, Figure 3.5) mentioned thus far.

Table 3.7 Average $\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+)$ (kJ mol^{-1}) in [C_4C_1im][NTf_2] calculated from computed free energy values according to Equation 3.34.

Level of Theory	*B3LYP/6-31G**	*M05-2X/6-31+G***
CBS-QB3	-1054	-1053
G3B3(MP2)	-1056	-1055
G4(MP2)	-1056	-1055
DLPNO-CCSD(T)/cc-pVTZ[a]	-1066	-1065

[a]DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP

The average across all levels of theory, gas and solvent phase, is approximately -1058 kJ mol^{-1} with 5 kJ mol^{-1} variation, about 21 kJ mol^{-1} higher than Xue et al.'s predicted value. Only slight variation is seen between solvent methods, with B3LYP/6-31G* producing slightly lower values overall. CBS-QB3 is higher in energy than the other composite methods by only a small margin (2 kJ mol^{-1}). DLPNO-CCSD(T) on the other hand, is far lower than all the other methods, by ~ 10 kJ mol^{-1}. However, it should be mentioned that this method was chosen for its computational efficiency, with composite methods used as a comparison to assess its relative accuracy and precision. To this end, DLPNO-CCSD(T) was not applied with a large basis set equivalent to those used by composite methods (or extrapolated to the complete basis set limit) and only approximates the correlation limit with a truncated basis set. Furthermore, optimisation and frequency calculations are built into each composite method with appropriate scaling, along with empirically determined error corrections whereas the DLPNO-CCSD(T) calculations made use of TPSS/def2-TZVP optimised geometries and thermal corrections. Thus, substantial variation could be expected between the composite methods and DLPNO-CCSD(T), with greater significance to be attributed to the composite method predicted average $\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+)$ values.

Comparing these values to Table 3.6, the best prediction of the proton solvation free energy would be closest to the composite method average of -1055 ± 1 kJ mol^{-1}. The closest

approximate is -1060 kJ mol^{-1} calculated with a one (n =1) anion ($[NTf_2]^-$) cluster at M05-2X/6-31+G**, regardless of cycle.

Summary tables were created to better assess the effects of each aspect of the cycle (LoT, solvent molecule representation, number of solvent molecules, etc.) individually as well as in combination with other aspects (Tables 3.8-3.10). The tables are set up in the same manner, with a row total (light shading), column total (light) and cumulative overall cycle total (bold, darker shading). The averages are, however, over five acids, as values for the reference acids (Figure 3.5) of the PX cycle were included along with the original model system N-base amine conjugate acids (Figure 3.4). The explicit individual error results for each of these acids according to each gas and solvent phase LoT are shown in Table 3.9.

Table 3.8 Mean absolute error (MAE), standard deviation (SD) and minimum/maximum errors (MIN/MAX) for pK_a prediction of N-base amine conjugate acids (Figure 3.4) and reference acids (Figure 3.5) using a direct (D) cycle at the various level of theory (LoT) combinations.

Phase LoT	*B3LYP/6-31G**				*M05-2X/6-31+G***				***Total***			
	MAE	**SD**	**MIN/MAX**		**MAE**	**SD**	**MIN/MAX**		**MAE**	**SD**	**MIN/MAX**	
CBS-QB3	5.42	3.90	0.16	12.52	3.33	2.51	0.19	8.01	4.38	3.42	0.16	12.52
G3B3(MP2)	5.06	3.78	0.01	11.99	3.52	2.68	0.38	8.68	4.29	3.34	0.01	11.99
G4(MP2)	5.01	3.77	0.03	11.95	3.53	2.71	0.41	8.72	4.27	3.34	0.03	11.95
[b]DLPNO-CCSD(T)[a]	4.07	3.02	0.01	10.18	4.38	3.42	0.07	10.24	4.23	3.21	0.01	10.24
Total	4.89	3.63	0.01	12.52	3.69	2.85	0.07	10.24	**4.29**	**3.31**	**0.01**	**12.52**

[b]DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP

Table 3.9 Direct cycle calculated gas phase reaction free energy (ΔG_{gas}) and pK_a error[a] for all N-base amine conjugate acids and reference systems in $[C_4C_1im][NTf_2]$ using cluster (CC) and monomer cycle (MC) generated $\Delta G^*_{soly}(H^+)$ values.

			*B3LYP/6-31G**								*M05-2X/6-31+G***							
			1				***2***				***1***				***2***			
			Anion only		*Full Ion Pair*		*Anion only*		*Full Ion Pair*		*Anion only*		*Full Ion Pair*		*Anion only*		*Full Ion Pair*	
HA	**Gas Phase LoT[b]**	**ΔG_{gas} (kJ mol^{-1})**	**MC[c]**	**CC[d]**	**MC[c]**	**CC[d]**	**MC[c]**	**CC[d]**	**MC[c]**	**CC[d]**	**MC[c]**	**CC[d]**	**MC[c]**	**CC[d]**	**MC[c]**	**CC[d]**	**MC[c]**	**CC[d]**
Pyridin-1-ium	*CBS-QB3*	895.13	7.6	7.6	1.2	1.2	6.9	-	0.7	10.9	0.7	0.7	-6.9	-6.9	2.9	-	-3.3	-1.7
	G3B3(MP2)	895.34	7.6	7.6	1.2	1.2	6.9	-	0.7	10.9	0.7	0.7	-6.9	-6.9	2.8	-	-3.3	-1.8
	G4(MP2)	896.47	7.4	7.4	1.0	1.0	6.7	-	0.5	10.7	0.5	0.5	-7.1	-7.1	2.6	-	-3.5	-2.0
	DLPNO-CCSD(T)[e]	907.39	5.5	5.5	-0.9	-0.9	4.7	-	-1.4	8.8	-1.4	-1.4	-9.0	-9.0	0.7	-	-5.4	-3.9
Piperidin-1-ium	*CBS-QB3*	919.87	8.4	8.4	2.0	2.0	7.7	-	1.5	11.7	1.6	1.6	-6.0	-6.0	3.8	-	-2.4	-0.8
	G3B3(MP2)	922.79	7.9	7.9	1.5	1.5	7.2	-	1.0	11.2	1.1	1.1	-6.5	-6.5	3.3	-	-2.9	-1.3
	G4(MP2)	922.73	7.9	7.9	1.5	1.5	7.2	-	1.0	11.2	1.1	1.1	-6.5	-6.5	3.3	-	-2.9	-1.3
	DLPNO-CCSD(T)[e]	930.10	6.6	6.6	0.2	0.2	5.9	-	-0.3	9.9	-0.1	-0.1	-7.8	-7.8	2.0	-	-4.2	-2.6
Morpholin-4-ium	*CBS-QB3*	894.11	9.2	9.2	2.8	2.8	8.5	-	2.3	12.5	2.1	2.1	-5.5	-5.5	4.3	-	-1.9	-0.3
	G3B3(MP2)	897.12	8.7	8.7	2.3	2.3	8.0	-	1.8	12.0	1.6	1.6	-6.0	-6.0	3.7	-	-2.4	-0.9
	G4(MP2)	897.35	8.7	8.7	2.2	2.2	7.9	-	1.8	11.9	1.6	1.6	-6.1	-6.1	3.7	-	-2.5	-0.9
	DLPNO-CCSD(T)[e]	907.45	6.9	6.9	0.5	0.5	6.2	-	0.0	10.2	-0.2	-0.2	-7.8	-7.8	1.9	-	-4.2	-2.7

[a]Difference between experimental pK_a and calculated pK_a (Error = expl pK_a – calcd pK_a)
[b]Level of theory
[c]Monomer cycle
[d]Cluster cycle
[e]DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP

Table 3.9 continued.

			*B3LYP/6-31G**								*M05-2X/6-31+G***							
			1				***2***				***1***				***2***			
			Anion only		*Full Ion Pair*		*Anion only*		*Full Ion Pair*		*Anion only*		*Full Ion Pair*		*Anion only*		*Full Ion Pair*	
HA	**Gas Phase LoT[b]**	**ΔG_{gas} (kJ mol^{-1})**	**MC[c]**	**CC[d]**	**MC[c]**	**CC[d]**	**MC[c]**	**CC[d]**	**MC[c]**	**CC[d]**	**MC[c]**	**CC[d]**	**MC[c]**	**CC[d]**	**MC[c]**	**CC[d]**	**MC[c]**	**CC[d]**
2-piperazin-4-ium-1-ylethanol	*CBS-QB3*	899.63	6.6	6.6	0.2	0.2	5.9	-	-0.3	9.9	-0.4	-0.4	-8.0	-8.0	1.8	-	-4.4	-2.8
	G3B3(MP2)	903.50	5.9	5.9	-0.5	-0.5	5.2	-	-1.0	9.2	-1.1	-1.1	-8.7	-8.7	1.1	-	-5.1	-3.5
	G4(MP2)	903.73	5.9	5.9	-0.6	-0.6	5.1	-	-1.0	9.1	-1.1	-1.1	-8.7	-8.7	1.0	-	-5.1	-3.6
	DLPNO-CCSD(T)[e]	912.39	4.3	4.3	-2.1	-2.1	3.6	-	-2.6	7.6	-2.6	-2.6	-10.2	-10.2	-0.5	-	-6.6	-5.1
piperazin-4-ium-1-carbaldehyde	*CBS-QB3*	875.31	7.5	7.5	1.1	1.1	6.7	-	0.6	10.8	0.2	0.2	-7.4	-7.4	2.3	-	-3.8	-2.3
	G3B3(MP2)	878.57	6.9	6.9	0.5	0.5	6.2	-	0.0	10.2	-0.4	-0.4	-8.0	-8.0	1.8	-	-4.4	-2.8
	G4(MP2)	878.71	6.9	6.9	0.5	0.5	6.1	-	0.0	10.2	-0.4	-0.4	-8.0	-8.0	1.7	-	-4.4	-2.9
	DLPNO-CCSD(T)[e]	889.02	5.1	5.1	-1.4	-1.4	4.3	-	-1.8	8.3	-2.2	-2.2	-9.8	-9.8	-0.1	-	-6.2	-4.7

[b]Level of theory
[c]Monomer cycle
[d]Cluster cycle
[e]DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP

Table 3.8 shows the effect of each level of theory, allowing for the comparative accuracy of DLPNO-CCSD(T)/cc-pVTZ and the most accurate (lowest MAE) and most precise (lowest SD) LoT protocol to be determined.

DLPNO-CCSD(T) is the most accurate as well as precise gas phase method, producing the lowest total MAE (4.23 pK_a total MAE) and SD (3.21 pK_a total SD). For precision, it far outperformed the composite methods, being 0.13 pK_a units lower than the Gn methods and 0.21 pK_a units lower than CBS-QB3. CBS-QB3 is thus the least accurate and precise gas phase LoT overall.

In terms of accuracy, the solvent phase LoT choice had a significant impact on overall performance. M05-2X/6-31+G** far outperformed B3LYP/6-31G*, which produced MAEs 1.2 pK_a units larger than that of M05-2X, with SD values differing by about 1 pK_a unit. While B3LYP by and large was the least accurate, it performed significantly better in combination with DLPNO-CCSD(T); however, this greatly affected the other protocols, causing increases of ~ 1.5-2 pK_a units in each. This is very noticeable in the case of CBS-QB3, which performed the best in combination with M05-2X, but the worst with B3LYP, skewing its overall accuracy. Consequently, the overall lowest error, of 3.3 ± 2.5 pK_a units, was in fact produced by CBS-QB3 [B3LYP/6-31G* (SMD)], despite being the lowest performing gas phase LoT overall.

The direct cycle's overall performance using cluster-calculated proton solvation values is poor, with a total error of 4.3 ± 3.3 pK_a units. Errors could be a result of the various levels of theory used; however, given the accuracy and precision noted for the proton exchange cycle this is unlikely. The increased error is thus a result of the proton solvation techniques. To this end, results were summarised to indicate the effect of each aspect of the proton solvation process. Table 3.10 summarises results according to the specific thermodynamic cycles displayed in Scheme 3.6.

Table 3.10 Mean absolute error (MAE), standard deviation (SD) and minimum/maximum errors (MIN/MAX) for pK_a prediction of N-base amine conjugate acids and reference acids using direct (D) cycles according to $\Delta G^*_{solv}(H^+)$ determination conditions.

		Monomer Cycle				*Cluster Cycle*				***Total***			
S	*n*	**MAE**	**SD**	**MIN/MAX**		**MAE**	**SD**	**MIN/MAX**		**MAE**	**SD**	**MIN/MAX**	
Anion only	*1*	4.06	3.21	0.14	9.22	4.06	3.21	0.14	9.22	4.06	3.19	0.14	9.22
	2	4.31	2.40	0.07	8.51	-	-	-	-	-	-	-	-
Full Ion Pair	*1*	4.38	3.39	0.16	10.24	4.38	3.39	0.16	10.24	4.38	3.37	0.16	10.24
	2	2.48	1.82	0.01	6.64	6.37	4.23	0.34	12.52	4.43	3.78	0.01	12.52
Total		3.81	2.86	0.01	10.24	4.94	3.75	0.14	12.52	**4.29**	**3.31**	**0.01**	**12.52**

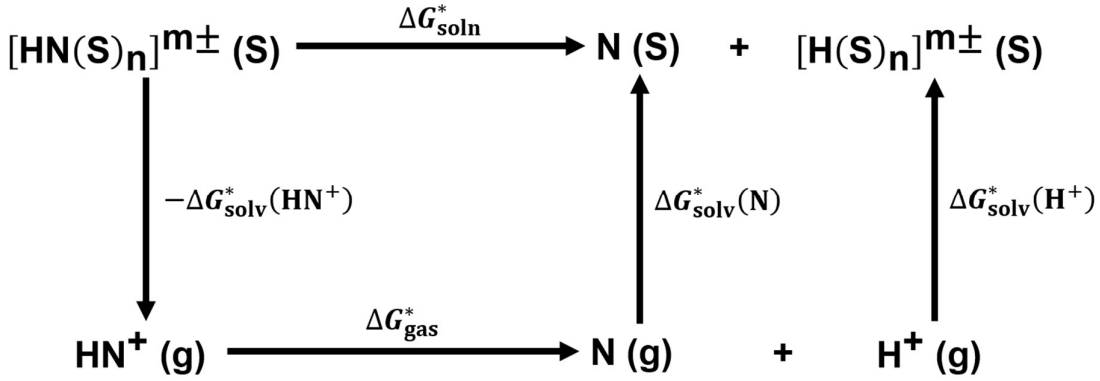
Greater errors in both accuracy and precision are seen for the cluster cycle, regardless of how the solvent cluster is composed. This correlates with conclusions found for proton solvation literature and "experimental" comparisons which favoured the monomer cycle. Missing data points are due to the n=2 solvent cluster optimisation complications encountered and thus make comparisons of solvent cluster composition between the monomer and cluster cycles difficult. This is circumvented by averaging across both cycles (monomer and cluster), such that the effect of solvent cluster composition (S and n) can be assessed (Table 3.11). These results contradict Xue et al.'s results, showing that the best performance is achieved using one anion solvent clusters (n=1, S = $[NTf_2]^-$).[168] Overall, the results mainly show great variation generally caused by cycle choice, leading to n=2 full ion pair clusters being both the best (monomer) and worst (cluster) results. One can see the composition of the solvent cluster has far less of an impact than the cycle choice, varying by less than 0.50 pK_a units compared to the 1.13 seen between cycles.

Focusing specifically on solvent cluster composition, regardless of cycle choice, shows some minor variation (Table 3.11). How the solvent is represented seems to affect accuracy far more than cluster size (*n*), with accuracy differing by only 0.17 pK_a units compared to solvent representation's 0.26 pK_a units. Precision follows the same trend and is once again affected more by cluster representation, with anion cluster SDs consistently being lower.

Table 3.11 Mean absolute error (MAE), standard deviation (SD) and minimum/maximum errors (MIN/MAX) for pK_a prediction of N-base amine conjugate acids and reference acids using direct (D) cycles according to solvent cluster (nS or S_n) composition.

	1				*2*				***Total***			
	MAE	**SD**	**MIN/MAX**		**MAE**	**SD**	**MIN/MAX**		**MAE**	**SD**	**MIN/MAX**	
Anion only	4.06	3.19	0.14	9.22	4.31	2.40	0.07	8.51	4.14	2.94	0.07	9.22
Full Ion Pair	4.38	3.37	0.16	10.24	4.43	3.78	0.01	12.52	4.40	3.57	0.01	12.52
Total	4.22	3.27	0.14	10.24	4.39	3.38	0.01	12.52	**4.29**	**3.31**	**0.01**	**12.52**

Regardless, error values are high relative the PX cycle and from Tables 3.8-3.11 no major improvement seems possible through alteration of the current proton solvation conditions. Furthermore, these high errors occur despite the similarity of $\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+)$ values generated during this study to available literature values and use of high accuracy methods proven to be effective with other cycles and in literature.[142,208–210] This indicates a larger problem with the direct cycle, aside from the approximation of $\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+)$. This could also be a result of the inherent errors that occur when implicitly solvating charged solutes, an effect that would affect not only the proton but also the charged conjugate acid systems used. Thus, further calculations were done using pyridin-1-ium (Table 3.12), whereby all charged components are solvated using a cluster-continuum method (Scheme 3.7).



Scheme 3.7 Direct cycle for a generic N-base amine conjugate ($^+$NH), with all charged reactants and products solvated using cluster-continuum solvation thermodynamic cycles (monomer or cluster).

Table 3.12 Direct cycle calculated pK_a variation[a] from experimental for pyridine-4-ium in [C_4C_1im][NTf_2] using cluster (CC) and monomer cycle (MC) generated $\Delta G^*_{solv}(H^+)$ and ΔG^*_{solv}(pyridine-1-ium) values.

	*B3LYP/6-31G**								*M05-2X/6-31+G***							
	1				***2***				***1***				***2***			
	Anion only		*Full Ion Pair*		*Anion only*		*Full Ion Pair*		*Anion only*		*Full Ion Pair*		*Anion only*		*Full Ion Pair*	
Gas LoT	**MC[b]**	**CC[c]**	**MC[b]**	**CC[c]**	**MC[b]**	**CC[c]**	**MC[b]**	**CC[c]**	**MC[b]**	**CC[c]**	**MC[b]**	**CC[c]**	**MC[b]**	**CC[c]**	**MC[b]**	**CC[c]**
CBS-QB3[d]	-167.1	-167.9	-170.3	-171.6	-165.5	-	-165.8	-160.8	-162.6	-162.6	-169.2	-168.6	-161.8	-	-161.1	-159.8
G3B3(MP2)[e]	-145.1	-145.9	-148.3	-149.6	-143.5	-	-143.8	-138.8	-140.6	-140.6	-147.2	-146.6	-139.8	-	-139.1	-137.8
G4(MP2)	-134.9	-135.7	-138.1	-139.4	-133.3	-	-133.6	-128.6	-130.4	-130.4	-137.1	-136.4	-129.6	-	-129.0	-127.6
DLPNO-CCSD(T)[d]	-208.5	-209.3	-211.7	-213.0	-206.9	-	-207.2	-202.2	-204.0	-204.0	-210.6	-210.0	-203.2	-	-202.5	-201.2

[a]Difference between experimental pK_a and calculated pK_a (Error = expl pK_a – calcd pK_a)
[b]Monomer cycle
[c]Cluster cycle
[d]DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP

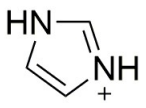
As is evident from the values noted in Table 3.12, clustering of all charged components in the reaction greatly decreases the accuracy of the cycle, producing errors that are magnitudes higher than those noted for the proton only cluster direct cycle (Table 3.9).

From these results (Table 3.12), as well as those for clustering only the proton (Table 3.9), one can assume that large errors are introduced during the clustering process. This could be due to poor representation of the intermolecular forces by a single solvent shell and implicit bulk solvent, or incorrect placement of solvent molecules leading to results further from what is seen in actual solvation. Nonetheless, comparatively the direct cycle performs far worse than the proton exchange cycle. Hence, its use for any predictive experiments is unadvisable, without first undergoing further investigation into cluster generation perhaps using fully explicit methodology or larger sampling.

3.3.2 The Acidity of Brønsted-functionalised Acidic Ionic Liquid Cations

As has been previously stated, no experimental pK_a data is available for the acidic ionic liquids of interest in an ionic liquid solvent (Figure 3.1) and hence the choice of methodology should follow from the benchmarking results discussed above. As a predictive tool, less reliance on empirical experimental results would be preferable and would favour the direct cycle. However, direct cycle tests performed in this work deemed it unsuitable. The PX cycle, on the other hand, produced far better results overall, with an overall average error of 0.96 ± 0.56 pK_a units. This is almost five times smaller than the direct cycle overall error of 4.29 ± 3.31 pK_a units. It was thus decided to predict the pK_a of the acidic ionic liquids using the proton exchange cycle.

