Computational chemistry studies of wood-derived lignans

Thomas Sandberg

Quantum Chemistry and Molecular Spectroscopy
Centre of Excellence for Functional Materials
Laboratory of Physical Chemistry
Department of Natural Sciences
Åbo Akademi University
Åbo 2013

Supervisor

Professor Matti Hotokka
Åbo Akademi University

Pre-examiners

Professor Dage Sundholm
University of Helsinki

Associate Professor Aurora Clark
Washington State University

Opponent for the public defense

Professor Toomas Tamm
Tallinn University of Technology

ISBN 978-952-12-2873-5
Painosalama – Åbo 2013

"Our greatest glory is not in never falling,
but in rising every time we fall."
– *Konfucius*

"There is nothing permanent except change."
– *Heraclitos*

Förord på svenska
”Other Side Of The Story”

Enligt 27 § i ”Instruktion för forskarutbildningen vid Åbo Akademi (20.10.2011)” bör en artikelavhandling (som denna) ”bilda en helhet som i regel innefattar 1–4 artiklar för licentiat- och 3–7 artiklar för doktorsavhandling.” ”Målsättningen är att disputation kan hållas och examen utfärdas inom fyra år efter antagning till utbildning.” Verkligheten är dock ofta en annan, och detta är en sådan historia.

Redan då jag började mitt pro gradu-arbete med Max Johansson som närmaste handledare i praktiken, ville min övervakande handledare prof. Matti Hotokka att jag skulle bli ”Mr MD” på Åbo Akademi. Det var en fin vision, men vägen dit visade sig vara stenig. Det var en lycka att jag sommaren 2004 beslöt att utvidga min första ”övningsuppgift” i kvantkemiska beräkningar på några toluenderivat till en pro gradu-avhandling, för annars hade jag suttit kvar som kandidat ända till år 2009.

Hösten 2005 blev jag magister och det var dags att ”börja på fyke” på riktigt. Som vanligt är det två skilda världar vad man önskar jobba med och vad man hittar finansiering för. Kanske delvis p.g.a. detta dilemma blev min första forskningsuppgift ett mer eller mindre ”omöjligt projekt” som tyvärr lär vara alltför vanliga för nya doktorander som ofta själva saknar helhetsuppfattning. Att inte heller ha funnit en lämplig rytm mellan forskning, egna studier och avslutad idrottskarriär, gjorde att jag under våren 2006 drabbades av yrselattacker förorsakade av en skadad hörselnerv, med stress som en mycket möjlig utlösande faktor.

I det här skedet hade redan finansieringsformen hunnit ändra flera gånger om, alltifrån korta stipendiatperioder till kursassistentarvoden för laborationshandledning. Finansieringscirkusen är för övrigt något som fortsatt genom hela denna resa. Från hösten 2006 till 2010 fick jag åtnjuta en doktorandplats i forskarskolan LasKeMo – dock såtillvida rationaliserad att jag erhöll ca nio månader lön per år, men med en ständig ovisshet om hur nästa år skulle se ut. Min livlina

under dessa år blev laborationerna i fysikalisk kemi som förutom en mental paus i forskningen gav mig ca två månader lön per år. Den sista finansieringsmånaden gick under dessa år oftast att pussla ihop med FPA:s ersättingar, eftersom mina viktigaste delar i livet, Melvin och Lova, föddes 2007 resp. 2009 ♡.

Efter att undervisnings- och kulturministeriet inte beviljade fortsatt finansiering för vår forskarskola gällde det att börja söka pengar på annat håll. Jag har haft lyckan att få napp på flera håll, och däremellan räddats av spetsforskningsanslag på mitt eget laboratorium som med öppet hjärta tagit emot mig trots att jag är teoretiker och egentligen inte ingår i laboratoriets forskningsmiljö. Man kan tycka att det är en självklarhet att varje doktorand kan skriva sin egen forskningsplan och sälja sig till finansiärerna, men då bör man minnas att detta är ett synnerligen tidskrävande arbete. Och då meningen är att leverera en färdig bok efter blott fyra år, finns det inte mycket tomt utrymme. Även om det tidvis varit jobbigt, kan jag samtidigt skatta mig lycklig att jag inte fått fyra års finansiering på silverfat, utan i stället tvingats vänja mig vid vardagen efter disputation. Men det har tagit sin tid. Jag har stigit på varje trappsteg, frivilligt tagit ut varje examen på vägen, och det är jag stolt över.

För att återknyta till historien om arbetet med just denna avhandling var det dags för en två veckors sjukledighet p.g.a. ryggbesvär under hösten 2007 med efterföljande rehabilitering. Tacksamt nog har både vårt laboratorium börjat satsa på personalens ergonomi och även Åbo Akademi insett nyttan med förebyggande vård i stället för att tvingas sjukskriva personalen under långa perioder. Detta kunde man tycka är en självklarhet, men det har krävts många offer på vägen.

År 2008 var dock ett slags vändpunkt för mig i bedrövelsen. En bidragande bakomliggande orsak till detta kan ha varit att jag deltog i den ettåriga kursen i Universitetspedagogik, där undervisningsmetoden i huvudsak var socialkonstruktivistisk, vilket gjorde att jag träffade många doktorander från andra ämnesområden, men i övrigt i motsvarande situation som jag med småbarn (nyfödda eller på kommande), osäkerhet i finansiering och ibland också i forsk-

ningen kombinerat med en ångest över livet efter avhandlingen. Att därtill få dela den underbart vackra majmånaden med min ettåriga Melvin, och att efter tre års ansträngning slutligen lyckas genomföra (underförstått publicera) molekyldynamiksimuleringsarbetet (artikel I), var en belöning som fick mig att orka kämpa vidare.

”För säkerhets skull” skrev jag även ihop en licentiatavhandling under sommaren 2009. Fortsatt finansiering har som sagt varit det största ständiga hotet under denna resa. För tillfället är jag uppe i 24 kontrakt/stipendieperioder sedan starten för åtta år sedan. Många handledare tycker att licentiatexamen är slöseri med tid. Jag kan hålla med om det förutsatt att (som det tyvärr blivit på många ämnen) doktorsexamens omfattning motsvarar min licentiatexamen. Den blev klar på fyra år trots motgångar, ett ”bortkastat” första doktorandår och föräldraledigheter. På Fysikalisk kemi har dock nivån alltid varit betydligt högre än så, och det är bara beklagligt att allt flera sänker nivån för att klara målsättningarna.

I samband med LasKeMo:s begravning valde jag att stanna hemma med barnen under hösten 2010, och det var helt klart den finaste tid jag haft hittills. Samtidigt både publicerade jag (artikel IV) och deltog i Vinterskolan i teoretisk kemi i Helsingfors tillsammans med min ettåriga Lova.

Sommaren 2011 såg det faktiskt ut som att jag skulle kunna disputera ett år senare, men den ursprungliga arrangören för konferensen 17^{th} International Workshop on Quantum Systems in Chemistry and Physics (QSCP-XVII) 2012 drog sig ur, och vi ”blev tvungna” att ta emot utmaningen ett år tidigare än planerat. Med hjälp av våra experimentella kollegor i FyKe-familjen lyckades vi dock genomföra en lyckad tillställning redan i augusti 2012. Bland ytterligare faktorer som bidrog till fördröjningen av mitt arbete kan även nämnas en vattenskada i Gadolinia som förstörde ca tre rum per våning och bl.a. tvingade oss i rum 306 att flytta ut i några veckor. Sådant här ”borde inte få hända” och naturligtvis är ”ingen ansvarig”, men likväl är resultatet detsamma. Vi får bara tacka högre makter att skadorna inte blev ekonomiskt mer betungande.

Senaste höst (2012) har jag dock skalat bort störande moment, tackat nej till extra uppdrag och fokuserat på sammanfattningen som håller ihop denna bok. Inte heller det fick förstås gå på räls, utan i slutet av september drabbades jag av en allvarlig variant av lunginflammation (pneumoni) som höll mig däckad i ett par veckor. Som kronan på verket pajade den knappt ett år gamla hårdskiva som jag arbetade på, men det är förstås mera kutym än undantag. Men nu är det klart.