Limited experimental data for pK_a acidity in ionic liquids, meant a reference system for the acidic ionic liquids could only be found in a [C_4C_1im][NTf_2] solvent system (Figure 3.11). This limited pK_a acidity predictions to the singular solvent system of [C_4C_1im][NTf_2].



12.1

Figure 3.11 Structure and experimental pK_a in [C_4C_1im][NTf_2] of acidic ionic liquid proton exchange reference acid, 1H-imidazol-3-ium.

Furthermore, butane-1-sulfonic acid and sulfuric acid (Figure 3.12) are also assessed to provide non-ionic comparisons to the cations and allow for the conceptualisation of the ionic liquids' acidity relative to acids of known strength. Butane-1-sulfonic acid is the molecular equivalent to the AIL cation sulfonyl side-chain functionalisation and sulfuric acid is a strong acid that is commonly used as a cellulose hydrolysis catalyst.[41,211–213]

(i) **(ii)**

Figure 3.12 Structures of non-ionic or neutral common acids **(i)** butane-1-sulfonic acid and **(ii)** sulfuric acid.

The results for all three AILs and the two molecular acids are shown in Table 3.13. The composite methods proved applicable (in terms of computational time) to the AILs and thus are included as comparative references to DLPNO-CCSD(T) predictions. However, their comparative time costs increased significantly compared to the N-base amine conjugate acids (Figure 3.6), serving to further indicate poor applicability to cellobiose protonation energy predictions using acidic ionic liquids.

Table 3.13 PX cycle calculated and relative pK_as, gas (ΔG^*_{gas}) and solvent (ΔG^*_{soln}) phase reaction free energies of neutral and cationic acids (Figures 3.1 & 3.12) with reference base/acid 1H-imidazole/1H-imidazol-3-ium (Figure 3.11) in [C_4C_1im][NTf_2].

				*M05-2X/6-31+G***	
HA	**Gas Phase Level of Theory**	**ΔG^*_{gas} (kJ mol^{-1})**	**ΔG^*_{soln} (kJ mol^{-1})**	**Calcd pK_a**	**Relative pK_a[a]**
Butane-1-sulfonic acid	*DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP*	406.55	-2.0	11.8	3.6
	CBS-QB3	383.41	-25.1	7.7	3.8
	G3B3(MP2)	388.10	-20.4	8.5	3.9
	G4(MP2)	388.67	-19.9	8.6	3.8
[IL-OH]$^+$	*DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP*	269.85	139.4	36.5	28.4
	CBS-QB3	250.35	119.9	33.1	29.2
	G3B3(MP2)	255.61	125.2	34.0	29.4
	G4(MP2)	212.64	82.2	26.5	21.7
[IL-COOH]$^+$	*DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP*	162.15	46.3	20.2	12.0
	CBS-QB3	148.03	32.2	17.7	13.8
	G3B3(MP2)	143.86	28.0	17.0	12.4
	G4(MP2)	144.37	28.5	17.1	12.3
[IL-SO_3H]$^+$	*DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP*	109.40	12.3	14.2	6.1
	CBS-QB3	52.35	-44.8	4.3	0.3
	G3B3(MP2)	96.02	-1.1	11.9	7.3
	G4(MP2)	95.17	-2.0	11.8	7.0
Sulfuric acid	*DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP*	384.05	-22.5	8.2	-
	CBS-QB3	359.96	-46.6	3.9	-
	G3B3(MP2)	363.91	-42.6	4.6	-
	G4(MP2)	406.55	-2.0	11.8	-

[a]pK_a relative to sulfuric acid (pK_a(HA) – pK_a(sulfuric acid))

Overall, the acidity trend shows sulfuric acid to be the most acidic overall, as would be expected. This is followed by butane-1-sulfonic acid, the sulfonyl-functionalised AIL ([IL-SO_3H]$^+$, Figure 3.1 (i)), the carbonyl-functionalised AIL ([IL-COOH]$^+$, Figure 3.1 (iii)) and finally the hydroxyl-functionalised cationic acid ([IL-OH]$^+$, Figure 3.1 (ii)). The sulfonyl-functionalised ionic liquid is far more acidic than its cationic counterparts, being closer in pK_a value to the traditionally stronger neutral molecular acids. However, the results still show [IL-SO_3H]$^+$ to be a far weaker acid than sulfuric acid. This general acidity trend correlates well with trends in other acidity measurements (Figure 3.13), like Hammett acidity functions defined as

$$H_0 = \mathrm{p}K(\mathrm{I})_{\mathrm{aq}} + \log\left(\frac{[\mathrm{I}]_{\mathrm{S}}}{[\mathrm{HI}^+]_{\mathrm{S}}}\right) \qquad \textbf{(3.35)}$$

where $\mathrm{p}K(\mathrm{I})_{\mathrm{aq}}$ is the aqueous pK_a of the uncharged indicator base and $[\mathrm{I}]_{\mathrm{S}}$ and $[\mathrm{HI}^+]_{\mathrm{S}}$ the molar concentrations the indicator/protonated indicator in a solvent, S.[132]

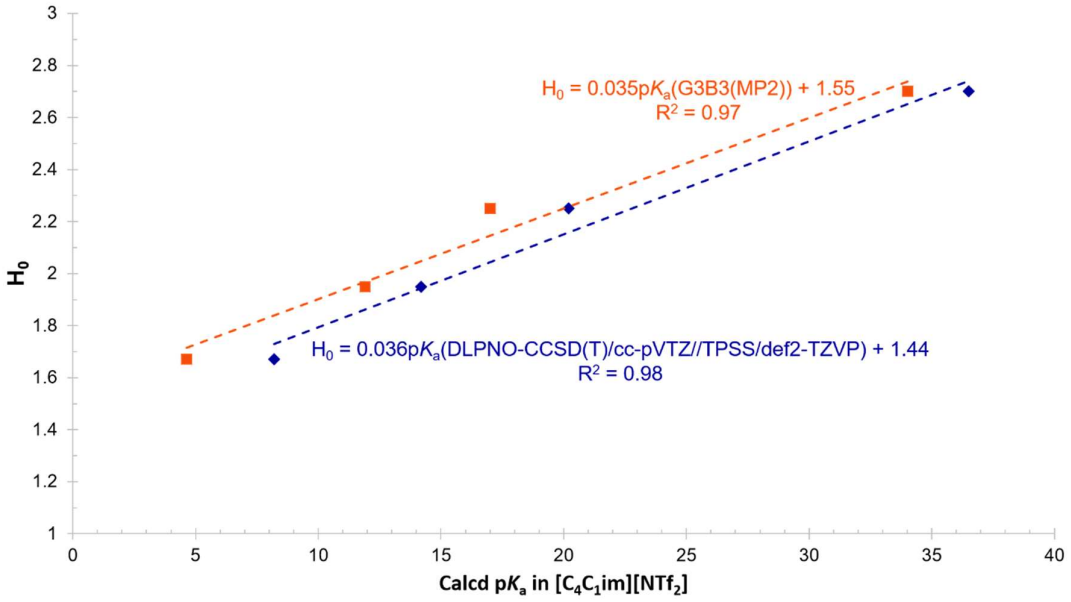


Figure 3.13 Plot illustrating correlation between experimentally determined Hammett acidity function (H_0, in water using 4-nitroaniline) and calculated pK_a in [C_4C_1im][NTf_2] using G3B3(MP2) and DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP [M05-2X/6-31+G** (SMD)] for the AILs and sulfuric acid.[38,214]

By comparing the predicted pK_a values to Hammett acidity functions found experimentally for the AILs of interest and sulfuric acid, regardless of LoT combination, the trend in acidity and

relative relationship between the acids of interest is well replicated in the predicted values, lending further support to the methodology applied.

While relation to experimental trends and thus relative acidity may not vary much between LoT, in terms of absolute values, significant variations can be seen. DLPNO-CCSD(T), on average, consistently produces higher pK_a values than the composite methods by 3-4 pK_a units. This almost doubles for $[IL\text{-}SO_3H]^+$ when compared to the CBS-QB3 prediction. However, the result using CBS-QB3 is questionable as it varies significantly from G3B3(MP2) and typical behaviour observed for the other acids. The reaction free energy in solution is far lower than the other LoT combinations. Given that the same solvent LoT and molecular structures were used throughout, the source of error is very likely the gas phase reaction free energy. This value is significantly lower than the other predictions, indicating a concern with the CBS-QB3 accuracy. This is not unexpected, as CBS-QB3 performed poorly in terms of accuracy, regardless of solvent LoT for the N-base amine conjugate acids (column total MAE of 1.05 pK_a units, Table 3.2).

Returning once again to relative trends, when measured relative to the strongest acid, i.e. sulfuric acid with a consistently low pK_a value and strong hydrolysis performance, the DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP [M05-2X/6-31+G** (SMD)] LoT combination performance significantly improves. Predictions using this protocol varies by little over 1 pK_a from G3B3(MP2)-based predictions (Table 3.13).

Furthermore, as was stated in Section 3.1.1, pK_a values are determined according to the equilibrium of a proton transfer reaction. This reaction generally involves the generation or neutralisation of a charged structure. The stability of these charged structures often shifts the equilibrium of the proton transfer reaction and thus affects the strength of the acid.[38] Polar/charged solvent systems can thus cause large changes in pK_a values, through stabilisation of these charged reactants or products. Due to the charged nature of the individual components of ionic liquids, Coulombic interactions occur that may stabilise (or destabilise) charged structures. With regards to the acid series studied here, the traditional acids would generate negatively-charged bases which would be stabilised by the cations, and the cationic acidic ionic liquids would initially be stabilised by anions and then once deprotonated generate an overall non-charged zwitterionic conjugate base (Figure 3.14).

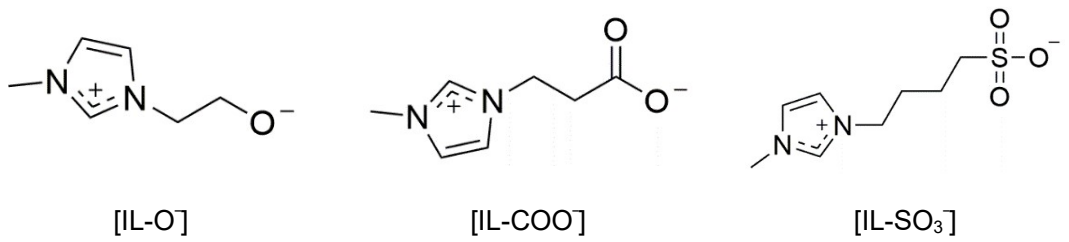


Figure 3.14 Acid-functionalised ionic liquid cations' conjugate base structures.

Hence, solvation would favour the neutral acids by stabilising the conjugate base produced but negatively affect the cationic acids through stabilisation of the positively charged reactant more than the neutral conjugate base. This would indicate higher pK_a values for the cations but lower for the neutral acids, as is observed in Table 3.13. However, having used one solvent system this does not fully address the effect the polarity or charge composition of a solvent would have on a proton transfer or dissociation reaction.

3.4 Conclusion

Initial N-base amine benchmarking tests showed a proton exchange cycle to be far more accurate and precise than a direct cycle for ionic liquid pK_a prediction. G3B3(MP2) and CBS-QB3 gas phase levels of theory combined with a M05-2X/6-31+G** solvent phase LoT were favoured in terms of accuracy and precision. However, execution time analysis illustrates the high computational cost associated with these methods that make them less suitable for larger systems (see Figure 3.6). Comparatively, DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP, included for its computational efficiency, performed relatively well in terms of precision and accuracy.

The dependence of the direct cycle on the accuracy of the proton solvation free energy greatly disfavoured its use in pK_a prediction, with many complications encountered in the solvation free energy thermodynamic cycle protocols. Nevertheless, predicted $\Delta G^*_{\mathrm{solv}}(\mathrm{H}^+)$ trends matched that found in literature, favouring anion only n = 2 clusters, although the absolute values were somewhat higher than those found in literature. Thus, overall direct cycle performance was extremely poor in terms of accuracy and precision, regardless of the phase level of theory combination chosen.

Acid-functionalised ionic liquid pK_a values were thus predicted using a proton exchange cycle with a 1H-imidazol-3-ium reference acid. Trends noted between gas phase levels of theory indicate that DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP consistently overpredicts

compared to PX cycle favoured composite methods (G3B3(MP2) and CBS-QB3); however, in terms of relative acidity (to sulfuric acid in this case) performance remains consistent. Thus, regardless of gas phase LoT, the ordering of the acids remained the same. As was expected, the molecular acids were far more acidic than the AILs, with sulfuric acid having the lowest predicted pK_as. Furthermore, within the AILs results the sulfonic acid functionalised AIL was shown to be far more acidic than the other functionalities tested. $[IL\text{-}SO_3H]^+$ pK_a value were on average closer to butane-1-sulfonic acid and sulfuric acid than $[IL\text{-}COOH]^+$ and $[IL\text{-}OH]^+$. This matched experimental Hammett acidity trends, which show $[IL\text{-}SO_3H]^+$ to fall closer in magnitude to sulfuric acid when measured under the same conditions.

However, given the variation between ionic liquid solvents systems noted for the N-base amine conjugate acid PX results, further analysis into the effect of the solvent on the thermodynamics of the protonation energy and the reaction as a whole needs to be done.

4.

Implicit Solvation Effects on Cellobiose Acid Hydrolysis Mechanism

The effects of ionic liquid solvation on the thermodynamics of the aqueous cellobiose hydrolysis mechanism proposed by Loerbroks et al. were assessed using the SMD implicit solvent model modified for ionic liquids. Effects were analysed in terms of free energy requirements for the reaction and compared to aqueous and gaseous pathways. The findings of these solvation investigations showed electrostatics to dominate the solvation effect, making dielectric constant magnitude a key determinant of stabilisation/solvation ability. Consequently, water gave the lowest intermediate activation barriers and final reaction energy. Implicit solvation would poorly represent the hydrogen bonding abilities of the ionic liquids, skewing their performance according to electrostatics and not hydrogen bonding capabilities. However, amongst the ionic liquid reaction profile, no clear connection with dielectric constant magnitude or hydrogen bonding ability could be established, with $[C_4C_1im][MeSO_4]$ outperforming $[C_4C_1im][Cl]$, which gave similar results to $[C_4C_1im][NTf_2]$. Overall, due to the dominating effect of electrostatics in implicit solvation behaviour predictions, ionic liquids, having lower dielectric constants compared to water, unfavourably raised the protonation energy (highest energy requirement) as well as the conformational change barrier (rate-determining step of the mechanism) compared to water.

4.1 Background

For use in biorefinery processes, cellulose needs to be depolymerised or hydrolysed to intermediates more susceptible to chemical and/or biological change. One such intermediate is glucose. Glucose generation from cellulose requires a saccharide conversion, whereby the *β*–(1,4)-glycosidic bond linkages in the cellobiose polymer repeating units are cleaved.[8] This can be achieved through acidic hydrolysis, with the use of a relatively strong acid in a suitable solvent.[3] However, this decreases the "green" nature of the process and increases risks and costs, making it inefficient for industry use. This is generally attributed to cellulose's intermolecular hydrogen-bonding networks, which limit solubility in common solvents.[215]

It is well-known that the hydrolysis reaction is hampered by the structure of the cellulose or cellobiose monomers. Using an appropriate solvent allows greater access to the glycosidic linkages through disruption of the extensive hydrogen bonding network. The naturally occurring chair conformations allow specific hydrogen bonds to form between the glucose subunits.[24] Effective solvation has been shown to break these hydrogen bonds through formation of solute-solvent hydrogen bonds.[29,56] This further has the effect of decreasing electron sharing abilities of endocyclic oxygens, reducing anomeric effects.[8] Overall, this affords a more flexible structure that would be more prone to hydrolysis.

Investigations found ILs with strong hydrogen-bond acceptor anions (Cl, carboxylates) to be good dissolution agents for cellulose, as they can effectively break up the hydrogen bonding networks.[28,216] This results in better yields and milder reaction conditions. It is thought that this is through increased access to hydrolysis reaction sites. However, this does explain the effect of these solvents on the thermodynamics of the mechanism.

Literature descriptions of the cellobiose acid hydrolysis mechanism are generally based around three main steps. The glycosidic linkage oxygen atom is first protonated, then cleaved to form glucose and a charged cyclic intermediate, which undergoes the nucleophilic addition of a hydroxyl using water generating glucose.[3] Given the extensive hydrogen-bonding network surrounding the glycosidic linkage (Figure 1.3), great significance has been placed on the initial protonation step, often denoted the highest energy requirement step.[15,24] However, it is a barrierless process and thus only forms part of the total activation energy, albeit a significant part. The remaining energy is a result of the conformational changes and bond breaking/bond forming energy required to convert the protonated cellobiose to glucose. Solvent stabilisation could greatly affect these barriers, resulting in reduced rates or unfavourable energy requirements.[217] As such, many solvation studies have been conducted for cellulose acid hydrolysis, assessing the mechanism of action and effect of the solvent system on the mechanism.[23,29,49,218,219] Most theoretical studies make use of explicit solvent models, as they explicitly include solvent-solute intermolecular bonding and thus more realistically represent their effect. However, these methods are very computationally demanding and the importance of the explicitly represented hydrogen bonding interaction largely related to the breaking up of the supramolecular structure. In terms of mechanism thermodynamics and kinetics these interactions may not play as much of a role. Thus, to reduce cost, implicit solvent models expanded to allow inclusion of ionic liquids, may be used.[18] Overall this study does not aim to

study the cellulose solvation ability of the ionic liquids, which is determined by hydrogen bonding abilities,[8,15,55,213,220] but rather their effect on the thermodynamics of the hydrolysis reaction.

4.1.1 A Summary of Cellulose Acid Hydrolysis Mechanisms Thus Far

As was detailed in Chapter 1 Section 1.3.1-2, the exact mechanism of cellobiose acid hydrolysis is not known with certainty; however, many computational studies have been done to test the viability of proposed mechanisms in specific solvents. Generally, biomimetic mechanisms are favoured, which are based on the mechanism of action of glycoside hydrolase enzymes that convert lignocellulose to sugar derivatives in nature.[5,45,221,222] However, these enzymes create and stabilise convenient/reactive conformations that may not be possible under catalytic conditions; specifically with regards to the protonation of the glycosidic linkage, where enzymatic studies have shown activation of the glycosidic linkage through forced conformation shifts in the glucose ring subunits.[5,222] Furthermore, enzymatically the cellobiose monomer is held in position where acid functionalities/proton donors can interact with both the glycosidic linkage oxygen and the anomeric carbon.[5,6,223] This allows for the effective protonation of the linkage, through either a concerted or stepwise double displacement mechanism (Scheme 1.1, see Chapter 1 Section 1.3.1 for more details).[6]

Under non-enzymatic conditions, the orientation of the reaction coordinates is not fixed and thus acidic protonation of the glycosidic linkage is dependent on availability. Proton availability depends largely on acid strength and concentration. However, unlike enzymes catalysts have no active site, and thus are less specific, allowing for protonation of other basic sites on the cellobiose. This is further complicated by the cellobiose conformations that occur in the crystal form of the cellulose found most prominently in nature. In the crystalline domains of cellulose I_β, two conformations, related to the hydrogen bonding patterns, dominate (Figure 4.1).[24]

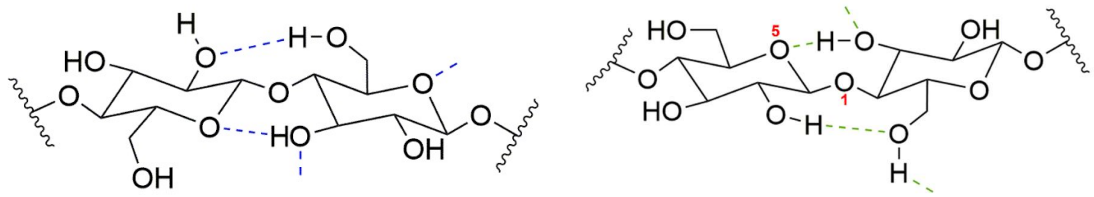
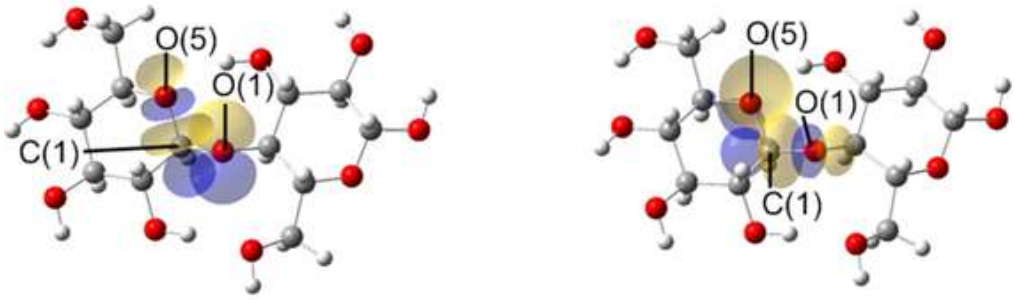


Figure 4.1 Representations of cellobiose conformers found in cellulose I_β sheets, according to the hydrogen-bonding patterns found (shown in blue and green).[24]

Natural bond orbital (NBO) analysis of oxygen atoms surrounding the linkage in these structures has shown the linkage oxygen to be the least basic by a large margin.[24] This is further affirmed by experimental pK_b values of protonated acetal and hydroxyl groups, showing the hydroxyl oxygen sites to be significantly more basic than the glycosidic linkage oxygen atoms.[224] Thus, without direct forced interaction, all the surrounding hydroxyl and pyranic oxygen atoms will be preferentially protonated. This can explain the necessity of a strong acid catalyst, which would also protonate less basic sites such as the glycosidic linkage. In addition, this also links to alternatively proposed mechanisms, where the pyranic O_5 endocyclic oxygen is protonated instead of the glycosidic linkage oxygen atom, O_1.

However, pyranic protonation mechanisms proceed through an acyclic intermediate (Scheme 1.2) and are not favoured in literature, as they are found to lead to higher energy transition states and thus overall higher barriers.[3,24] This may in part be due to electronic effects present in the conformers. O_5 and O_1 are linked by anomeric effects, with protonation of either causing charge to be transferred between the oxygen lone pairs (n_{O_1}, n_{O_5}) and C_1-O_1 or C_1-O_5 antibonding orbitals ($\sigma^*_{C_1O_1}$, $\sigma^*_{C_1O_5}$) (Figure 4.2).[24] Protonation of O_1 lowers the energy of $\sigma^*_{C_1O_1}$causing a charge transfer between n_{O_5} and $\sigma^*_{C_1O_1}$. This increases the occupancy of $\sigma^*_{C_1O_1}$, weakening the bond and causing it to lengthen. Now weakened/elongated, this bond is easier to break, facilitating hydrolysis. Protonation of the endocyclic O_5 oxygen atom, on the other hand, significantly lengthens the C_1-O_5 bond, while slightly shortening the C_1-O_1 bond. This is through protonation once again lowering the energy of the connected antibonding orbital ($\sigma^*_{C_1O_5}$), facilitating a $n_{O_1} \rightarrow \sigma^*_{C_1O_5}$ charge transfer.[8,24]



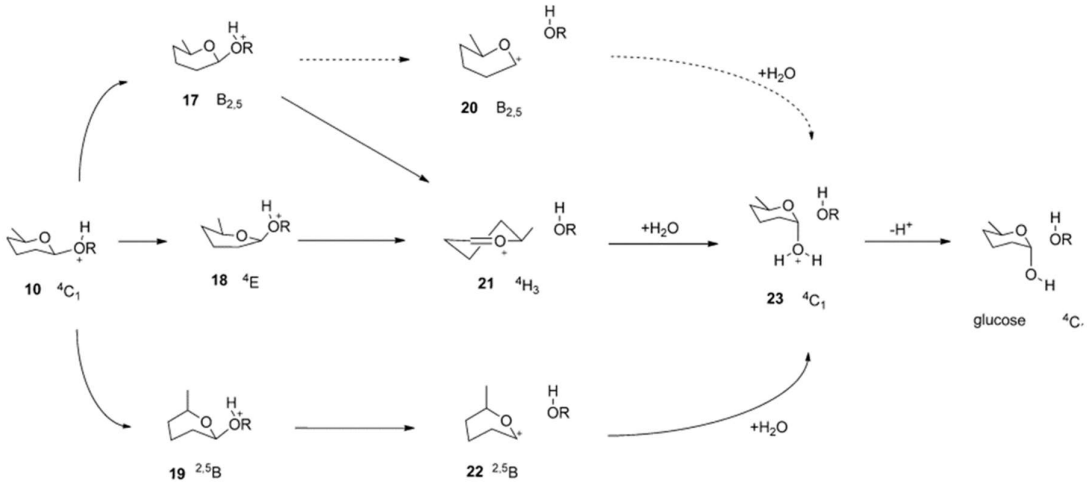
$n_{O_1} \rightarrow \sigma^*_{C_1O_5}$, exo-anomeric effect`1 $n_{O_5} \rightarrow \sigma^*_{C_1O_1}$, endo-anomeric effect

Figure 4.2 Figures from Loerbroks et al. (2013) illustrating anomeric effects present in cellobiose.[24]

From this activated protonated intermediate the glycosidic bond can be broken to form a cyclic carbocation intermediate, which is converted to glucose by the nucleophilic addition of a hydroxyl at the anomeric carbon. The protonation and nucleophilic addition steps are both thought to be barrierless, with only the bond cleavage having a barrier. However, some computational studies have suggested that the bond cleavage is preceded by a conformation change.[24,225]

In a 2013 implicit water study, Loerbroks et al. found that altering the conformation of the glucose rings in the protonated cellobiose from a chair conformation, greatly reduces the activation barrier. They linked this to anomeric effects present in the cellobiose structure, stating that changing the ring conformation can enhance $n_{O_5} \rightarrow \sigma^*_{C_1O_1}$ donation. As with the initial glycosidic linkage oxygen protonation, this would weaken or activate the glycosidic linkage resulting in a lower overall energy barrier. Implementing this conformational change into the mechanism, they were able to replicate experimental thermodynamic data for aqueous sulfuric acid cellobiose hydrolysis, using implicit solvation. When calculated using just the crystalline structure chair conformation, both Loerbroks et al. and other computational studies were unable to achieve the same accuracy,[15,24] further validating their proposed aqueous mechanism. This suits enzymatic studies, which show the cellulose reactant substrate to be forced into less stable ring conformations in the activated complex, linking ring conformational change to weakening of the glycosidic linkage and effective breakdown of the cellobiose into glucose.

After investigating three different conformation intermediates ($B_{2,5}$, ^4E and $^{2,5}B$), chosen to facilitate $n_{O_5} \rightarrow \sigma^*_{C_1O_1}$ charge transfer, the lowest energy path was found to proceed through a ^4E conformer intermediate (Str 18, Scheme 1.3).[24]



Scheme 4.1 Repeat of Scheme 1.3.

Cleavage of the glycosidic linkage resulted in the formation of a $^4H_3$ half-chair oxocarbenium intermediate (Str 21) and glucose. Water then attacks the C1 or anomeric carbon atom of the oxocarbenium intermediate and is quickly deprotonated by the acid catalyst residue, bring the mechanism to completion by forming the final glucose molecule (Scheme 1.3). The other suggested conformational changes resulted in higher energy barriers relative to the Str 18 path, which when combined with the protonation energy, determined using a thermodynamic cycle, resulted in reaction energy of 141.0 kJ mol^{-1}.[24] This mechanism is however, for aqueous solvation conditions and implicit solvation. Given the charged nature of the ionic liquid components and drastic reduction in activation energy seen experimentally when they are used as the reaction medium, there is a chance that ionic liquid solvation may induce a different mechanistic path. However, in an expansion to their original study, Loerbroks et al. (2015) used molecular dynamics and metadynamics simulations to investigate the effect of solvation on the mechanism, in which water and $[C_2C_1im][MeCO_2]$ were compared for cellobiose hydrolysis, according to their optimal mechanistic pathway of Str 10 → Str 18 → Str 21 → Str 23. They found evidence that $[C_2C_1im][MeCO_2]$ facilitates activation of the glycosidic linkage through stabilisation of non-chair conformers and increasing rotational freedom around the glycosidic linkage through disruption of hydrogen

bonding and exo-anomeric effects.[217] This seems to indicate that ionic liquid solvation would enhance a mechanistic pathway containing conformational changes.

Alternative mechanisms have been shown to be possible in ionic liquids. Nishimura et al. have proposed a mechanism that neglects the conformational change activation step.[15] They used a hybrid quantum chemical/reference interaction site solvent model to represent [C_1C_1im][Cl] solvation of cellobiose acid hydrolysis. They were able to calculate a activation energy in good agreement with experimental findings for a different ionic liquid ([C_2C_1im][Cl]),[14] however their comparative water results were extremely poor.[15] Given this variation amongst solvent results as well as the robust performance of the Loerbroks et al. (2013) mechanism in various solvent and solvent model systems, it is more likely that the mechanism proceeds through a conformational change. This correlates with proposed enzymatic mechanisms and given the nature of an ionic liquids and how they interact with cellulose/cellobiose (as will be discussed in the following section), stabilisation of less stable ring conformers is highly likely, facilitating an enzymatic mechanism.[5–7]

4.1.2 Solvation Effects on Cellobiose Hydrolysis

Solvents can be characterised by their physical properties (e.g. boiling point, density) and chemical properties (e.g. hydrogen bonding abilities, polarity).[226] However, a solvent is also a medium made up of individual molecules which interact with the solute and one another on an intermolecular level. This can include polarisation effects as well as stabilisation through intermolecular interactions such as hydrogen bonding. This is a key feature for cellulose/cellobiose solvents, where access to the reaction sites (glycosidic linkages), through solvation, predominantly requires the disruption of extensive hydrogen-bonding networks.[28] Many of the ionic liquids shown to be good cellobiose solvents have good hydrogen bonding abilities, typically through the anions.[34,56] Studies show that effective dissolution of cellulose in ionic liquids is predominantly determined by the nature of the anion. Anions predominantly interact with the cellulose hydroxyls, closely surrounding the cellobiose monomers, followed by the cations. This is a phenomenon called preferential solvation, where interactions with one component of a solvent or solvent mixture produces larger negative Gibbs solvation free energies, causing preferential binding/interaction. In the case of imidazolium based ionic liquids, this is due to favourable hydrogen bonding interactions between the anions and the solute hydroxyls. As strong hydrogen bond acceptors, the anions form stronger hydrogen bonds with the hydroxyls, displacing the previous inter- and intra-molecular hydrogen bonds between

the various ring oxygens.[227] Many computational studies have illustrated this through explicit solvent modelling, showing anions predominantly occupy the first solvation shell and strongly hydrogen bond with hydroxyl groups.[8,44,216] This was confirmed in $^{35/37}$Cl and ^{13}C NMR studies, which indicate a stoichiometric interaction between both groups.[55] $[Cl]^-$ is an anion of choice, given the significance of hydrogen bonding. It is a small, strong hydrogen bond acceptor and thus when coupled with the appropriate cation, is one of the best cellulose dissolution agents. However, it does lead to the formation of relatively viscose liquids, constraining its ability and usage, possibly necessitating harsh reaction conditions. Increasing the temperature of a liquid can generally decrease its viscosity. However, at higher temperatures by-products can form, reaction vessel choice is limited, and process costs increase. Other anions such as $[NTf_2]^-$ result in lower viscosities but are weaker hydrogen bond acceptors. Key to the use of the appropriate anion is thus also the choice of cation, which can greatly alter these physical solvent properties.

Methylimidazolium or methylpyridinium cores with butyl to hexyl alkyl chains are favoured as cellulose solvents. This is mainly due to effect of molecular symmetry on melting point and viscosity. Lower symmetry cations such as $[C_4C_1im]^+$, have lower melting points, while still maintaining a small enough molecular weight to be mobile and short enough alkyl chain to limit attractive Van der Waals interactions. They also form strong C-H···halide hydrogen bonds, interacting well with anions such as $[Cl]^-$. This once again leads to increased viscosity but also lower vapour pressure and better thermal stability, increasing safety and applicability to biorefinery procedures.[57]

This does not exclude the cation from playing a role in the solvation process. Cation choice and subsequent interaction, with both the anion and cellulose solute, can play an important role in the dissolution process. A computational explicit solvent study done by Lu et al. showed that identity of the cationic species in an ionic liquid greatly affected the dissolution ability of the solvent.[58] These effects are centred around the presence of acidic protons on the heterocyclic ring of the cations. Usually found in conjugated imidazolium rings, these acidic protons form weak hydrogen bonds with the hydroxyl groups or ether oxygen atoms present in cellulose, increasing solubility in much the same manner as the anions. However, the cations act as hydrogen bond donors, with the acidic ring protons interacting directly with the cellulose oxygens in a C-H···$O_{cellulose}$ hydrogen bond. The strength of this bond is affected by how acidic the protons are and this can be influenced by substituents present in the ring, particularly on

the alkyl sidechain. Electron donating groups such as -OCH_3 and -OH decrease the acidity of the ring protons, reducing their hydrogen bonding ability. Nonetheless, explicit computational studies show the anions to form far more and far stronger hydrogen bonds to the ring hydroxyls, about 0.3-0.4 Å shorter than the cation hydrogen bonds in the case of [C_4C_1im][Cl].[57] This implies that the anion is still mainly responsible to the dissolution process; however, the cation's interaction with the anion can affect its ability to interact with cellulose, with stronger anion-cation interactions leading to a decrease in solubility. Furthermore, increased cation size can also lead to steric hinderance, which has also shown to decrease cellulose solubility. Lindman et al. have also proposed that the cation's effect may be linked more to hydrophobic interactions with the sugar polymer, given that both are amphiphilic in nature.[29] However, these suggestions are not widely considered, with most studies focusing on the anion, given the overwhelming experimental (NMR, etc.) and computational evidence illustrating the significance of its effect.

Functionalisation of the cation (e.g. with an acidic moiety) provides the possibility of a one-pot synthetic process, where an ionic liquid would act as both the acid catalyst and the solvent. This is discussed and assessed in more detail in Chapter 3. Given the significance of hydrogen-bonding in the dissolution, a large emphasis has been placed on explicit solvent studies. However, this approach can be time consuming and resource heavy, making exploration of multiple solvents and various mechanistic paths costly. An attractive alternative is presented in implicit solvents, which reduce both time and resource costs, but can still be representative of specific solvent characteristics. These models, detailed in Chapter 2 Section 2.6, reduce the solvent to a dielectric continuum, negating the inclusion and optimisation of multiple individual solvent molecules. Solvation effects are largely driven by electrostatic interaction of the continuum, but the influence of explicit interactions are still included through parameterisation of an atomic surface tension term, that represents cavitation, dispersion and solvent structural (CDS) affects. Hydrogen-bonding is specifically included in this factor and represented using Abraham parameters. Unfortunately, these are not often found in literature for ionic liquids; rather, the hydrogen bonding abilities of ionic liquids are usually characterised using Kamlet and Taft parameters. The parameters differ on how they measure the ability but more importantly how they treat the solvent. Abraham hydrogen bond acidity and basicity values treat the molecule of interest as a solute, while Kamlet and Taft treat it as a solvent. Bernales et al. developed a correction between the values, using 19 organic solvents where both values were determined experimentally. This subsequently allows application of the SMD model to

various ionic liquids as well. However, it should be mentioned that while the correlation for hydrogen bond acidity or donor strength was high (R^2 = 0.94), the correlation between hydrogen bond acceptor and basicity was only 0.61.[18] Nonetheless, the model allows for the simulation of ionic liquid solvation implicitly and has been used in literature to varying levels of success in the calculation of solvation energies.

4.2 Computational Details

4.2.1 General

All optimisations were run using Gaussian09.[174] Tight convergence criteria and an ultrafine grid for numerical integration of the exchange-correlation energy were used for optimisations. Local minima and transition states were confirmed using harmonic force constant (frequency) analysis.

For comparison, replication of the Loerbroks et al. study was done at BB1K/6-31++G**,[228] in accordance with the original conditions.[24] Additionally, protonation free energies were calculated in accordance with protocols established and detailed in Chapter 3. Thus, gas phase energies were determined from TPSS/def2-TZVP optimised structures using DLPNO-CCSD(T)/cc-pVTZ.[71,171,172] Solvation free energies were calculated at the M05-2X/6-31+G** level (including the geometry optimization), according to Equation 3.22.[184] This is the best LoT combination to predict acidity of acidic ionic liquids in ionic liquids using a proton exchange cycle. For further information on the specifics of each phase leg see Chapter 3 Sections 3.2.1 and 3.2.2.

All other steps, for the rest of the mechanism, were calculated using the BB1K functional, but with a larger basis set more specific to DFT calculations, def2-TZVP.[172]

4.2.2 Solvent Models

All solvent systems, unless otherwise stated, were represented using the SMD model,[127] with [C_4C_1im][Cl] represented using SMD-PGP as no experimental refractive index could be found.[18] The other parameters, for all the ionic liquid solvents used, are the same as those listed in Section 3.2.4 of Chapter 3. Default Gaussian09 implemented parameters were used for SMD water.[127]

To replicate the Loerbroks et al. study BB1K/6-31++G** water calculations were also done using the default Gaussian09 implementation of the C-PCM model.[121,126] This makes use of universal force field (UFF) radii, scaled by a factor of 1.1, for creation of the solvent cavity.

4.2.3 RMSD Calculation, Hydrogen Bond Analysis and Graphics

See Chapter 3 Sections 3.2.5 and 3.2.6.

4.3 Results and Discussion

This study focuses on the mechanism put forth by Loerbroks et al. for cellobiose acid hydrolysis in an aqueous solvent.[24] Their approach splits the hydrolysis reaction into two parts, protonation and bond breaking/formation. The total free energy profile is then determined by the combination of both these parts

The efficiency of the protonation step depends on the activity of the acid catalyst, which can conveniently be measured in terms of a pK_a value or Hammett acidity function, as done in Chapter 3. While these expressions are convenient to compare acid strengths in various solvents, they do not include a measure of the thermodynamic stability of the protonated reactive species, in this case cellobiose. The discussion therefore commences by first addressing the protonation equilibrium in ionic liquid solvent.

4.3.1 Cellobiose Protonation Free Energy in Ionic Liquids

Activation of the glycosidic linkage always involves the protonation of the glycosidic oxygen atom, which can be followed by a conformation change as suggested by Loerbroks et al.[24] Conversion to glucose becomes far more likely from the protonated starting product, which is "activated" towards scission of the glycosidic linkage through elongation of exocyclic C-O bonds and disruption of electronic stabilisation effects such as the anomeric effect.[7,15,24] This reduces the activation energy required for the hydrolysis, favouring higher glucose and TRS yields.

Consequently, to determine the full effect of ionic liquids as catalysts and solvents, the free energies of protonation of cellobiose (Figures 3.1 and Figure 3.12 (ii)) with each acid in $[C_4C_1im][NTf_2]$, and two known effective cellulose dissolution solvents, $[C_4C_1im][Cl]$ and $[C_4C_1im][MeSO_4]$, were calculated (Table 4.1). These are compared to gas phase and aqueous (SMD) phase protonation free energies calculated in the same manner. Equilibrium constants

were calculated at 298.15 K using the standard state free energy change with all species at 1 mol L^{-1} concentration:

$$K = \exp\left(\frac{-\Delta G^*_{soln}}{RT}\right) \quad \textbf{(4.1)}$$

Table 4.1. Free energy and equilibrium constant estimates for the protonation of cellobiose by sulfuric acid and acid-functionalised cationic acids (HA) in gas, water and multiple ionic liquid solvents, calculated using DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP [M05-2X/6-31+G** (SMD)].

Phase thermodynamic quantities		HA			
		[IL-OH]$^+$	*[IL-COOH]$^+$*	*[IL-SO$_3$H]$^+$*	*Sulfuric acid*
Gas	ΔG^*_{gas} **(kJ mol^{-1})**	477.42	369.71	316.96	591.61
	K	2.29×10^{-84}	1.70×10^{-65}	2.95×10^{-56}	2.30×10^{-104}
[C$_4$C$_1$im][NTf$_2$]	ΔG^*_{soln}(**Str 10**) **(kJ mol^{-1})**	381.82	288.67	254.65	219.90
	K	1.28×10^{-67}	2.67×10^{-51}	2.44×10^{-45}	2.99×10^{-39}
[C$_4$C$_1$im][Cl]	ΔG^*_{soln}(**Str 10**) **(kJ mol^{-1})**	389.21	292.83	259.41	215.92
	K	6.50×10^{-69}	5.00×10^{-52}	3.58×10^{-46}	1.48×10^{-38}
[C$_4$C$_1$im][MeSO$_4$]	ΔG^*_{soln}(**Str 10**) **(kJ mol^{-1})**	376.53	282.24	247.93	208.38
	K	1.08×10^{-66}	3.57×10^{-50}	3.67×10^{-44}	3.12×10^{-37}
H$_2$O	ΔG^*_{soln}(**Str 10**) **(kJ mol^{-1})**	-147.34	-127.50	-114.31	-430.03
	K	6.51×10^{25}	2.17×10^{22}	1.06×10^{20}	2.18×10^{75}

These energies are the reaction free energies in solution of a proton transfer reaction between each acid and cellobiose and thus can be accurately predicted with an optimised PX thermodynamic cycle protocol. Hence, DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP [M05-2X/6-31+G** (SMD)] PX cycles were used. Of the acidic ionic liquids, [IL-SO_3H]$^+$ shows the lowest protonation free energy in solution and gas, implying that it produces the most stabilised conjugate base structure, which matches the trends in acidity calculated earlier.