Åbo, 5 april 2013
Thomas Sandberg

Preface

The work behind this thesis was conducted in Prof. Matti Hotokka's research group in Quantum Chemistry and Molecular Spectroscopy at the Department of Natural Sciences (until 2009 the Faculty of Mathematics and Natural Sciences) during the years 2005–2012. The research I have conducted during this time has been nuanced and somewhat diverging. I have collaborated with several research groups that have been in need of theoretical calculations in their experimental work. Two trends have, though, crystallized. One of them is summarized in this thesis, and the second can be seen in the supporting publications.

A basis for all computational studies is structural optimization of the studied molecules, but I have chosen to focus on classical molecular dynamics in the methods chapter as my main interest of science lies there. All figures in chapter 3.3 are made using the program Xfig from a model of [1] except the figures 3.1 and 3.5 for which the program Gnuplot was used. The type setting is made using LaTeX 2_ε. Part of this PhD thesis is published in Swedish as my Licentiate thesis [2].

Acknowledgements

I express my deepest gratitude for all funding I have received for my research during these years, especially from the Graduate School of Computational Chemistry and Molecular Spectroscopy - LasKeMo († R.I.P.), the Centre of Excellence for Functional Materials - FunMat and Victoriastiftelsen, as well as from Stiftelsens för Åbo Akademi forskningsinstitut, Magnus Ehrnrooths stiftelse, Ellida och Tor Magnus Ljungbergs stipendiefond and Berndt Olof Grönbloms stipendiefond. Computational and information services provided by CSC – IT Center for Science are kindly acknowledged. Especially I would like to thank FD Atte Sillanpää and FD Nino Runeberg at the Chemistry services at CSC.

I sincerly thank the reviewers of this thesis, Prof. Dage Sundholm and

Ass. Prof. Aurora Clark, for their valuable comments to help me improve this work, as well as Lecturer Colette Gattoni for language checking and evaluation and Lecturer Ream C. Barclay for the profitable sessions of individual academic text consultation provided by the Centre for Language & Communication at Åbo Akademi University. Dr. Fredrik Jansson is thanked for all kind of help especially regarding the type setting in LaTeX 2_ε.

The most relevant person during this journey is of course Prof. Matti Hotokka. When I began my university studies at ÅAU, my plan was to choose organic chemistry as my major and graduate as a teacher. At the time of my BSc thesis, I considered changing major to mathematics. However, I ended up with you, Matti. It cannot have been easy to act at all levels one can find out in a scientific relation: teacher, instructor, advisor, supervisor, head of department, dean and even remover during the floor renovation in Gadolinia in the summer of 2007, but I feel we made it good enough on each of them. Our last battle was to survive the arrangements (with a huge help of the staff at fyke) of the 17^{th} International Workshop on Quantum Systems in Chemistry and Physics (QSCP-XVII) in August 2012. And we did it. When I came to your group, you probably expected something fast and dirty, but science seldom works on demand. I took my time to climb every rung of the ladder. So here we are now, and that is what matters.

One advantage with scientific calculations is the possibility to predict things before actually doing them in the lab. Calculations without scientific relevance, however, are quite useless. It is not a coincidence that I ended up collaborating with Docent Patrik Eklund. This was the field, where I felt that my expertice was most useful to the experimentalists, and for some reason it has always been easier for me to discuss science with you, Patrik, than with anyone else. You are honest, esteeming and not judging a young scientist like me. Thank you! I really appreciate this and hope our collaboration will continue for long.

To make my living and rest my nerves, I have been working as a teaching assistant, instructing students in lab work in physical chemistry every year from

February–April, between 2006–2010 together with Dr. Gun Hedström and the last two years all by myself. I would like to thank you, Gun, for being the best colleague one can get, always trying to do the best out of the reality. Not many have the interest and strength like you to struggle against the conservative way of thinking about teaching in our field.