Water facilitates the lowest solution protonation free energy; whereas [C_4C_1im][Cl] gives the highest solution protonation free energies of the ionic liquids as well as overall. Given that

solvation was represented implicitly, electrostatic contributions are of greater significance thus it is more likely that solvents with higher dielectric constants will stabilise the structure more. Under these conditions, water ($\varepsilon_r = 78$) would be expected to produce the lowest energies. This distinction is less clear when applied to the ionic liquids, as $[C_4C_1im][Cl]$ has the highest value ($\varepsilon_r = 15$,[190] compared to 11.52[191] and 14.80[190] for $C_4C_1im][NTf_2]$ and $C_4C_1im][MeSO_4]$, respectively), but still produces the highest energies.

These values imply $[C_4C_1im][Cl]$ is a poor solvent but it should also be kept in mind that a key determinant in cellulose solvation is the hydrogen bonding ability, which is moderately represented in the SMD model in the CDS contribution using Abraham parameters. The CDS component is only approximately 5% of the overall solvation free energy, however the correlation between Kamlet-Taft acceptor parameters and Abraham basicity parameters is weak ($R^2 = 0.61$).[18] Thus, the true effect of the anion choice may not be well represented in the implicit solvent model, making it difficult to distinguish between the various ionic liquids tested based on hydrogen-bonding abilities.

4.3.2 Implicit Solvation Effects on Bond Breaking and Formation Steps

Starting structure coordinates were taken from the Loerbroks et al.[24] supporting information provided and reoptimized accordingly at the levels of theory used here. An exemplar of conformation numbering is represented below (Figure 4.3) using the protonated cellobiose reactant (Str 10), to allow for easy referencing in the following sections.

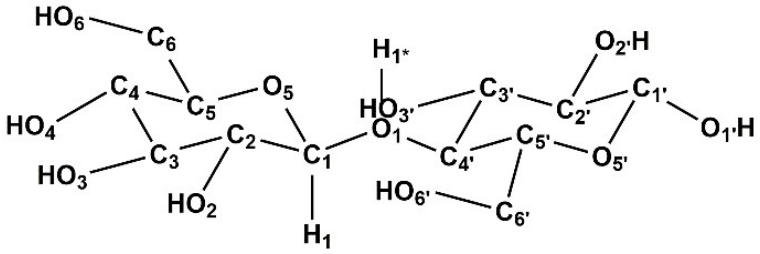


Figure 4.3 Example of structure numbering used.

Only the preferred pathway is studied as the focus was on the relative effect of ionic liquid solvation on the energetics of the hydrolysis mechanism. Thus, focus was on stability relative to the original, implicit water computations, with analysis focusing on the changes in the individual free energy, activation barriers and corresponding structural changes.

The structure naming/numbering of the source paper has been retained here, thus the initial cellobiose is Str 1, protonated form is Str 10, and so on. Furthermore, each glucose monomer of cellobiose is termed as either non-reducing or reducing according to the following illustration (Figure 4.4).

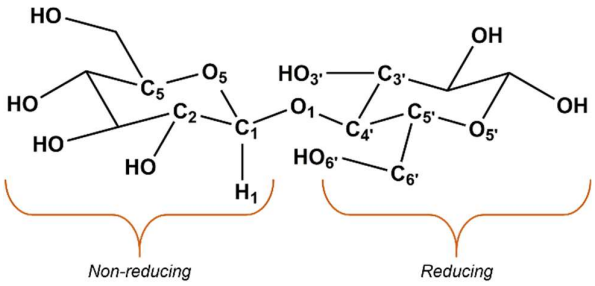


Figure 4.4 Cellobiose monomer with glucose subunits labelled as is referenced in the text.

4.3.2.1 Study Replication and Comparison of Selected Implicit Solvation Models

Calculations were performed with an alternative basis set to that used in the original Loerbroks et al. paper.[24] The DFT level of theory was maintained as BB1K as it has been shown to perform well for transition state structure prediction and energy barrier heights in hydrogen-transfer reaction benchmark studies and sugar hydrolysis studies.[24,223,229,230] For comparison to ionic liquids the solvent model was also changed to SMD. Given these changes an initial comparison study was conducted to evaluate what effect, if any, these changes would have. Loerbroks et al.'s parameters, of BB1K/6-31++G** and C-PCM water were replicated and compared to SMD water BB1K/6-31++G** and BB1K/def2-TZVP; with the original Loerbroks et al. (2013) paper free energy values included as well.[24] The effects of these changes are illustrated in Figures 4.5, which shows a series of reaction profiles with all free energy values relative to the relevant Str 10 for the specified calculation conditions, such that the effect on the reaction energy barriers and overall profile can be assessed.

The profile shown in grey was created using the energy values given in the original paper. The profile shown in purple is the replication, with the supporting information structures used as initial structural guesses. The replicated profile is slightly lower in general, by about 3 kJ mol^{-1} (Figure 4.5).

RMSDs and comparisons of hydrogen bonding for the mechanism structures show insignificant variation (Figure 4.6), leading to the conclusion that the free energy variation could be a result of the convergence criteria used, which is not stated in the paper. Changes in

basis set as expected, increased the free energy but cause very little structural variation, with an average RMSD of 0.042 Å (Figures 4.6-7).

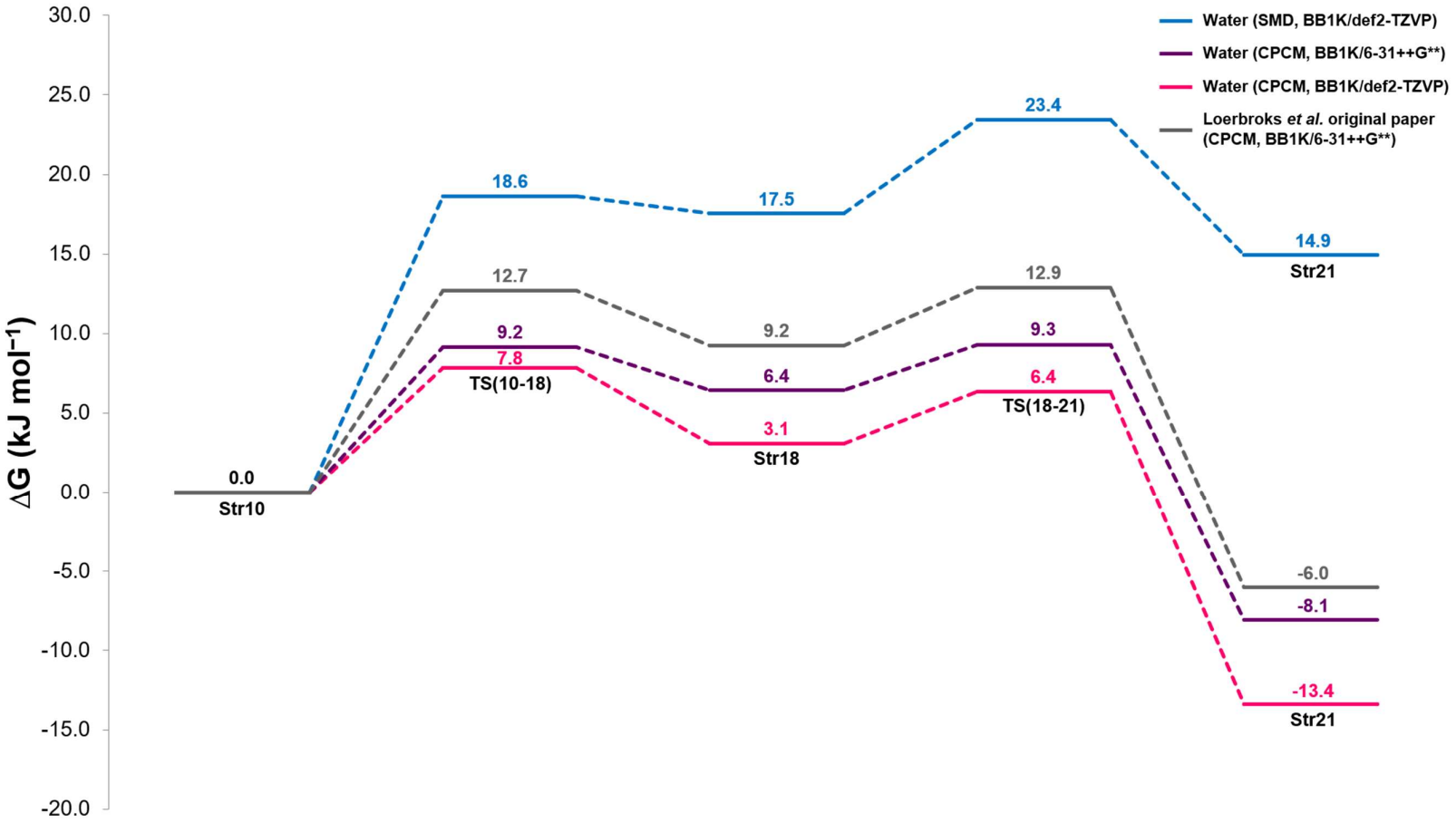


Figure 4.5 Reaction profiles of the Loerbroks et al. preferred cellobiose hydrolysis mechanism at BB1K/6-31++G** and BB1K/def2-TZVP in aqueous phase using SMD and C-PCM implicit models, relative to their respective Str 10 free energies.

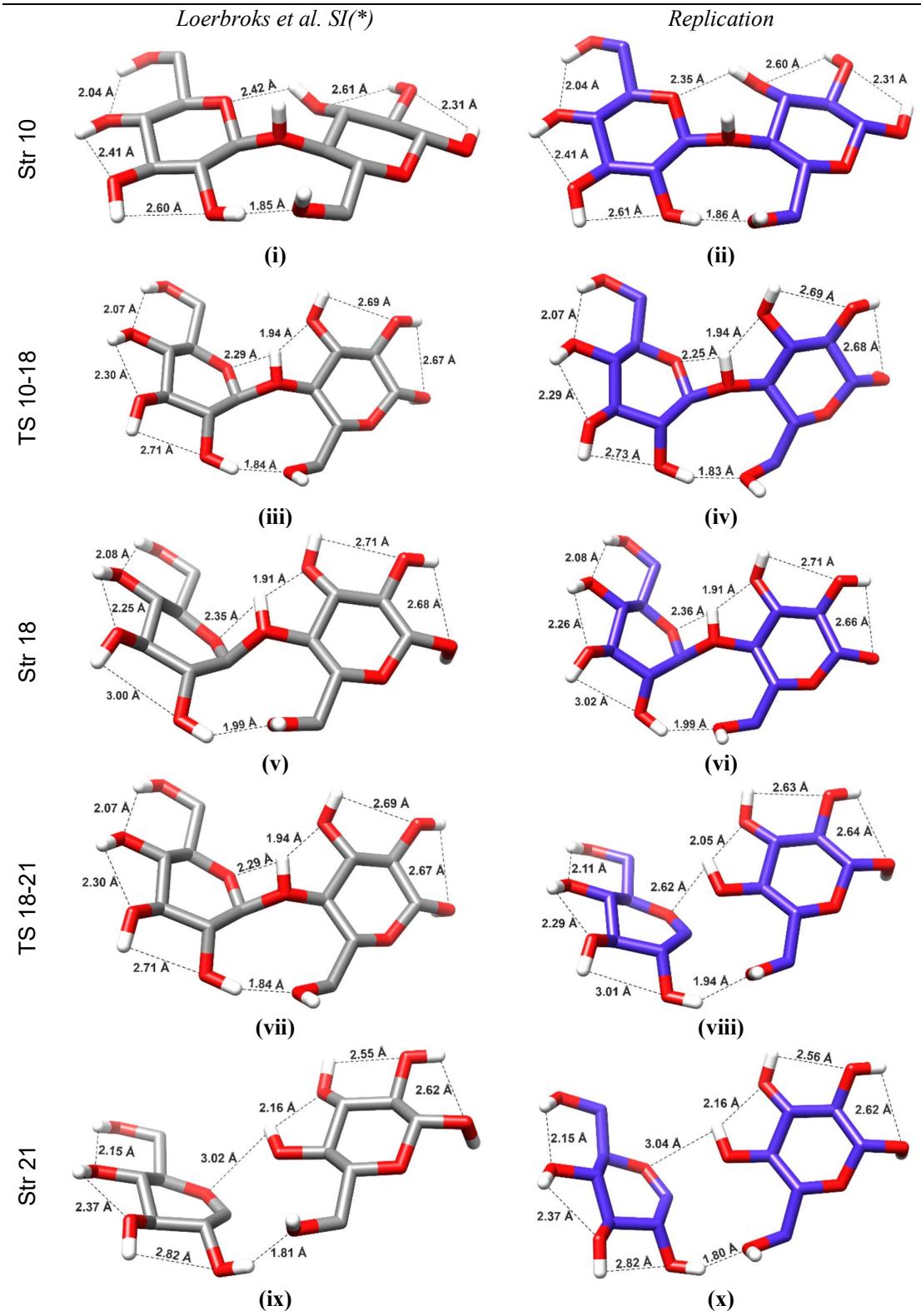


Figure 4.6 Original Loerbroks et al. (2013) structures (left) compared to structures replicated using same conditions of BB1K/6-31++G** (right).[24]

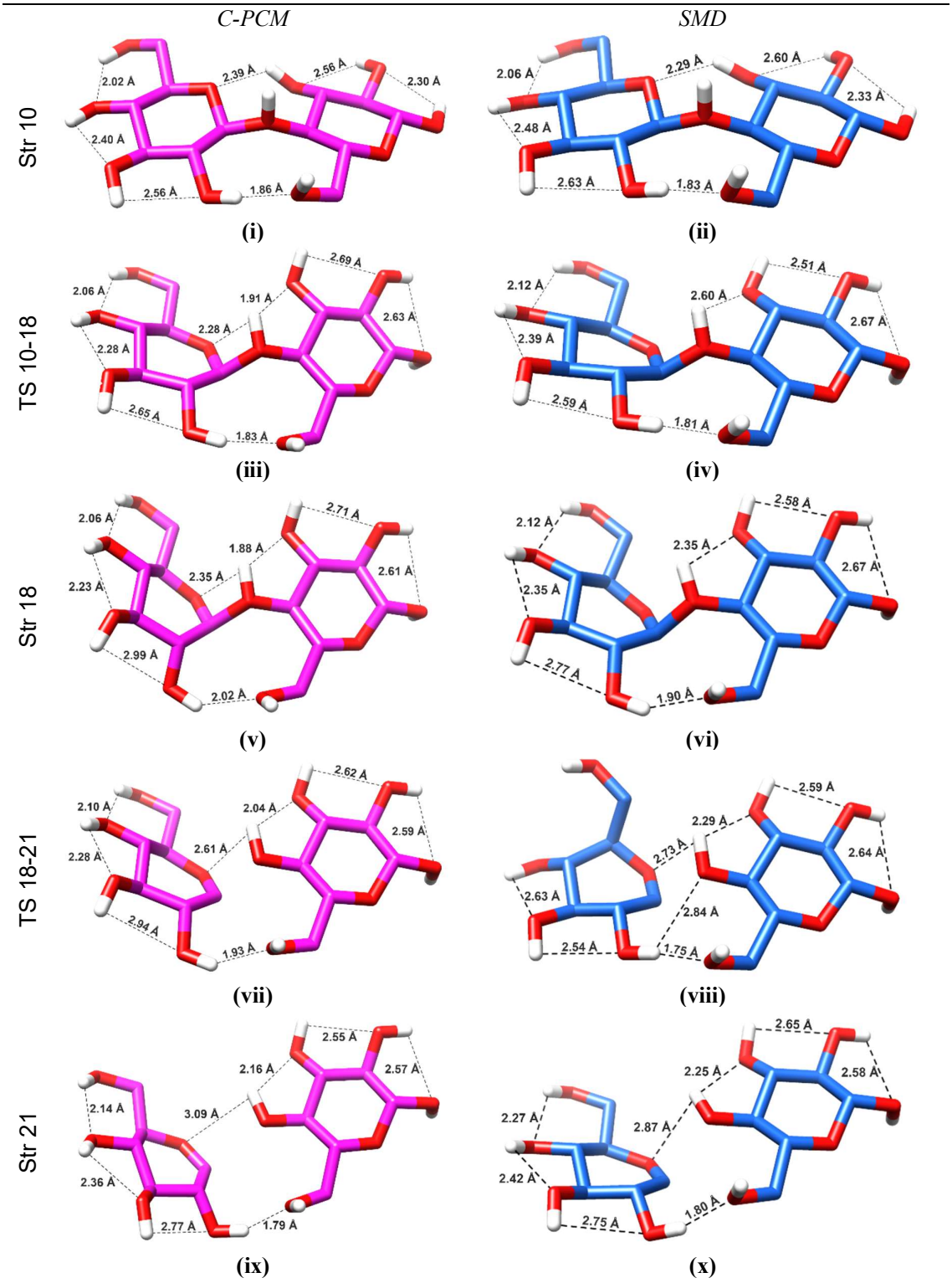


Figure 4.7 Comparative structures, optimised at BB1K/def2-TZVP using C-PCM (left) or SMD (right) water.

Changing the implicit solvent model from C-PCM to SMD results in the greatest changes in free energy. Solvation free energy calculated using the SMD model is on average 113.4 kJ mol^{-1} more negative than the C-PCM values. Str 10 is stabilised more using SMD than C-PCM, compared to the proceeding structures, which consequently leads to a higher conformational change activation barrier of 18.6 kJ mol^{-1}.

Large structural variation is also seen due to a shift in the orientation of the two glucose rings, illustrated in Figure 4.8 below, where the SMD structures are superimposed on the C-PCM structures.

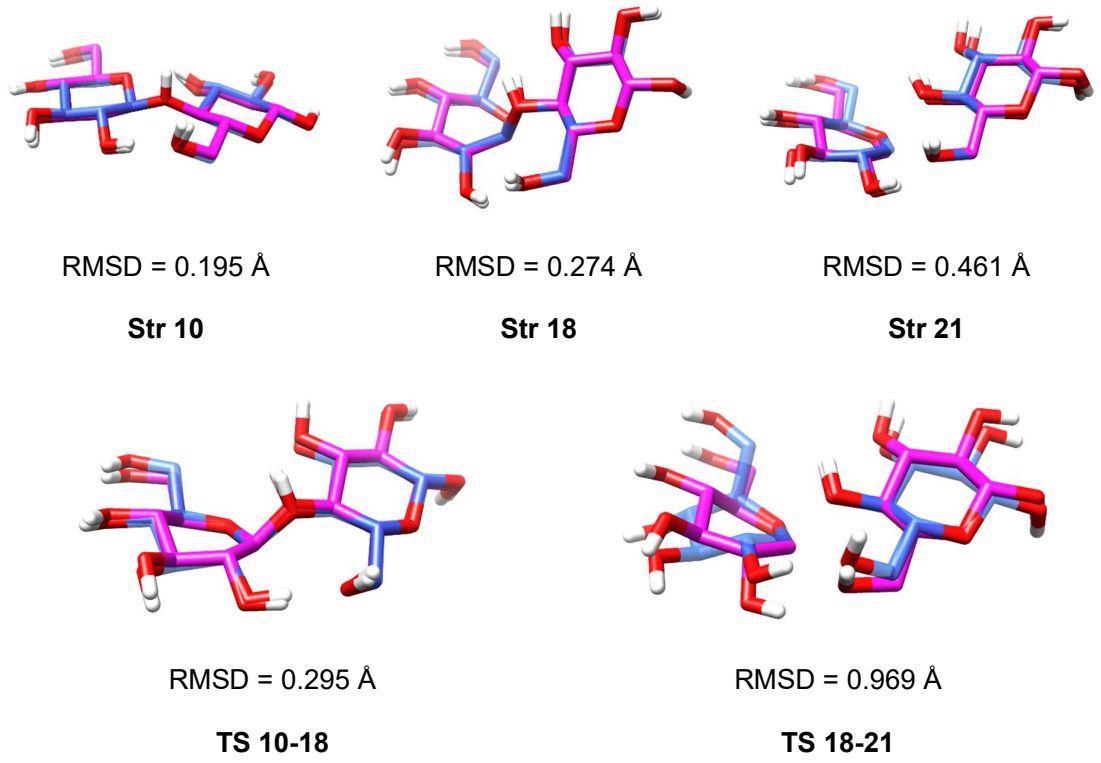


Figure 4.8 SMD (blue) water BB1K/def2-TZVP generated structures superimposed on C-PCM (pink) water generated structures calculated at the same level of theory.