I also like to thank Prof. Emer. Jarl Björn "Nalle" Rosenholm and his successor, Prof. Jouko Peltonen for accepting me in the FyKe-family despite the fact that I am a theoretician. This has meant much to me, both financially and mentally, especially since LasKeMo was ended.

Since the change in the organization structure at Åbo Akademi University from the beginning of 2010, it has been a turbulent time at the university in general and at the Department of Natural Sciences in particular. My best regards to Docent Jan-Henrik Smått and Dr. Tom Lundin. It has been great three years in the Department Council, and I have to say that I am proud of what our team accomplished. Thank you also, as well as Dr. Santeri Puranen, for all your support regarding research plans etc. on my way to find my own path as a researcher. Without the last two employees at fyke, Christina Luojola and Kenneth Stenlund the lab would be a mess. Thank you for keeping the lab in order and helping out with whatever it might be.

Countless are the people that I have got hints from on my way to finish the thesis. My first advisor, Max Johansson, is greatly acknowledged for getting me into "the right path", writing shell scripts, constructing z-matrices and using Gamess, beginning on 30.10.03. However, my last resort, when I have got stuck with the research has many times been my former colleague, Dr. Antti Taskinen. Thanks for always trying to find solutions to my problems, even if you must have thought "RTFM". The last but not least person solving my computer problems, who is like a father of my linux computer hermes, is Dennis Holtlund. Thanks for always being willing to help me, most often at very short notice, when I am already deep in trouble.

Besides, my warmest regards goes to all my co-authors for all the struggling

and success, my room mates in Gado 306, Gunnar and Helka for never letting me sink down too deeply into my own thoughts, and finally Dennis and Pia for the loyal company at the coffee breaks.

Athletics, especially running, has been essential for three quarters of my life, at times managing the whole of it. Since the autumn of 2004 I have had the opportunity to practice together with Axel Meierjohann, which has most often made me feel calm and easier in my mind, even though I have been very frustrated when coming to the training. Especially our morning workouts and sauna in Kuppis really took me through the struggling with manuscripts and finally the writing process of this thesis, so thanks!

There is also a group of my closest friends since my pre-university life: Torulf, Tove, Martin and Niklas, that deserve special regards for always being there for me, no matter how little time we have to meet.

Finally I would like to give my warmest hugs to my parents Bengt and Benita, who have always encouraged me on my way to fulfil myself both in athletics and in my education, lately also helping out with child-care, and my wife Lotta for sharing your life with me for better or for worse. I would never have succeeded without you. I also like to thank my father-in-law Matts for always welcoming us to your peaceful cottage in the Närpes forrest. It is the only place, where I can totally forget all thoughts about work and just rest. And last but not least, the meaning of my life, Melvin and Lova ♡. Thanks for all your joy and energy, especially when I feel bad, and for keeping up the good mood even when "dad has to work".

Åbo, April 5, 2013
Thomas Sandberg

Abstract

This thesis is based on computational chemistry studies on lignans, focusing on the naturally occurring lignan hydroxymatairesinol (HMR) (Papers I–II) and on TADDOL-like conidendrin-based chiral 1,4-diol ligands (LIGNOLs) (Papers III–V).

A complete quantum chemical conformational analysis on HMR was previously conducted by Dr. Antti Taskinen. In the works reported in this thesis, HMR was further studied by classical molecular dynamics (MD) simulations in aqueous solution including torsional angle analysis, quantum chemical solvation effect study by the COnductorlike Screening MOdel (COSMO), and hydrogen bond analysis (Paper I), as well as from a catalytic point of view including protonation and deprotonation studies at different levels of theory (Paper II).

The computational LIGNOL studies in this thesis constitute a multi-level deterministic structural optimization of the following molecules: 1,1-diphenyl (**2Ph**), two diastereomers of 1,1,4-triphenyl (**3PhR**, **3PhS**), 1,1,4,4-tetraphenyl (**4Ph**) and 1,1,4,4-tetramethyl (**4Met**) 1,4-diol (Paper IV) and a conformational solvation study applying MD and COSMO (Paper V). Furthermore, a computational study on hemiketals in connection with problems in the experimental work by Docent Patrik Eklund's group synthesizing the LIGNOLs based on natural products starting from HMR, is shortly described (Paper III).

Sammanfattning på svenska

I denna avhandling behandlas beräkningskemiska studier på lignaner, med fokus på den naturligt förekommande lignanen hydroximatairesinol (HMR) (artiklarna I–II), och på TADDOL-liknande kirala 1,4-diolligander (LIGNOL:er) baserade på conidendrin (artiklarna III–V).

En fullständig kvantkemisk konformationsanalys av HMR gjordes tidigare av FD Antti Taskinen. I de arbeten som presenteras i denna avhandling fortsatte jag forskningen på HMR med klassiska molekyldynamiksimuleringar (MD) i vattenlösning inklusive torsionsvinkelanalys, en kvantkemisk studie av lösningsmedelseffekter med lösningsmedelsmodellen COnductorlike Screening MOdel (COSMO) och en vätebindningsanalys (artikel I). Därtill utförde jag protonerings- och deprotoneringsstudier på olika teorinivåer ur en katalytisk synvinkel (artikel II).

De beräkningskemiska studierna av LIGNOL:er i denna avhandling består av en deterministisk strukturoptimering på olika teorinivåer av följande molekyler: 1,1-difenyl (**2Ph**), två diastereomerer av 1,1,4-trifenyl (**3PhR**, **3PhS**), 1,1,4,4-tetrafenyl (**4Ph**) och 1,1,4,4-tetrametyl (**4Met**) 1,4-diol (artikel IV) och en konformationsstudie i lösningsmedel med MD och COSMO (artikel V). Ytterligare beskrivs kort en beräkningsstudie av hemiketaler i anslutning till problem vid det experimentella arbetet av Docent Patrik Eklunds grupp med att syntetisera LIGNOL:erna som baserar sig på naturliga produkter med HMR som utgångsmaterial (artikel III).

List of original publications

This thesis is based on the following original publications, which are referred to in the text by Roman numerals.

I Sandberg, T. and Hotokka, M. "Conformational analysis of hydroxymatairesinol in aqueous solution with molecular dynamics" *Journal of Computational Chemistry* **30** (2009) 2666–2673.

II Markus, H., Plomp, A. J., Sandberg, T., Nieminen, V., Bitter, J. H. and Murzin, D. Yu. "Dehydrogenation of hydroxymatairesinol to oxomatairesinol over carbon nanofibre-supported palladium catalysts" *Journal of Molecular Catalysis A* **274** (2007) 42–49.

III Brusentsev, Y., Sandberg, T., Hotokka, M. Sjöholm, R. and Eklund, P. "Synthesis and structural analysis of sterically hindered chiral 1,4-diol ligands from the lignan hydroxymatairesinol" *Tetrahedron Letters*, Accepted 2012.

IV Sandberg, T., Brusentsev, Y., Eklund, P. and Hotokka, M. "Structural analysis of sterically hindered 1,4-diols from the naturally occurring lignan hydroxymatairesinol. A quantum chemical study" *International Journal of Quantum Chemistry*, **111** (2011) 4309–4317.

V Sandberg, T., Eklund, P. and Hotokka, M. "Conformational solvation studies of LIGNOLs with molecular dynamics and Conductor-like screening model" *International Journal of Molecular Sciences*, **13** (2012) 9845–9863.

The author of the thesis made all the theoretical modelling in all the papers mentioned above, wrote the manuscripts I, IV and V , as well as the theoretical part of papers II and III.

List of supporting publications

S1 Granqvist, B., Sandberg, T. and Hotokka, M. "Adsorption of organic probes on silica through Lewis interactions: A comparison of experimental results and quantum chemical calculations" *Journal of Colloid and Interface Science* 310 (2007) 369–376.