The ring shifts begin from the first transition state. Str 10 remains largely the same in terms of ring orientation. Thus, the hydrogen bonding remains the same for both solvent models but on average the SMD model produces longer hydrogen bonds, with the major exception being the $O_{3'}H_{3'}\cdots O_5$ bond, which shortens significantly (0.1 Å). This correlates with its low RMSD (C, O, H) relative to the other shifted structures. More crucial is that both TS 10-18 and Str 18 show a shift in the orientation of the reducing ring and its hydroxyl groups between solvent

models. This shifts the glycosidic linkage such that it can no longer form a hydrogen bond with the non-reducing ring cyclic oxygen (O_5) and consequently the SMD structures have one hydrogen bond less.

Large variation in hydrogen bond lengths is seen, specifically for TS 10-18, where bond lengths in SMD solvent are ~ 0.23 Å longer. A further possible change in ring orientation can be observed for the bond breaking transition state and product. This shift affects the non-reducing ring more; specifically, the hydroxy-alkyl sidechain ($C_6O_6H_6$), as the C_5 shifts upwards instead of the C_4 as seen in C-PCM (Figure 4.10.). Consequently, no hydrogen bond is formed between the $O_6H_6\cdots O_4$. However, a $O_1H_{1*}\cdots O_5$ bond is form, which was previously missing in SMD structures. This change is carried through to the final structure, Str 21. Str 21 reforms the $O_6H_6\cdots O_4$ hydrogen bond, making it so that the solvent models produce the same hydrogen bonding network.

4.3.2.2 Effect of Ionic Liquid Solvation on Free Energy Barriers

The previous section closely examined the effects of each of the theoretical components of the computation, indicating how a change in basis set or solvation model can affect the relative energy and structure of a specific step. Of interest was the comparison of C-PCM to the SMD model, as the latter is the preferred solvent model in this work and the only one that has been parameterised for ionic liquids. In this section, the effect of different solvent systems on the cellobiose hydrolysis mechanism is studied at the BB1K/def2-TZVP level, using an SMD or SMD-PGP ionic liquid solvent model. In using the SMD model, the solvent effects can be correlated to shifts in the set of parameters listed in Section 3.2.4 of Chapter 3, as well as subsequent conformational changes. The profiles in ionic liquid are compared to that in water, as this is the solvent for which the mechanism was originally proposed for.

All Loerbroks et al. supporting information structures were initially optimised in gas phase at the BB1K/def2-TZVP level of theory.[24] These gas phase geometries were then used as the input geometries for the various solvent phase optimisations, allowing for an assessment of conformational changes as well as energy changes. However, during this process, it emerged that the gas phase optimised structures varied from the original Loerbroks et al. (2013) structures (Figures 4.6 and 4.9).[24]

The major change stemmed from changes in hydroxyl group(s) orientation that occurred in the initial reactant Str 10. The reducing ring hydroxyls shifted in the newly generated gas structure,

denoted Str 10†. These changes were then maintained throughout solvent phase optimisation, facilitating the creation of two different reaction profiles based on input geometry origins (Figures 4.10). Most importantly, this alternate Str 10 was lower in energy, consistently, in all solvents, while the other minima (Str 18 and Str 21) remained the same.

RMSD comparison of the solvent optimised † and * structures showed that the input geometry only affected the transition states and Str 10 significantly, with the other relative minima remaining the same. These variations correlate with the shifts in barriers that cause the profiles to deviate, specifically favouring † conformers for ionic liquids and * for water.

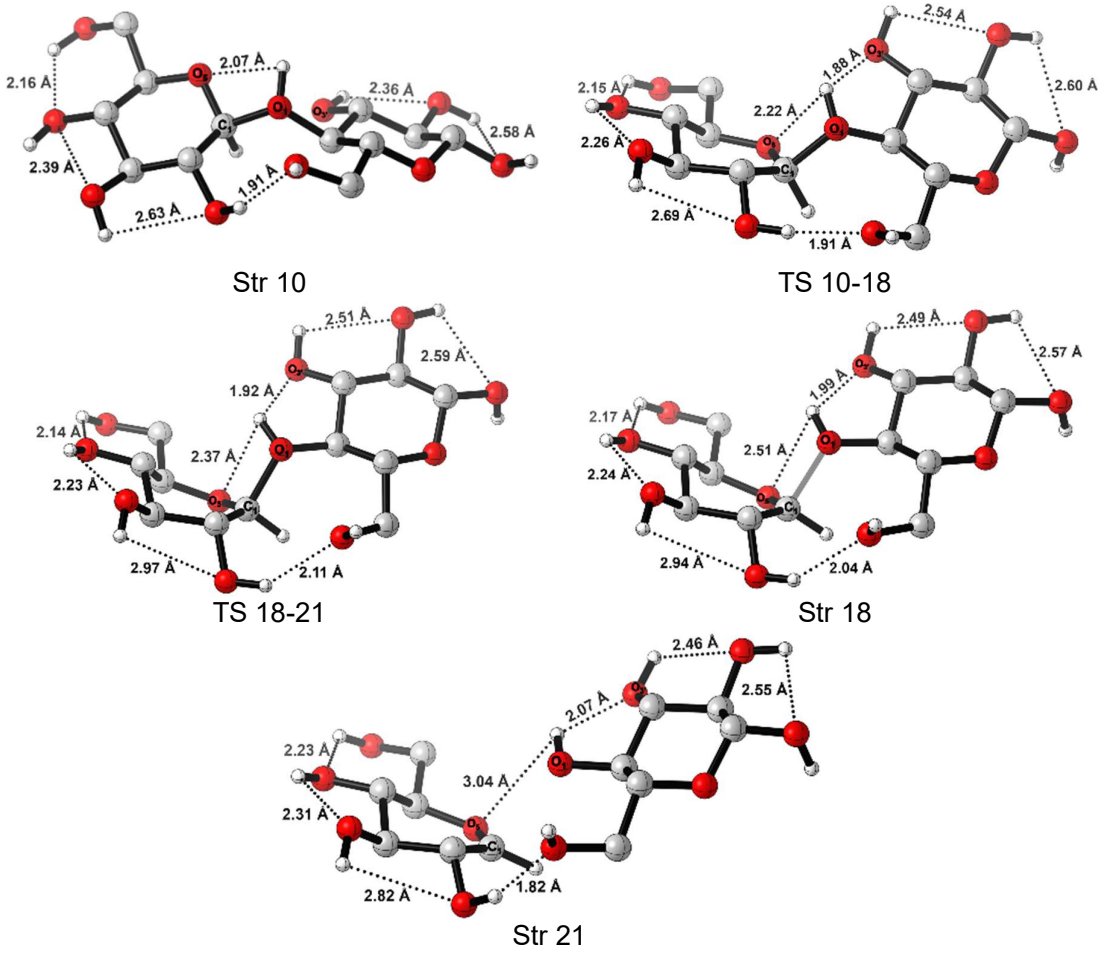


Figure 4.9 Gas phase structures optimised at BB1K/def2-TZVP.[24]

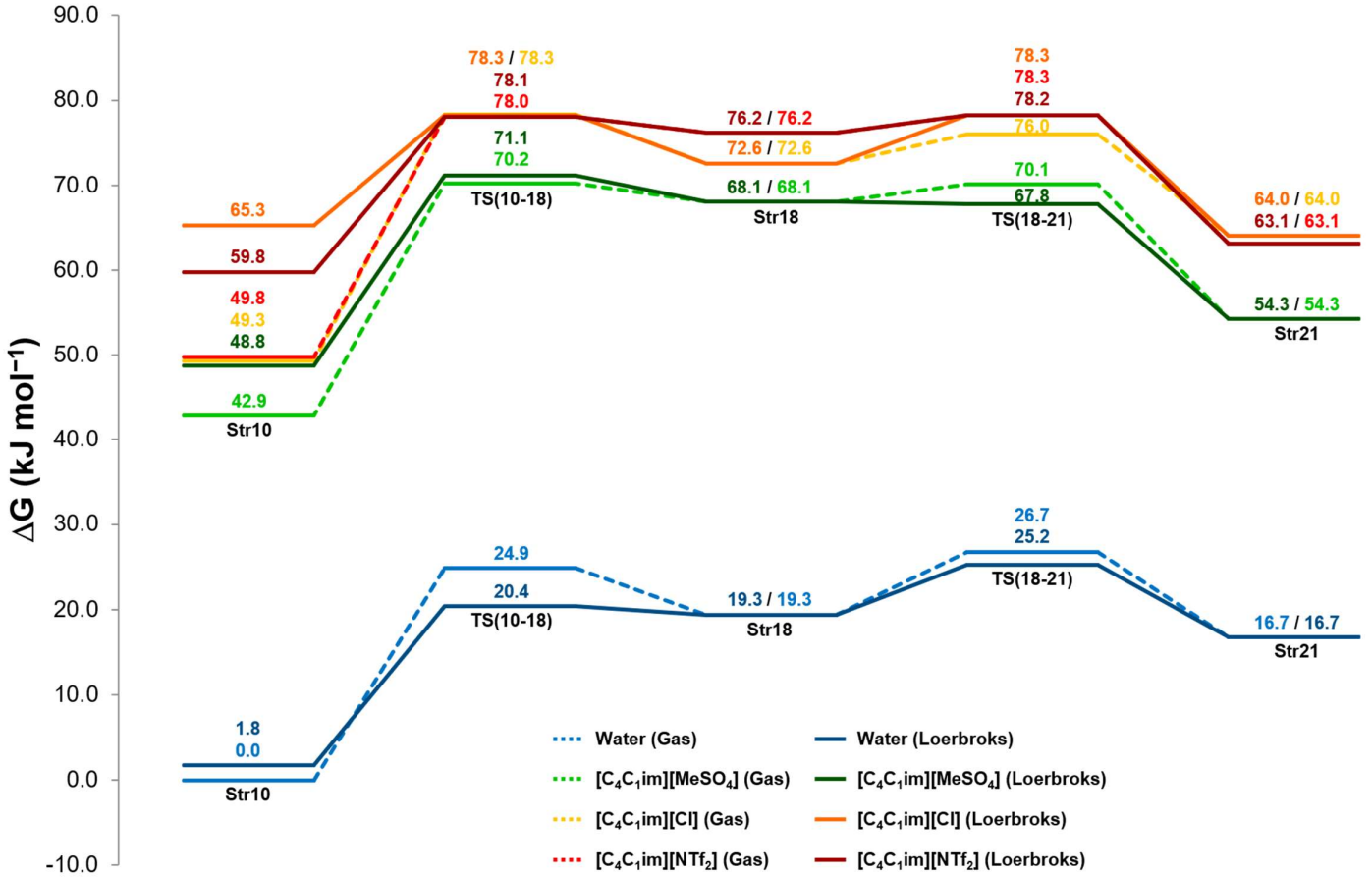


Figure 4.10 Comparison of reaction profiles of the Loerbroks et al. preferred cellobiose hydrolysis mechanism at BB1K/def2-TZVP in various SMD modelled implicit solvents, optimised from gas optimised structural coordinates (---) and Loerbroks et al. (2013) supporting information structural coordinates (—).[24] Structures plotted relative to lowest energy conformer, Str 10†, in water.

Subsequent to this analysis, the lowest energy conformations for each step, regardless of input geometry, were used to generate an overall minimum reaction free energy reaction profile within each solvent (Figure 4.11). All structures (Str 10-21, including transition states) were plotted relative to the lowest energy structure overall, which is water optimised Str 10.

The reaction is endergonic overall, with the reactant (Str 10) 14 kJ mol^{-1} lower on average. This contradicts Loerbroks et al.'s original findings where Str 21 (product) was 6 kJ mol^{-1} lower in energy, making the reaction exergonic (Figure 4.5). The conformational change is endergonic with an average activation barrier of 28.2 kJ mol^{-1} for the ionic liquids and 20.4 kJ mol^{-1} for water. This step is far more energy intensive generally ~ 23 kJ mol^{-1} greater than the proceeding exergonic bond breaking.

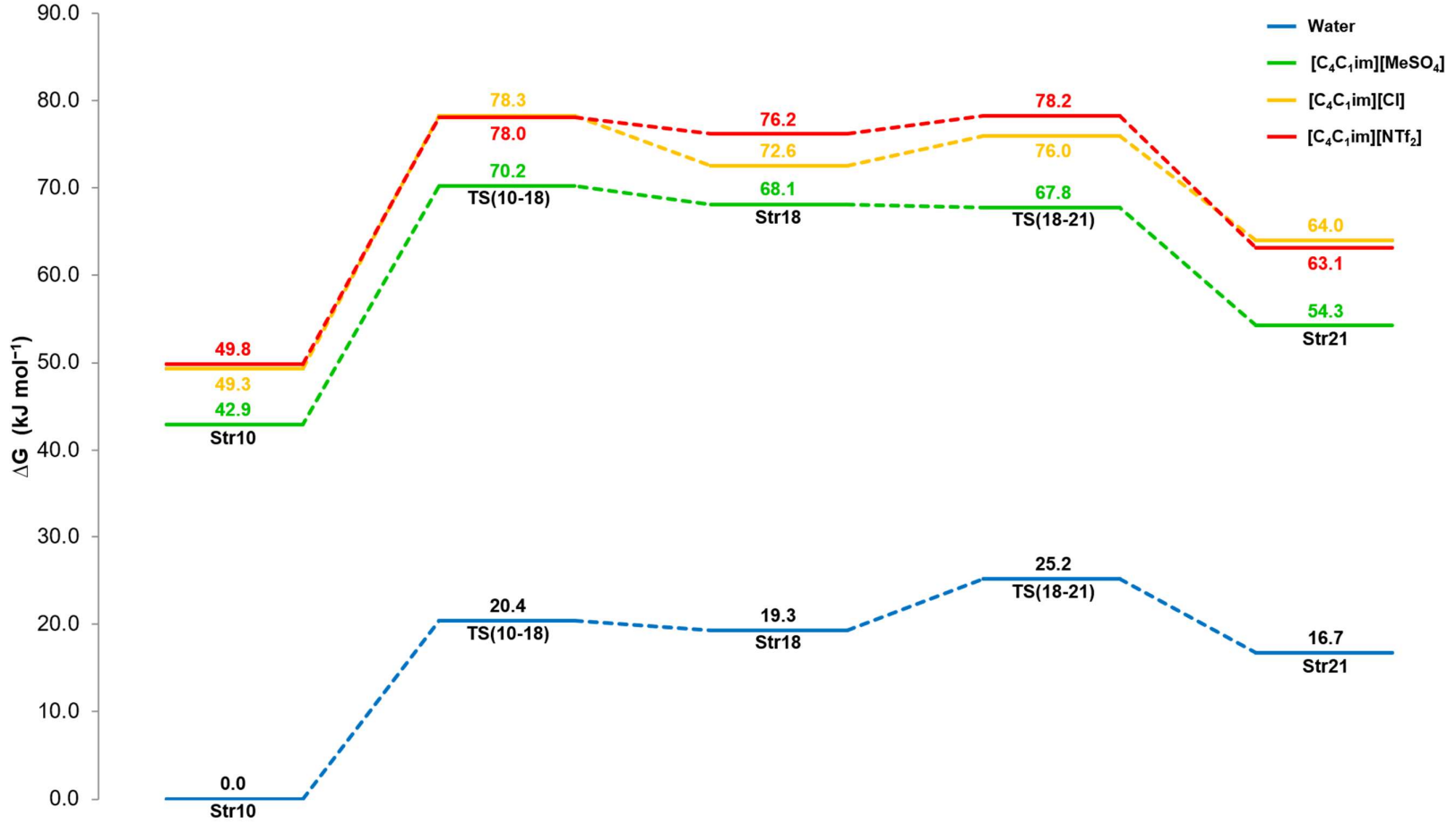


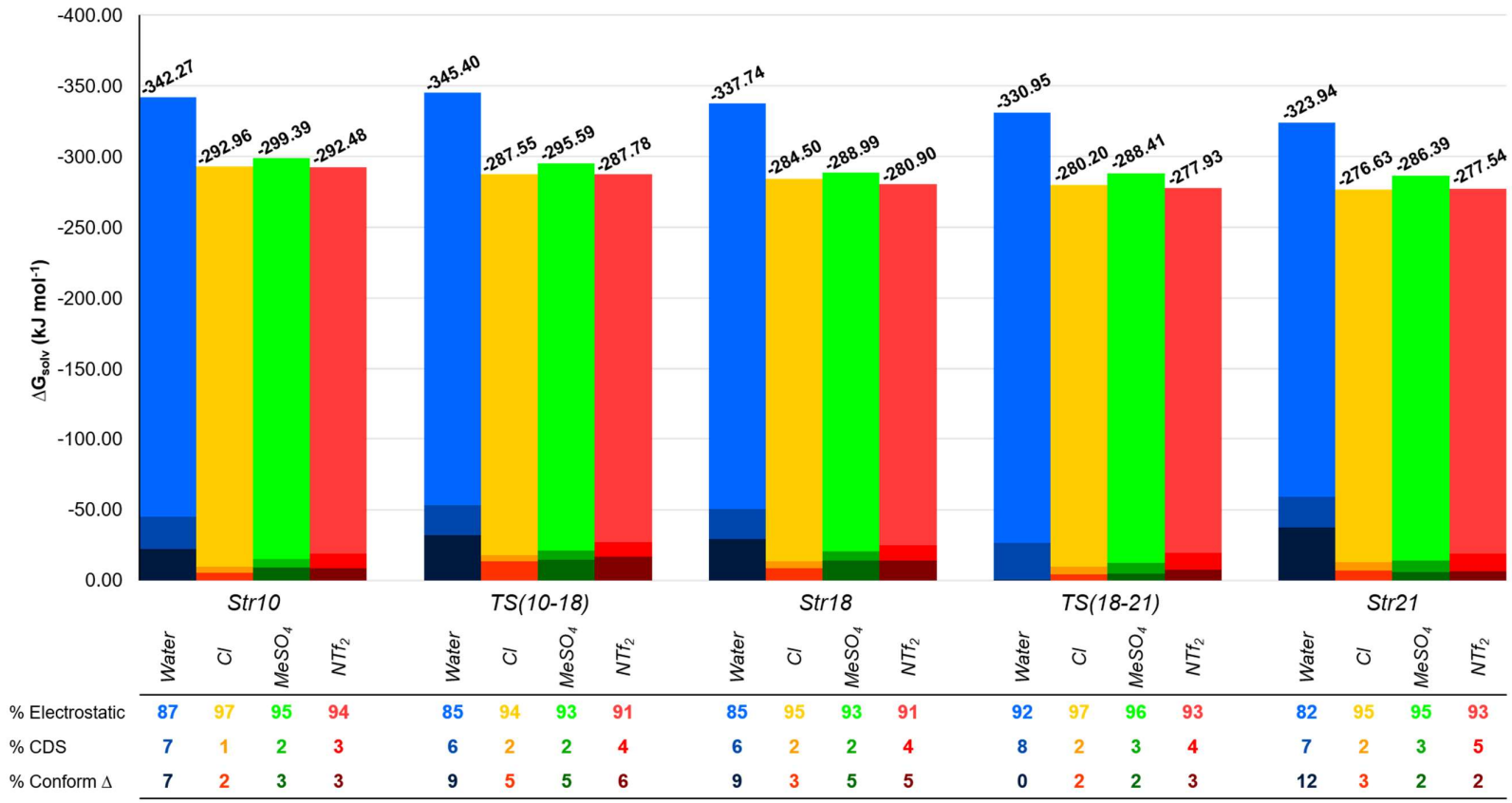
Figure 4.11 Reaction profile of the Loerbroks et al. preferred cellobiose hydrolysis mechanism at BB1K/def2-TZVP in various SMD modelled implicit solvents.[24]

Solvation greatly stabilises the structures compared to gas, lowering each structure by ~300 kJ mol^{-1} on average (Figure 4.12). High variation between gas and solvent phase is expected, given the charged nature of the structures. The water optimised structures are the lowest energy for each step, 40-50 kJ mol^{-1} lower in energy than the ILs (Figure 4.11), resulting in ~ 100 kJ mol^{-1} larger solvation energies thus greater stabilisation (Figure 4.12). This is unsurprising given its high dielectric constant of 78 and the large molecular dipole moments of the structures (Table 4.2). This correlation breaks down for the ILs where $[C_4C_1im][NTf_2]$ and $[C_4C_1im][Cl]$ produce similar values and $[C_4C_1im][MeSO_4]$ is significantly larger than $[C_4C_1im][Cl]$, which do not match the changes in dielectric constant.