S2 Sandberg, T., Rosenholm, J. and Hotokka, M. "The molecular structure of disulfiram and its complexation with silica. A quantum chemical study" *Journal of Molecular Structure (THEOCHEM)* 861 (2008) 57–61.

List of conference presentations

C1 Sandberg, T. and Hotokka, M. "Conformational analysis of hydroxymatairesinol in aqueous solution with molecular dynamics" *17th International Workshop on Quantum Systems in Chemistry and Physics*, Turku, 19–25.8.2012, *13th ICQC International Congress of Quantum Chemistry*, Helsinki, 22-27.6.2009, *FunMat Seminar*, Helsinki, 3.3.2009, *Nanoscience Days*, Jyväskylä, 29–30.10.2009. Poster presentations.

C2 Sandberg, T., Brusentsev, Y., Eklund, P. and Hotokka, M. "Structural analysis of sterically hindered 1,4-diols from the naturally occurring lignan HMR" *17th International Workshop on Quantum Systems in Chemistry and Physics*, Turku, 19–25.8.2012, *14th ICQC International Congress of Quantum Chemistry*, Boulder, Colorado, USA, 25–30.6.2012, *FunMat Annual Symposium*, Turku, 24–25.8.2011, *27th Winter School in Theoretical Chemistry*, Helsinki, 12–15.12.2011. Poster presentations.

C3 Sandberg, T., Rosenholm, J. and Hotokka, M. "Interactions between Antabus® and silica surfaces" *17th International Workshop on Quantum Systems in Chemistry and Physics*, Turku, 19–25.8.2012. Poster presentation.

C4 Sandberg, T., Hedström, G., Granqvist, B., Rosenholm, J. B. and Hotokka, M. "Interaction of benzyl probes with silica in cyclohexane. Ab initio calculations, adsorption and microcalorimetry" *17th International Workshop on Quantum Systems in Chemistry and Physics*, Turku, 19–25.8.2012. Poster presentation.

C5 Sandberg, T., Markus H., Nieminen V. and Murzin, D. Yu. "Dehydrogenation of Hydroxymatairesinol to Oxomatairesinol" *17th International Workshop on Quantum Systems in Chemistry and Physics*, Turku, 19–25.8.2012. Poster presentation.

C6 Sandberg, T., Eklund, P. and Hotokka, M. "Conformational Solvation Studies of LIGNOLs with Molecular Dynamics and COSMO" *17th International Workshop on Quantum Systems in Chemistry and Physics*, Turku, 19–25.8.2012, *14th ICQC International Congress of Quantum Chemistry*, Boulder, Colorado, USA, 25–30.6.2012, *FunMat Annual Symposium*, Turku, 23.5.2012. Poster presentations.

C7 Sandberg, T., Eklund, P. and Hotokka, M. "Conformational Chemistry Studies of LIGNOLs" *PARA 2012: Workshop on State-of-the-Art in Scientific and Parallel Computing*, Helsinki, 10–13.6.2012, In *Applied Parallel and Scientic Computing*, Manninen, P.; Öster, P., Ed., volume 7782 of *Lecture Notes in Computer Science*, pp. 553–557. Springer Berlin Heidelberg, 2013. Oral presentation.

List of abbreviations

6-31G*	a basis set consisting of six primitive Gauss functions for the inner atomic orbitals and two contracted functions with three and one Gauss functions, respectively, in the valence area, as well as polarization functions for all atoms except hydrogen and helium
B3LYP	Becke's non-local three parameter exchange functional combined with Lee-Yang-Parr's non-local correlation functional in DFT
BP86	the combination of Becke's non-local exchange functional and Perdew's non-local correlation functional in DFT
COSMO	COnductorlike Screening MOdel - a solvation effect method
DFT	Density Functional Theory
HF	Hartree-Fock method
HMR	hydroxymatairesinol
LJ	Lennard-Jones (interactions)
MC	(Metropolis) Monte Carlo
MD	molecular dynamics
MM	(classical) molecular mechanics
μVT	grandcanonical ensemble (constant temp., volume and chemical potential)
NPT	isothermic-isobaric ensemble (constant temp., pressure and number of particles)
NVE	microcanonical ensemble (constant energy, volume and number of particles)
NVT	canonical ensemble (constant temp., volume and number of particles)
PCM	Polarizable Continuum Model - a solvation model
PME	particle-mesh Ewald - an algorithm for long-range forces
SCRF	Self-Consistent Reaction Field method
TZVP	a triple-ζ basis set emphasizing the valence area and extended with polarization functions