Table 4.2 Molecular dipole moments for each mechanism structure optimised in gas and the various implicit solvents of interest at BB1K/def2-TZVP. Dipole moments were calculated in the standard orientation with the centre of charge at the origin.

	Str 10	**TS 10-18**	**Str 18**	**TS 18-21**	**Str 21**
Gas	7.88	5.66	4.89	3.80	3.49
Water	11.15	7.09	6.52	1.83	3.57
$[C_4C_1im][Cl]$	10.33	6.50	5.89	3.31	3.90
$[C_4C_1im][MeSO_4]$	10.41	6.51	5.88	3.39	3.89
$[C_4C_1im][NTf_2]$	10.35	6.61	5.95	3.38	3.81

Figure 4.12 further shows that the solvation energy is dominated by electrostatic contributions, with a percentage contribution of greater than 80% throughout. In the case of the ionic liquids, this shifts to over 90%.



	Str10				TS(10-18)				Str18				TS(18-21)				Str21			
	Water	Cl	$MeSO_4$	NTf_2	Water	Cl	$MeSO_4$	NTf_2	Water	Cl	$MeSO_4$	NTf_2	Water	Cl	$MeSO_4$	NTf_2	Water	Cl	$MeSO_4$	NTf_2
% Electrostatic	87	97	95	94	85	94	93	91	85	95	93	91	92	97	96	93	82	95	95	93
% CDS	7	1	2	3	6	2	2	4	6	2	2	4	8	2	3	4	7	2	3	5
% Conform Δ	7	2	3	3	9	5	5	6	9	3	5	5	0	2	2	3	12	3	2	2

Figure 4.12 Solvation free energies for each structure in the mechanism in various solvents, broken down into various contributions. Contributions listed in the table below the bar chart figure are: electrostatics (Electrostatic); cavitation, dispersion and structural rearrangement of solvent (CDS), and free energy of any conformational changes occurring upon solvation of the gas phase structure (Conform Δ).

All solvent structures have changed significantly under implicit solvation conditions compared to gas phase conditions (Figures 4.9 and 4.13); however, water consistently shows the greatest deviation from the gas phase structures (average RMSD = 0.579 Å). $[C_4C_1im][MeSO_4]$ generally follows (avg RMSD from gas = 0.330 Å), showing greatest resemblance to structures obtained in water (avg RMSD from water = 0.292 Å). However, overall the structures optimised in different IL are all very similar in structure (avg RMSD between $[C_4C_1im][MeSO_4]$ and $[C_4C_1im][Cl]$ or $[C_4C_1im][NTf_2]$ = 0.052 Å and = 0.060 Å). This is despite $[C_4C_1im][Cl]$ and $[C_4C_1im][NTf_2]$ generally bearing greater similarity to the gas phase structures (avg RMSD $[C_4C_1im][Cl]$ = 0.319 Å, and $[C_4C_1im][NTf_2]$ = 0.298 Å) and larger variation from water (avg RMSD $[C_4C_1im][Cl]$ = 0.305 Å, and $[C_4C_1im][NTf_2]$ = 0.317 Å) than noted for $[C_4C_1im][MeSO_4]$. Therefore it can be postulated that they would be higher in energy given these structural findings, and that perhaps $[C_4C_1im][MeSO_4]$ deviates from the expected trend according to dielectric magnitudes for the ionic liquids due to better structural resemblance to the low energy aqueous structures.

Attempts to expand the model to include hydrogen bonding effects by introducing a small cluster of explicit solvent molecules into the implicitly solvated system were unsuccessful. The explicit solvent molecules are intended to represent the first solvation shell and thus perhaps capture some of the key solute-solvent interactions that may not be accurately represented under implicit solvation. The cations and anions were placed according to an electrostatic potential surface. However, upon optimisation the anion moves closer to glycosidic linkage where the proton is transferred to the anion. This was irrespective of anion placement, cation presence and number of solvent molecules present. This is most likely due to low basicity of the glycosidic linkage Oxygen atom (O_1), with the conjugate acids formed (HCl, CH_4O_4S) more stable. Thus, no profile could be established for comparison to the implicit solvent to properly assess extent hydrogen bonding would affect cellobiose solvation energy.

It should however be mentioned that accurate representation of these properties would only relate to the breakup of the full cellulose crystal and not solvation of the individual monomers (cellobiose). Water is unable to break apart cellulose but very capable of dissolving and stabilising cellobiose. These results merely indicate the effect of each solvent system on the thermodynamics of the hydrolysis mechanism, showing IL solvation to be less stabilising resulting in an increase in the key conformation change barrier.

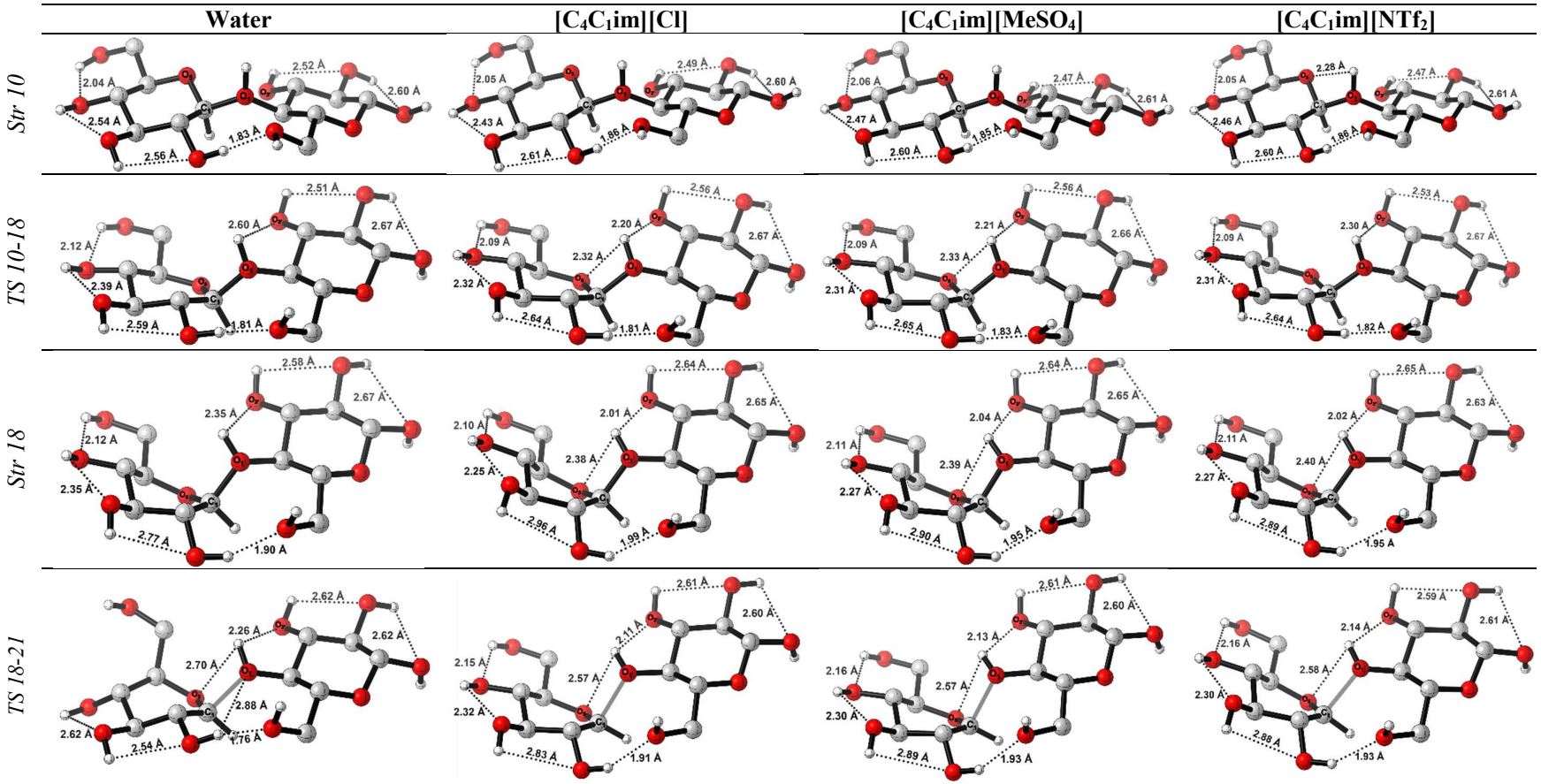


Figure 4.13 Structures with hydrogen bond contact distances shown, optimised at BB1K/def2-TZVP in water, $[C_4C_1im][Cl]$, $[C_4C_1im][MeSO_4]$ and $[C_4C_1im][NTf_2]$. See below for separate figure containing Str 21.

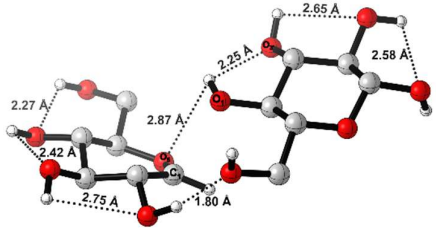
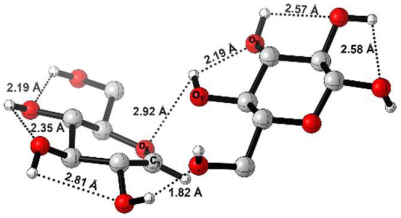
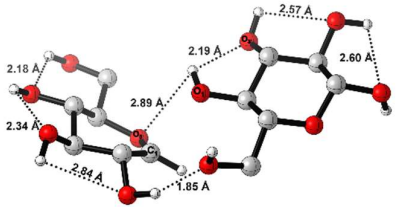
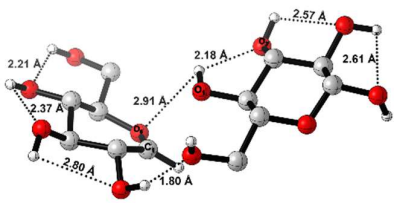
	Water	$[C_4C_1im][Cl]$	$[C_4C_1im][MeSO_4]$	$[C_4C_1im][NTf_2]$
Str. 21				

Figure 4.13 continued.

4.3.3 Correlation to Experimental Yields for Cellulose Hydrolysis

AIL and sulfuric acid [C_4C_1im][Cl] rate-determining free energy barrier (ΔG^{rds}) was approximated by adding the conformational change energy barrier and the protonation free energies together. These values are compared to experimental TRS yields found in literature, [11,38,213] to see if a relationship exists (Table 4.3 and Figures 4.14).

Table 4.3 Experimental total reducing sugar percentage yields and reaction free energies in [C_4C_1im][Cl] calculated using DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP [M05-2X/6-31+G** (SMD)] and BB1K/def2-TZVP for cellobiose acid hydrolysis by sulfuric acid and acid-functionalised cationic acids (HA).

HA	[C_4C_1im][Cl]	
	ΔG^{rds} (kJ mol^{-1})	TRS % Yield[a]
[IL-OH]$^+$	418.21	10[b]
[IL-COOH]$^+$	321.83	25[c]
[IL-SO_3H]$^+$	288.41	85[d]
Sulfuric acid	244.92	92[e]

[a]TRS yields for microcrystalline (MCC) in [C_4C_1im][Cl]. Reaction conditions vary as best possible yield for the same cellulose reactant was used. Values for similar conditions to IL-SO3H shown in brackets. See individual references listed below for further reaction conditions.
[b]Ref 38
[c]Ref 38
[d]Ref 38
[e]Ref 213

Reaction conditions were chosen to optimise yield as no exact match in reaction conditions could be found for sulfuric acid, making direct comparison impossible It has been found that reaction conditions other than catalyst and solvent also play a significant role in performance. Variations in temperature and duration have both been shown to cause significant variations in performance for an acid/solvent system to a certain point. This limit is generally imposed by degradation reactions, where yields lower as the glucose or sugar derivatives produced are further reacted to other derivatives.

Under optimum reaction conditions, sulfuric acid produces the highest total reducing sugar yield and has the lowest overall barrier. This is followed by [IL-SO_3H]$^+$, [IL-COOH]$^+$ and lastly [IL-OH]$^+$. This trend is mimicked in terms of acidity, where the lower the Hammett acidity function (higher acidity), the higher the TRS yield (Figure 4.14).

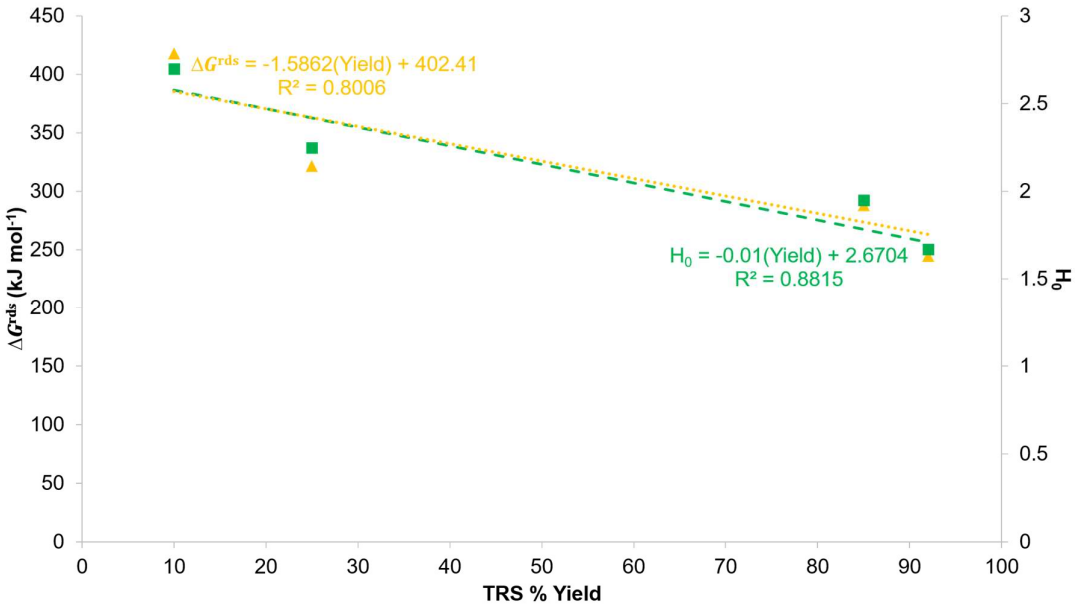


Figure 4.14 Plot illustrating correlation between optimum cellobiose hydrolysis total reducing sugar percentage yields for the AILs and sulfuric acid in [C_4C_1im][Cl] and experimentally determined Hammett acidity function (H_0, in water using 4-nitroaniline) and ΔG^{rds} calculated using a DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP [M05-2X/6-31+G** (SMD)] PX cycle and BB1K/def2-TZVP (SMD).[38,213,214]

This shows an inverse relationship for both, implying the greater the Hammett acidity magnitude the lower the yield and the lower the yield the higher the ΔG^{rds}. This points toward a link between catalyst acidity and ΔG^{rds} through the similar behaviour related to the yield. Comparison of both experimental (Hammett acidity functions, H_0) and predicted acidity measures (PX predicted pK_as from Chapter 3) to the calculated ΔG^{rds} confirm this relationship, with a near linear correlation (Figure 4.15).

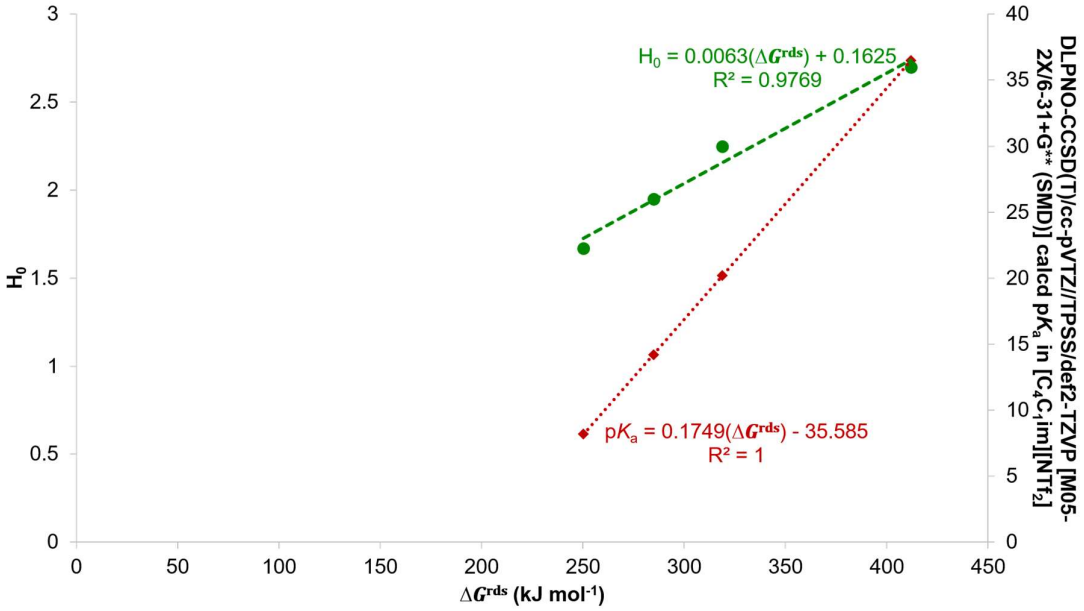


Figure 4.15 Plot illustrating correlation between DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP [M05-2X/6-31+G** (SMD)] PX cycle and BB1K/def2-TZVP (SMD) calculated cellobiose hydrolysis ΔG^{rds} in [C_4C_1im][NTf_2] and experimentally determined Hammett acidity function (H_0, in water using 4-nitroaniline) and DLPNO-CCSD(T)/cc-pVTZ//TPSS/def2-TZVP [M05-2X/6-31+G** (SMD)] PX cycle predicted pK_as. [38,213,214]

The sulfonyl functionalised AIL ([IL-SO_3H]$^+$) more closely resembles sulfuric acid in TRS yield performance than the other AILs. This matches pK_a predictions, which showed [IL-SO_3H]$^+$ to be far more acidic than the other AILs, more closely resembling butane-1-sulfonic acid and sulfuric acid. This correlates with experimental studies which show sulfonyl-functionalised ionic liquid acid catalysts to generate far higher yields than other acidic functionalities.[11,38,131,231] However, the sulfonic AIL a yield in [C_4C_1im][Cl] very close to that of sulfuric acid, despite the relatively large difference in acidity. A possible reason could be the aforementioned reaction conditions or the increased flexibility or mobility of the acidic group in the [IL-SO_3H]$^+$. The sulfonyl group is attached at the end of a four Carbon atom alkyl chain which allows separation from the bulkier imidazolium ring and increased rotational flexibility, possibly resulting in easier access to glycosidic linkages.[11,39]

4.4 Conclusion

Replication of the Loerbroks et al. study showed changes in the solvent model to most greatly affect the free energy. SMD solvation resulted in a far more stabilised reactant, leading to an

increase in the conformation change free energy barrier and the overall reaction energy change. Originally exergonic, the SMD stabilised Str 10 was lower than the product (Str 21) making the reaction endergonic, which was maintained across all SMD modelled solvents. The energetics of the barriers themselves remained the same, with conformational change being endergonic and far larger than the exergonic bond breaking step.

As expected for implicit models, electrostatic contributions dominate solvation energies, leading to water producing the lowest energy conformations and most negative solvation free energies. This translated into the key energy intensive steps, protonation and conformational change, being significantly lower for water compare to the ILs. However, amongst the ILs the same trend could not be seen, resulting in little variation between ILs in terms of barrier heights and protonation energies. Given the size of the protonation contribution relative to conformational and bond breaking barriers, it is clearly the greatest determinant of the reaction thermodynamic viability. This translates to trends in experimental yields, which correlate well with AIL acidity trends.

5.

Final Remarks

This study aimed to investigate, through quantum chemical computations, the effect of various ionic liquids on cellulose acid hydrolysis. Using the cellobiose repeating unit as a representative model and selecting a specific mechanistic pathway from literature, the reaction could be separated into two parts based on energy requirements. The first addressed catalytic activity in terms of acidity or proton donor ability, as protonation of the glycosidic linkage is thought to be the greatest contributor to the activation energy. This involved predicting pK_a values for Brønsted functionalised acidic ionic liquids that have been used as catalysts in cellulose hydrolysis. Given the lack of experimentally determined acidic ionic liquid pK_a values *in* ionic liquids, a prediction technique needed to be established and benchmarked. Thus, a series of thermodynamic cycle protocols, consisting of a gas and solvent phase level of theory combination and specific TC type, were benchmarked using a series of N-base amine conjugate acids. Overall the proton exchange cycle proved most accurate, with the direct cycle greatly inhibited by inaccuracies in the predicted proton solvation free energies. The most accurate protocol consisted of a proton exchange cycle with a G3B3(MP2) gas leg and M05-2X/6-31+G** for implicit solvation. However, computational efficiency greatly favoured the use of DLPNO-CCSD(T)/cc-pVYZ//TPSS/def2-TZVP, which showed reasonable mean absolute errors in combination with M05-2X/6-31+G** for solvation. Predicted acidities matched trends in experimentally determined Hammett acidity functions. Moreover, the sulfonyl functionalisation was shown to greatly improve acidity, with an acidity similar to popular molecular acids.

Of the changes made from the original Loerbroks et al. study, the greatest effect was seen in the change from C-PCM to SMD. In terms of overall reaction free energy changes, SMD modelling predicted an endergonic reaction, greatly stabilising the reactant (Str 10), which contradicted the C-PCM exergonic profile. These can be linked to significant structural changes in the conformers. Compared to $[C_4C_1im][MeSO_4]$, $[C_4C_1im][Cl]$ and $[C_4C_1im][NTf_2]$, water solvated structures were far lower in energy, leading to an exergonic protonation and lower conformational change barrier. Combined these steps gives the rate-determining barrier. These

results could be a by-product of using an implicit solvent model, where solvation is predominantly determined by electrostatic contributions. However, barriers for the ionic liquids remained relatively similar, despite changes in dielectric constant and overall only differed from water by approximately 8 and 4 kJ mol^{-1} for the conformational and bond break steps, respectively. Of greater significance was variations in the significantly larger protonation free energy, where IL-SO_3H energies significantly lower in all solvents. This matched trends in experimental yields, indicating that the effect of the catalyst is far greater than that of the solvent, with issues in solvation primarily due to the supramolecular structure of cellulose and not the hydrolysis of the individual cellobiose glycosidic linkages.

REFERENCE LIST

(1) Uçkun Kıran, E.; Trzcinski, A. P.; Liu, Y. *J. Chem. Technol. Biotechnol.* **2015**, *90* (8), 1364–1379.

(2) Isikgor, F. H.; Becer, C. R. *Polym. Chem.* **2015**, *6* (25), 4497–4559.

(3) Rinaldi, R.; Schüth, F. *ChemSusChem* **2009**, *2* (12), 1096–1107.

(4) Sun, Y.; Cheng, J. *Bioresour. Technol.* **2002**, *83* (1), 1–11.

(5) Fleming, K. L.; Pfaendtner, J. *J. Phys. Chem. A* **2013**, *117* (51), 14200–14208.

(6) Collins, T.; Gerday, C.; Feller, G. *FEMS Microbiol. Rev.* **2005**, *29* (1), 3–23.

(7) Li, J. J.; Li, J. J.; Zhang, D.; Liu, C. *ChemPhysChem* **2015**, *16* (14), 3044–3048.

(8) Loerbroks, C.; Heimermann, A.; Thiel, W. *J. Comput. Chem.* **2015**, *36* (15), 1114–1123.

(9) Notley, S. M.; Pettersson, B.; Wa, L.; Wågberg, L. *J. Am. Chem. Soc.* **2004**, *126* (43), 13930–13931.

(10) Da Costa Lopes, A. M.; Bogel-Lukasik, R. *ChemSusChem* **2015**, *8* (6), 947–965.

(11) Amarasekara, A. S.; Owereh, O. S. *Ind. Eng. Chem. Res.* **2009**, *48* (22), 10152–10155.

(12) Anderson, J. L.; Armstrong, D. W.; Wei, G.-T. *Anal. Chem.* **2007**, *79* (11), 4247–4253.

(13) Swatloski, R. P.; Spear, S. K.; Holbrey, J. D.; Rogers, R. D. *J. Am. Chem. Soc.* **2002**, *124* (18), 4974–4975.

(14) Vanoye, L.; Fanselow, M.; Holbrey, J. D.; Atkins, M. P.; Seddon, K. R. *Green Chem.* **2009**, *11* (3), 390–396.

(15) Nishimura, Y.; Yokogawa, D.; Irle, S. *Chem. Phys. Lett.* **2014**, *603*, 7–12.

(16) Jiang, F.; Zhu, Q.; Ma, D.; Liu, X.; Han, X. *J. Mol. Catal. A Chem.* **2011**, *334* (1–2), 8–12.

(17) Liu, Y.; Xiao, W.; Xia, S.; Ma, P. *Carbohydr. Polym.* **2013**, *92* (1), 218–222.

(18) Bernales, V. S.; Marenich, A. V.; Contreras, R.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2012**, *116* (30), 9122–9129.

(19) Volynets, B.; Ein-Mozaffari, F.; Dahman, Y. *Green Process. Synth.* **2017**, *6* (1), 1–22.

(20) Aso, T.; Koda, K.; Kubo, S.; Yamada, T.; Nakajima, I.; Uraki, Y. *J. Wood Chem. Technol.* **2013**, *33* (4), 286–298.

(21) Zhu, H.; Luo, W.; Ciesielski, P. N.; Fang, Z.; Zhu, J. Y.; Henriksson, G.; Himmel, M. E.; Hu, L. *Chem. Rev.* **2016**, *116* (16), 9305–9374.

(22) da Costa Lopes, A. M.; João, K. G.; Rubik, D. F.; Bogel-Łukasik, E.; Duarte, L. C.; Andreaus, J.; Bogel-Łukasik, R. *Bioresour. Technol.* **2013**, *142*, 198–208.

(23) Payal, R. S.; Bharath, R.; Periyasamy, G.; Balasubramanian, S. *J. Phys. Chem. B* **2012**, *116* (2), 833–840.

(24) Loerbroks, C.; Rinaldi, R.; Thiel, W. *Chem. - A Eur. J.* **2013**, *19* (48), 16282–16294.

(25) Chen, P.; Nishiyama, Y.; Putaux, J.-L.; Mazeau, K. *Cellulose* **2014**, *21* (2), 897–908.

(26) Lee, H. V.; Hamid, S. B. A.; Zain, S. K. *Sci. World J.* **2014**, 1–20.

(27) Loerbroks, C. Acid Hydrolysis of Cellulose and the Anomeric Effect : A Computational Study. PhD Thesis, Heinrich Heine University Düsseldorf : Düsseldorf, Germany, 2014.

(28) Pinkert, A.; Marsh, K. N.; Pang, S.; Staiger, M. P. *Chem. Rev.* **2009**, *109* (12), 6712–6728.

(29) Lindman, B.; Karlström, G.; Stigsson, L. *J. Mol. Liq.* **2010**, *156* (1), 76–81.

(30) Brandt, A.; Gräsvik, J.; Hallett, J. P.; Welton, T. *Green Chem.* **2013**, *15* (3), 550.

(31) Welton, T. *Chem. Rev.* **1999**, *99* (8), 2071–2084.

(32) Amarasekara, A. S. *Chem. Rev.* **2016**, *116* (10), 6133–6183.

(33) Hallett, J. P.; Welton, T. *Chem. Rev.* **2011**, *111* (5), 3508–3576.

(34) Du, H.; Qian, X. *Carbohydr. Res.* **2011**, *346* (13), 1985–1990.

(35) Cao, Y.; Zhang, R.; Cheng, T.; Guo, J.; Xian, M.; Liu, H. *Appl. Microbiol. Biotechnol.* **2017**, *101* (2), 521–532.

(36) Gupta, K. M.; Jiang, J. *Chem. Eng. Sci.* **2015**.

(37) Gold, V. *Pure Appl. Chem.* **1979**, *51* (8), 1725–1801.

(38) Parveen, F.; Patra, T.; Upadhyayula, S. *Carbohydr. Polym.* **2016**, *135*, 280–284.

(39) Zhuo, K.; Du, Q.; Bai, G.; Wang, C.; Chen, Y.; Wang, J. *Carbohydr. Polym.* **2015**, *115*, 49–53.

(40) Hara, M. *Energy Environ. Sci.* **2010**, *3* (5), 601–607.

(41) Li, C.; Wang, Q.; Zhao, Z. K. *Green Chem.* **2008**, *10* (2), 177–182.

(42) Koshland, D. E. *Biol. Rev.* **1953**, *28* (4), 416–436.

(43) Biarnés, X.; Ardèvol, A.; Iglesias-Fernández, J.; Planas, A.; Rovira, C. *J. Am. Chem. Soc.* **2011**, *133* (50), 20301–20309.

(44) Satoh, H.; Hansen, H. S.; Manabe, S.; van Gunsteren, W. F.; Hünenberger, P. H. *J. Chem. Theory Comput.* **2010**, *6* (6), 1783–1797.

(45) Kayser, H.; Rodríguez-Ropero, F.; Leitner, W.; Fioroni, M.; María, P. D. de. *RSC Adv.* **2013**, *3* (24), 9273–9278.

(46) Beck, J. M.; Miller, S. M.; Peczuh, M. W.; Hadad, C. M. *J. Org. Chem.* **2012**, *77* (9), 4242–4251.

(47) Krässig, H.; Schurz, J.; Steadman, R. G.; Schliefer, K.; Albrecht, W.; Mohring, M.; Schlosser, H. Cellulose. In *Ullmann's Encyclopedia of Industrial Chemistry*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2004.

(48) Ghasemi, M.; Tsianou, M.; Alexandridis, P. *Bioresour. Technol.* **2017**, *228*, 330–338.

(49) Medronho, B.; Lindman, B. *Adv. Colloid Interface Sci.* **2015**, *222*, 502–508.

(50) Zhang, L.; Ruan, D.; Gao, S. *J. Polym. Sci. Part B Polym. Phys.* **2002**, *40* (14), 1521–1529.

(51) Klemm, D.; Philipp, B.; Heinze, T.; Heinze, U.; Wagenkencht, W. *Comprehensive Cellulose Chemistry. Volume 1. Fundamentals and Analytical Methods*; Wiley-VCH Verlag GmbH: Weinheim, Germany, 1998; Vol. 1.

(52) Klemm, D.; Heublein, B.; Fink, H. P.; Bohn, A. *Angew. Chemie - Int. Ed.* **2005**, *44* (22), 3358–3393.

(53) Graenacher, C. Cellulose Solution. 1943176, 1934.

(54) Rogers, R. D.; Seddon, K. R. *Science*. October 31, 2003, pp 792–793.

(55) Remsing, R. C.; Swatloski, R. P.; Rogers, R. D.; Moyna, G. *Chem. Commun.* **2006**, 1271–1273.

(56) Guo, J.; Zhang, D.; Duan, C.; Liu, C. *Carbohydr. Res.* **2010**, *345* (15), 2201–2205.

(57) Xu, H.; Pan, W.; Wang, R.; Zhang, D.; Liu, C. *J. Comput. Aided. Mol. Des.* **2012**, *26* (3), 329–337.

(58) Lu, B.; Xu, A.; Wang, J. *Green Chem.* **2014**, *16* (3), 1326–1335.

(59) Cramer, C. J. Quantum Mechanics. In *Essentials of Computational Chemistry: Theories and Models*; John Wiley & Sons: West Sussex, England, 2004; pp 4–5.

(60) Schrödinger, E. *Phys. Rev.* **1926**, *28* (6), 1049–1070.

(61) Born, M.; Oppenheimer, R. *Ann. Phys.* **1927**, *389* (20), 457–484.

(62) Jensen, F. The Adiabatic and Born-Oppenheimer Approximations. In *Introduction to Computational Chemistry*; Wiley: Chichester, UK, 2017; pp 90–94.

(63) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*, 2nd ed.; John Wiley & Sons, Ltd: West Sussex, England, 2004.

(64) Szabo, A.; Ostlund, N. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications: Mineola, N.Y, 1996.

(65) Jensen, F. Electron Correlation Methods. In *Introduction to Computational Chemistry*; Wiley: Chichester, UK, 2017; pp 124–125.

(66) Cremer, D. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3* (5), 482–503.

(67) Cramer, C. J. Perturbation Theory. In *Essentials of Computational Chemistry: Theories and Models*; John Wiley & Sons, Ltd: West Sussex, England, 2004; pp 216–224.

(68) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46* (7), 618–622.

(69) Binkley, J. S.; Pople, J. A. *Int. J. Quantum Chem.* **1975**, *9* (2), 229–236.

(70) Forsberg, B.; He, Z.; He, Y.; Cremer, D. *Int. J. Quantum Chem.* **2000**, *76* (3), 306–330.

(71) Riplinger, C.; Neese, F. *J. Chem. Phys.* **2013**, *138* (3), 034106.

(72) Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. *J. Chem. Phys.* **2016**, *144* (2), 024109.

(73) Neese, F.; Hansen, A.; Liakos, D. G. *J. Chem. Phys.* **2009**, *131* (6), 064103.

(74) Saitow, M.; Becker, U.; Riplinger, C.; Valeev, E. F.; Neese, F. *J. Chem. Phys.* **2017**, *146* (16), 164105.

(75) Boys, S. F. *Rev. Mod. Phys.* **1960**, *32* (2), 296–299.

(76) Liakos, D. G.; Sparta, M.; Kesharwani, M. K.; Martin, J. M. L.; Neese, F. *J. Chem. Theory Comput.* **2015**, *11* (4), 1525–1539.

(77) Neese, F.; Wennmohs, F.; Hansen, A. *J. Chem. Phys.* **2009**, *130* (11), 114108.

(78) Pople, J. A. *J. Chem. Phys.* **1965**, *43* (10), 228–230.

(79) Karplus, M. *J. Phys. Chem.* **1990**, *94* (14), 5435–5436.

(80) Vereecken, L.; Francisco, J. S. *Chem. Soc. Rev.* **2012**, *41* (19), 6259–6293.

(81) Jensen, F. Composite Extrapolation Procedures. In *Introduction to Computational Chemistry*; Wiley: Chichester, UK, 2017; pp 215–221.

(82) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1* (5), 810–825.

(83) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. *J. Chem. Phys.* **1989**, *90* (10), 5622–5629.

(84) Pople, J. A.; Luke, B. T.; Frisch, M. J.; Binkley, J. S. *J. Phys. Chem.* **1985**, *89* (11), 2198–2203.

(85) Curtiss, L. A.; Pople, J. A. *J. Phys. Chem.* **1988**, *92* (4), 894–899.

(86) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2005**, *123* (12), 124107.

(87) Lewars, E. G. Thermodynamics: High Accuracy Calculations. In *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*; Springer Netherlands: Dordrecht, 2011; pp 309–313.

(88) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109* (18), 7764–7776.

(89) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *Chem. Phys. Lett.* **1997**, *270* (5–6), 419–426.

(90) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Baboul, A. G.; Pople, J. A. *Chem. Phys. Lett.* **1999**, *314* (1–2), 101–107.

(91) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *127* (12), 124105.

(92) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1999**, *110* (10), 4703–4709.

(93) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *126* (8), 084108.

(94) Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **1999**, *110* (16), 7650–7657.

(95) Nyden, M. R.; Petersson, G. A. *J. Chem. Phys.* **1981**, *75* (4), 1843–1862.

(96) Petersson, G. A.; Al-Laham, M. A. *J. Chem. Phys.* **1991**, *94* (9), 6081–6090.

(97) Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A. *J. Chem. Phys.* **1996**, *104* (7), 2598–2619.

(98) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94* (11), 7221–7230.

(99) Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **2000**, *112* (15), 6532–6542.

(100) Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999**, *110* (6), 2822–2827.

(101) Montgomery, J. A.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1994**, *101* (7), 5900–5909.

(102) Koch, W.; Holthausen, M. C. The Thomas-Fermi Model. In *A Chemist's Guide to Density Functional Theory*; Wiley: Weinheim, FRG, 2001; pp 30–31.

(103) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136* (3B), B864–B871.

(104) Cramer, C. J. The Hohenberg-Kohn Existence Theorem. In *Essentials of Computational Chemistry: Theories and Models*; John Wiley & Sons, Ltd: West Sussex, England, 2004;

pp 252–254.

(105) Koch, W.; Holthausen, M. C. Orbitals and the Non-interacting Reference. In *A Chemist's Guide to Density Functional Theory*; Wiley: Weinheim, FRG, 2001; pp 41–43.

(106) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140* (4A), A1133–A1138.

(107) Jensen, F. Exchange-Correlation Functionals. In *Introduction to Computational Chemistry*; Wiley: Chichester, UK, 2017; pp 244–258.

(108) Kurth, S.; Perdew, J. P.; Blaha, P. *Int. J. Quantum Chem.* **1999**, *75* (4–5), 889–909.

(109) Perdew, J. P. In *AIP Conference Proceedings*; AIP, 2001; Vol. 577, pp 1–20.

(110) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91* (14), 146401.

(111) Becke, A. D. *Phys. Rev. A* **1988**, *38* (6), 3098–3100.

(112) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37* (2), 785–789.

(113) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58* (8), 1200–1211.

(114) Becke, A. D. *J. Chem. Phys.* **1993**, *98* (2), 1372–1377.

(115) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98* (45), 11623–11627.

(116) Jensen, F. Solvation Models. In *Introduction to Computational Chemistry*; Wiley: Chichester, UK, 2017; pp 502–511.

(117) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99* (8), 2161–2200.

(118) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55* (3), 379-400.

(119) Hermann, R. B. *J. Phys. Chem.* **1972**, *76* (19), 2754–2759.

(120) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799–805.

(121) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24* (6), 669–681.

(122) Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *Chem. Phys. Lett.* **1996**, *255* (4–6), 327–335.

(123) Pierotti, R. A. *Chem. Rev.* **1976**, *76* (6), 717–726.

(124) Floris, F.; Tomasi, J. *J. Comput. Chem.* **1989**, *10* (5), 616–627.

(125) Floris, F. M.; Tomasi, J.; Ahuir, J. L. P. *J. Comput. Chem.* **1991**, *12* (7), 784–791.

(126) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102* (11), 1995–2001.

(127) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113* (18), 6378–6396.

(128) Jensen, F. *Introduction to Computational Chemistry*, 3rd ed.; Wiley: Chichester, UK, 2017.

(129) Amarasekara, A. S.; Shanbhag, P. *Bioenergy Res.* **2013**, *6* (2), 719–724.

(130) Kuroda, K.; Inoue, K.; Miyamura, K.; Satria, H.; Takada, K.; Ninomiya, K.; Takahashi, K. *Catalysts* **2017**, *7* (4), 108.

(131) Amarasekara, A. S.; Wiredu, B. *Int. J. Carbohydr. Chem.* **2012**, *2012*, 1–6.

(132) Thomazeau, C.; Olivier-Bourbigou, H.; Magna, L.; Luts, S.; Gilbert, B. *J. Am. Chem. Soc.* **2003**, *125* (18), 5264–5265.

(133) Fei, Z.; Zhao, D.; Geldbach, T. J.; Scopelliti, R.; Dyson, P. J. *Chem. - A Eur. J.* **2004**, *10* (19), 4886–4893.

(134) Himmel, D.; Goll, S. K.; Scholz, F.; Radtke, V.; Leito, I.; Krossing, I. *ChemPhysChem* **2015**, *16* (7), 1428–1439.

(135) Ho, J.; Coote, M. L. *Theor. Chem. Acc.* **2010**, *125* (1–2), 3–21.

(136) Bryantsev, V. S.; Diallo, M. S.; Goddard III, W. A. *J. Phys. Chem. B* **2008**, *112* (32), 9709–9719.

(137) Ho, J.; Coote, M. L. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1* (5), 649–660.

(138) Ho, J. *Phys. Chem. Chem. Phys.* **2015**, *17* (4), 2859–2868.

(139) Alongi, K. S.; Shields, G. C. *Annu. Rep. Comput. Chem.* **2010**, *6* (C), 113–138.

(140) Ben-Naim, A. *J. Phys. Chem.* **1978**, *82* (7), 792–803.

(141) Sutton, C. C. R. R.; Franks, G. V.; da Silva, G. *J. Phys. Chem. B* **2012**, *116* (39), 11999–

12006.

(142) Toth, A. M.; Liptak, M. D.; Phillips, D. L.; Shields, G. C. *J. Chem. Phys.* **2001**, *114* (10), 4595-4606.

(143) Liptak, M. D.; Shields, G. C. *Int. J. Quantum Chem.* **2001**, *123* (30), 727–741.

(144) Ho, J.; Ertem, M. Z. *J. Phys. Chem. B* **2016**, *120* (7), 1319–1329.

(145) Camaioni, D. M.; Schwerdtfeger, C. A. *J. Phys. Chem. A* **2005**, *109* (47), 10795–10797.

(146) Ho, J.; Coote, M. L. *J. Chem. Theory Comput.* **2009**, *5* (2), 295–306.

(147) Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41* (6), 760–768.

(148) Pliego, J. R.; Riveros, J. M. *J. Phys. Chem. A* **2001**, *105* (30), 7241–7247.

(149) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105* (8), 2999–3094.

(150) Hehre, W. J.; Ditchfield, R.; Radom, L.; Pople, J. A. *J. Am. Chem. Soc.* **1970**, *92* (16), 4796–4801.

(151) Wheeler, S. E.; Houk, K. N.; Schleyer, P. von R.; Allen, W. D. *J. Am. Chem. Soc.* **2009**, *131* (7), 2547–2560.

(152) Wheeler, S. E. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2* (2), 204–220.

(153) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R. *J. Phys. Chem. A* **1998**, *102* (40), 7787–7794.

(154) Pliego Jr, J. R.; Riveros, J. M. *Phys. Chem. Chem. Phys.* **2002**, *4* (9), 1622–1627.

(155) Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95* (14), 5610–5620.

(156) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2006**, *110* (32), 16066–16081.

(157) Zhan, C.-G.; Dixon, D. A. *J. Phys. Chem. A* **2001**, *105* (51), 11534–11540.

(158) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, *110* (7), 2493–2499.

(159) da Silva, E. F.; Svendsen, H. F.; Merz, K. M. *J. Phys. Chem. A* **2009**, *113* (22), 6404–6409.

(160) Xue, X. S.; Wang, Y.; Yang, C.; Ji, P.; Cheng, J. P. *J. Org. Chem.* **2015**, *80* (18), 8997–

9006.

(161) Asthagiri, D.; Pratt, L. R.; Ashbaugh, H. S. *J. Chem. Phys.* **2003**, *119* (5), 2702–2708.

(162) Pliego, J. R.; Riveros, J. M. *J. Phys. Chem. A* **2002**, *106* (32), 7434–7439.

(163) Asthagiri, D.; Pratt, L. R.; Paulaitis, M. E.; Rempe, S. B. *J. Am. Chem. Soc.* **2004**, *126* (4), 1285–1289.

(164) Haworth, N. L.; Wang, Q.; Coote, M. L. *J. Phys. Chem. A* **2017**, *121* (27), 5217–5225.

(165) Deng, H.; Li, X.; Chu, Y.; He, J.; Cheng, J.-P. *J. Org. Chem.* **2012**, *77* (17), 7291–7298.

(166) Bini, R.; Bortolini, O.; Chiappe, C.; Pieraccini, D.; Siciliano, T. *J. Phys. Chem. B* **2007**, *111* (3), 598–604.

(167) Zhao, W.; Leroy, F.; Heggen, B.; Zahn, S.; Kirchner, B.; Balasubramanian, S.; Müller-Plathe, F. *J. Am. Chem. Soc.* **2009**, *131* (43), 15825–15833.

(168) Xue, X.-S.; Wang, Y.; Yang, C.; Ji, P.; Cheng, J.-P. *J. Org. Chem.* **2015**, *80* (18), 8997–9006.

(169) Dissanayake, D. P.; Senthilnithy, R. *J. Mol. Struct. THEOCHEM* **2009**, *910* (1–3), 93–98.

(170) Liakos, D. G.; Neese, F. *J. Chem. Theory Comput.* **2015**, *11* (9), 4054–4063.

(171) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91* (14), 146401.

(172) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7* (18), 3297–3305.

(173) Kesharwani, M. K.; Brauer, B.; Martin, J. M. L. *J. Phys. Chem. A* **2015**, *119* (9), 1701–1714.

(174) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.;

Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian09*, *Revision E.01*; Gaussian, Inc.: Wallingford, CT, 2016.

(175) Neese, F. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2* (1), 73–78.

(176) Dunning, T. H. *J. Chem. Phys.* **1989**, *90* (2), 1007–1023.

(177) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1994**, *100* (4), 2975–2988.

(178) Prascher, B. P.; Woon, D. E.; Peterson, K. A.; Dunning, T. H.; Wilson, A. K. *Theor. Chem. Acc.* **2011**, *128* (1), 69–82.

(179) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1993**, *98* (2), 1358–1371.

(180) Koput, J.; Peterson, K. A. *J. Phys. Chem. A* **2002**, *106* (41), 9595–9599.

(181) Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2002**, *116* (8), 3175–3183.

(182) Hättig, C. *Phys. Chem. Chem. Phys.* **2005**, *7* (1), 59–66.

(183) Hill, J. G.; Platts, J. A. *J. Chem. Phys.* **2008**, *128* (4), 044104.

(184) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2* (2), 364–382.

(185) Zhang, J.; Dolg, M. *Phys. Chem. Chem. Phys.* **2015**, *17* (37), 24173–24181.

(186) Grimme, S.; Bannwarth, C.; Shushkov, P. *J. Chem. Theory Comput.* **2017**, *13* (5), 1989–2009.

(187) Del Pópolo, M. G.; Kohanoff, J.; Lynden-Bell, R. M. *J. Phys. Chem. B* **2006**, *110* (17), 8798–8803.

(188) Trulove, P. C.; Osteryoung, R. A. *Inorg. Chem.* **1992**, *31* (19), 3980–3985.

(189) Wakai, C.; Oleinikova, A.; Ott, M.; Weingärtner, H. *J. Phys. Chem. B* **2005**, *109* (36), 17028–17030.

(190) Singh, T.; Kumar, A. *J. Phys. Chem. B* **2008**, *112* (41), 12968–12972.

(191) Daguenet, C.; Dyson, P. J.; Krossing, I.; Oleinikova, A.; Slattery, J.; Wakai, C.; Weingärtner, H. *J. Phys. Chem. B* **2006**, *110* (25), 12682–12688.

(192) Tariq, M.; Forte, P. A. S.; Gomes, M. F. C.; Lopes, J. N. C.; Rebelo, L. P. N. *J. Chem. Thermodyn.* **2009**, *41* (6), 790–798.

(193) Miran Beigi, A. A.; Abdouss, M.; Yousefi, M.; Pourmortazavi, S. M.; Vahid, A. *J. Mol. Liq.* **2013**, *177*, 361–368.

(194) Huddleston, J. G.; Visser, A. E.; Reichert, W. M.; Willauer, H. D.; Broker, G. A.; Rogers, R. D. *Green Chem.* **2001**, *3* (4), 156–164.

(195) Crowhurst, L.; Mawdsley, P. R.; Perez-Arlandis, J. M.; Salter, P. A.; Welton, T. *Phys. Chem. Chem. Phys.* **2003**, *5* (13), 2790–2794.

(196) Fukaya, Y.; Sugimoto, A.; Ohno, H. *Biomacromolecules* **2006**, *7* (12), 3295–3297.

(197) Brandt, A.; Hallett, J. P.; Leak, D. J.; Murphy, R. J.; Welton, T. *Green Chem.* **2010**, *12* (4), 672.

(198) Ghatee, M. H.; Zolghadr, A. R. *Fluid Phase Equilib.* **2008**, *263* (2), 168–175.

(199) Pereiro, A. B.; Verdía, P.; Tojo, E.; Rodríguez, A. *J. Chem. Eng. Data* **2007**, *52* (2), 377–380.

(200) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612.

(201) Skarmoutsos, I.; Dellis, D.; Matthews, R. P.; Welton, T.; Hunt, P. A. *J. Phys. Chem. B* **2012**, *116* (16), 4921–4933.

(202) Legault, C. Y. Université de Sherbrooke 2009.

(203) Ramachandran, K. I.; Deepa, G.; Namboori, K. *Computational Chemistry and Molecular Modeling*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2008.

(204) Millán, D.; Rojas, M.; Santos, J. G.; Morales, J.; Isaacs, M.; Diaz, C.; Pavez, P. *J. Phys. Chem. B* **2014**, *118* (16), 4412–4418.

(205) *IUPAC Compendium of Chemical Terminology*, 2nd ed.; Nič, M., Jirát, J., Košata, B.,

Jenkins, A., McNaught, A., Eds.; IUPAC: Research Triagle Park, NC, 2009.

(206) Takano, Y.; Houk, K. N. *J. Chem. Theory Comput.* **2005**, *1* (1), 70–77.

(207) Vitz, J.; Erdmenger, T.; Haensch, C.; Schubert, U. S. *Green Chem.* **2009**, *11* (3), 417.

(208) Saracino, G. A. A.; Improta, R.; Barone, V. *Chem. Phys. Lett.* **2003**, *373* (3–4), 411–415.

(209) Dong, H.; Du, H.; Qian, X. *J. Phys. Chem. A* **2008**, *112* (49), 12687–12694.

(210) Fu, Y.; Liu, L.; Li, R.-Q.; Liu, R.; Guo, Q.-X. *J. Am. Chem. Soc.* **2004**, *126* (3), 814–822.

(211) Finkelstein, M.; Mcmillan, J. D.; Davison, B. H.; Evans, B.; Xiang, Q.; Lee, Y. Y.; Torget, R. W. *Issues)* **2004**, *113116*, 273–2289.

(212) Li, C.; Zhao, Z. K. *Adv. Synth. Catal.* **2007**, *349* (11–12), 1847–1850.

(213) Kassaye, S.; Pant, K. K.; Jain, S. *Fuel Process. Technol.* **2016**, *148*, 289–294.

(214) Matsagar, B. M.; Dhepe, P. L. *New J. Chem.* **2017**, *41* (14), 6137–6144.

(215) Himmel, M. E.; Ding, S.-Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. *Science (80).* **2007**, *315* (5813), 804–807.

(216) Matthews, R. P.; Ashworth, C.; Welton, T.; Hunt, P. A. *J. Phys. Condens. Matter* **2014**, *26* (28), 284112.

(217) Loerbroks, C.; Boulanger, E.; Thiel, W. *Chem. - A Eur. J.* **2015**, *21* (14), 5477–5487.

(218) Weerachanchai, P.; Kwak, S. K.; Lee, J.-M. M. *Bioresour. Technol.* **2014**, *170*, 160–166.

(219) Mayes, H. B.; Broadbelt, L. J. *J. Phys. Chem. A* **2012**, *116* (26), 7098–7106.

(220) Dee, S. J.; Bell, A. T. *ChemSusChem* **2011**, *4* (8), 1166–1173.

(221) Zhang, C.; Fu, Z.; Liu, Y. C.; Dai, B.; Zou, Y.; Gong, X.; Wang, Y.; Deng, X.; Wu, H.; Xu, Q.; Steven, K. R.; Yin, D. *Green Chem.* **2012**, *14* (7), 1928.

(222) *Hydrolysis of Cellulose: Mechanisms of Enzymatic and Acid Catalysis*; Brown, R. D., Jurasek, L., Eds.; Advances in Chemistry; American Chemical Society: Washington, D. C., 1979; Vol. 181.

(223) Brás, N. F.; Moura-Tamames, S. A.; Fernandes, P. A.; Ramos, M. J. *J. Comput. Chem.* **2008**, *29* (15), 2565–2574.

(224) Rinaldi, R.; Meine, N.; vom Stein, J.; Palkovits, R.; Schüth, F. *ChemSusChem* **2010**, *3* (2), 266–276.

(225) Liang, X.; Montoya, A.; Haynes, B. S. *J. Phys. Chem. B* **2011**, *115* (36), 10682–10691.

(226) Wilkes, J. S. *J. Mol. Catal. A Chem.* **2004**, *214* (1), 11–17.

(227) Yuan, X.; Cheng, G. *Phys. Chem. Chem. Phys.* **2015**, *17* (47), 31592–31607.

(228) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108* (14), 2715–2719.

(229) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Phys. Chem. A* **2007**, *111* (42), 10439–10452.

(230) Brás, N. F.; Fernandes, P. A.; Ramos, M. J. *J. Chem. Theory Comput.* **2010**, *6* (2), 421–433.

(231) Amarasekara, A. S.; Wiredu, B. *Ind. Eng. Chem. Res.* **2011**, *50* (21), 12276–12280.

(232) Koch, W.; Holthausen, M. C. The Schrödinger Equation. In *A Chemist's Guide to Density Functional Theory*; Wiley: Weinheim, FRG, 2001; p 4.

APPENDIX A

Thermodynamic cycles are a visual representation of the relationship between two states/reactions. In this study they are used to illustrate the relationship between two phases (gas and solvent) however they are not the means by which the relationship was established/derived.

The following algebraic derivations produce the relationships and equations that underlay the application of free energies and thus thermodynamic cycles to calculate pK_as. These are taken directly from derivations shown in literature and thermodynamic definitions of states and properties.

The chemical potential of a species X in solution ($\mu_{\mathrm{soln}}(\mathrm{X})$) can be written as

$$\mu^*_{\mathrm{soln}}(\mathrm{X}) = \mu^*_{\mathrm{gas}}(\mathrm{X}) + \Delta G^*_{\mathrm{solv}}(\mathrm{X}) + RT\ln[X] \qquad \textbf{(A.1)}$$

The superscript * indicates a standard state of 1 mol L^{-1}. $\mu^*_{\mathrm{gas}}(\mathrm{X})$, is the gas phase chemical potential of X, assuming ideal gas nature with a concentration of 1 mol L^{-1}. $\Delta G^*_{\mathrm{solv}}(\mathrm{X})$ is the free energy of solvation for X, defined as the free energy of transfer of 1 mole of solute X from a fixed position in an ideal gas state to a fixed position in an infinitely dilute solution.[140] The final term ($RT\ln[\mathrm{X}]$) is linked to the concentration of X in solution.

Thus, for an acid HA, with solution acidity defined by the following dissociation reaction:

$$\mathrm{HA\ (soln, 1M)} \rightarrow \mathrm{A^-\ (soln, 1M)} + \mathrm{H^+\ (soln, 1M)} \qquad \textbf{(A.2)}$$

the chemical equilibrium can be written, in terms of solution phase chemical potentials, as

$$\mu^*_{\mathrm{soln}}(\mathrm{A^-}) + \mu^*_{\mathrm{soln}}(\mathrm{H^+}) - \mu^*_{\mathrm{soln}}(\mathrm{AH}) = 0 \qquad \textbf{(A.3)}$$

Expanding each term according to Equation A.1 allows for the grouping of terms such that a relationship between free energy terms and concentration terms can be established, as shown below:

$$\Delta G^*_{\mathrm{gas}} + \Delta\Delta G^*_{\mathrm{solv}} = -RT\ln\frac{[\mathrm{A^-}][\mathrm{H^+}]}{[\mathrm{HA}]} \qquad \textbf{(A.4)}$$

$$\text{with } \Delta G^*_{\mathrm{gas}} = \mu^*_{\mathrm{gas}}(\mathrm{A^-}) + \mu^*_{\mathrm{gas}}(\mathrm{H^+}) - \mu^*_{\mathrm{gas}}(\mathrm{HA}) \qquad \textbf{(A.5)}$$

$$\Delta\Delta G^*_{\mathrm{solv}} = \Delta G^*_{\mathrm{solv}}(\mathrm{A}^-) + \Delta G^*_{\mathrm{solv}}(\mathrm{H}^+) - \Delta G^*_{\mathrm{solv}}(\mathrm{HA}) \quad \textbf{(A.6)}$$

Given that the equilibrium constant (K_a) for reaction A.2 is

$$K_\mathrm{a} = \frac{[\mathrm{A}^-][\mathrm{H}^+]}{[\mathrm{HA}]} \quad \textbf{(A.7)}$$

Equation A.4 can therefore be written as

$$e^{-\Delta G^*_{\mathrm{soln}}/RT} = K_\mathrm{a} \quad \textbf{(A.8)}$$

$$\text{where } \Delta G^*_{\mathrm{soln}} = \Delta G^*_{\mathrm{gas}} + \Delta\Delta G^*_{\mathrm{solv}} \quad \textbf{(A.9)}$$

Which taken to - $\log_{10}$ on both sides gives

$$\frac{\Delta G^*_{\mathrm{soln}}}{RT\ln(10)} = \mathrm{p}K_\mathrm{a} \quad \textbf{(A.10)}$$

$$\text{with } \mathrm{p}K_\mathrm{a} = -\log_{10} K_\mathrm{a} \quad \textbf{(A.11)}$$

APPENDIX B

Correct application of a thermodynamic cycles demands all components to be in the same standard state.[136] Commonly a standard state of 1 atmosphere (°) is used for gas phase free energies, as they are defined according to ideal gas conditions. This differs from the standard state used for solvation free energies of 1 M (*). Thus, a standard state correction term of

$$\Delta G^{\circ\rightarrow *} = \Delta \mathrm{n} RT \ln\left(\tilde{R}T\right) \qquad \textbf{(B.1)}$$

is applied to the gas phase leg free energy values to maintain standard state consistency. What follows is the derivation of this correction term, for $\Delta \mathrm{n} = 1$, according to literature and fundamental thermodynamic theory.

The Gibbs free energy change for 1 mole of ideal gas to a pressure of 1 atmosphere ($\mathrm{P_{atm}}$) and a pressure corresponding to 1 mol $\mathrm{L^{-1}}$ (1 M) concentration ($\mathrm{P_{1M}}$) are defined as follows:

$$\Delta G\ (\mathrm{p}, 1\ \mathrm{atm}) = \Delta G^{1\ \mathrm{atm}} = \mu^{\circ} + RT \ln \mathrm{P_{atm}} \qquad \textbf{(B.2)}$$

$$\Delta G\ (\mathrm{p}, 1\ \mathrm{M}) = \Delta G^{1\mathrm{M}} = \mu^{\circ} + RT \ln \mathrm{P_{1M}} \qquad \textbf{(B.3)}$$

where $\mu^{\circ}$is the standard molar chemical potential of the gas, $T$is temperature in Kelvin (K) and R the universal gas constant in units of J $\mathrm{K^{-1}}$ $\mathrm{mol^{-1}}$ (8.314472 (15) J $\mathrm{K^{-1}}$ $\mathrm{mol^{-1}}$).[205]

Using the ideal gas law, the pressure corresponding to a 1 M concentration ($\mathrm{P_{1M}}$) can be written as:

$$\mathrm{P_{1M}} = \frac{\mathrm{n}\tilde{R}T}{\mathrm{V}} = \frac{\mathrm{n}}{\mathrm{V}}\tilde{R}T = \frac{1\ \mathrm{mol}}{\mathrm{L}}\tilde{R}T \qquad \textbf{(B.4)}$$

$$\text{given } c = \frac{\mathrm{n}}{\mathrm{V}} = 1\frac{\mathrm{mol}}{\mathrm{L}} \qquad \textbf{(B.5)}$$

where n is the number of moles, T is temperature in Kelvin (K), V is volume in litres (L) and $\tilde{R}$ the universal gas constant in units of L atm $\mathrm{K^{-1}}$ $\mathrm{mol^{-1}}$ (0.082057 L atm $\mathrm{K^{-1}}$ $\mathrm{mol^{-1}}$). Inserting Equation B.4 into Equation B.3, the free energy difference between the two different pressure states can then be derived as follows:

$$\Delta G^{1\mathrm{M}} - \Delta G^{1\mathrm{atm}} = RT \ln \frac{1\ \mathrm{mol}}{\mathrm{L}}\tilde{R}T - RT \ln \mathrm{P_{atm}}$$

$$\Delta G^{\circ\to *} = RT \ln \frac{1 \text{ mol}/_{\text{L}}\, \tilde{R}T}{1 \text{ atm}}$$

$$\Delta G^{\circ\to *} = RT \ln \frac{1 \text{ mol}/_{\text{L}} \times 0.082057 \text{ L atm}/_{\text{K mol}} \times T}{1 \text{ atm}}$$

$$\Delta G^{\circ\to *} = RT \ln 0.082057/_{\text{K}} \times T \quad \textbf{(B.6)}$$

$$\Delta G^{\circ\to *} = RT \ln \tilde{R}T \quad \textbf{(B.7)}$$

$$\text{with } \Delta G^{\circ\to *} = \Delta G^{1\text{M}} - \Delta G^{1\text{atm}} \quad \textbf{(B.8)}$$

where R is the universal gas constant in units J K^{-1} mol^{-1}, T is temperature in Kelvin (K), and $\tilde{R}$ is the universal gas constant, of 0.082057 L atm K^{-1} mol^{-1}, in units of K^{-1} through unit cancelation. Thus, resulting in the standard state correction term for an overall reaction molar difference (Δn) of 1.