List of figures

Contents

Chapter 1

Introduction

1.1 Lignans → LIGNOLs

One of the easiest ways to receive optically pure, chiral compounds is by isolating natural products and synthesizing derivatives of these. Lignans are a group of phenols which occur in many different plant parts, consisting of two β–β linked phenyl propane units [3], as shown in Figure 1.1.

The richest lignan source in nature is Norway spruce (*Picea abies*), the knots of which contain up to 25 % lignans, out of which approximately 80 % is hydrox-

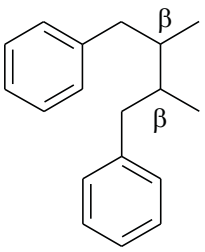


Figure 1.1: Schematic picture of lignans.

ymatairesinol (HMR) [4]. HMR can be used as a starting material for synthesis of many other lignans, as described in Papers II, III and IV. Several of these lignans have been confirmed to be strong antioxidants and to have anticarcinogenic effects (Paper II).

In Papers III and IV, the aim was to synthesize TADDOL-like conidendrin-based chiral ligands. Conidendrin is an intermediate in the synthesis of chiral catalysts (in this case 1,4-diols) from HMR. TADDOLs, *i.e.* α,α,α',α'-tetraaryl-1,3-dioxolan-4,5-dimethanols, have been used extensively as ligands in asymmetric catalysis, but the group of Docent Eklund has synthesized a new class of 1,4-diols based on natural products with the same catalytic functionality.

1.2 Computer simulations

Usually it takes several decades before inventions receive their applications, but where computer simulations are concerned, history is very different. During and after the Second World War, electronical calculation machines were built to compute the very heavy calculations needed to develop nuclear weapons and to break secret codes. Computer simulations started as a way to exploit these calculation machines. In the early 1950s, electronic computers became partly available even for those outside the military, which marked the beginning of the scientific field of computer simulations.

The first computer simulation of a liquid was performed at Los Alamos National Laboratories in USA with the so-called MANIAC [5]. The accurate technique based on random numbers that was used in this very first study is still in general use, and usually referred to as Metropolis Monte Carlo or MC.

To receive dynamic properties of many-particle systems a different technique is needed. Molecular dynamics (MD) is the term used to describe solving Newton's classical equations of motion for a number of molecules. This was first done by Alder and Wainwright [6, 7] on a system consisting of hard spheres. In this case, the particles move at constant speeds between totally elastic collisions,

and it is possible to solve the dynamic problem without any approximations. This happened many years before a successful attempt was made by Rahman [8] to solve the equations of motion for a number of Lennard-Jones particles. In this case, an approximative step-by-step method had to be applied, as the forces changes continuously as the particles move.

In Papers I and V, MD simulations have been performed on the naturally-occuring lignan HMR and on TADDOL-like conidendrin-based chiral ligands as an atomic system, where the internal degrees of freedom in the well-defined structures of the molecules has been allowed to vary freely.

1.3 Aims

The overall aim of the projects behind this thesis was to utilize the natural chirality of the available natural lignans, *e.g.* for chiral catalysts, originating from our vast, renewable source: the Finnish forest.

Each project has its own aims, and concerning HMR (Papers I and II) the aim was focused on the differences between the two diastereomeres (7R,8R,8R)-(–)-7-allo-hydroxymatairesinol (RRR) and (7S,8R,8R)-(–)-7-hydroxymatairesinol (RRS), which acted very differently. In the dehydrogenation reaction to 7-oxomatairesinol (oxoMAT) RRS reacted with a higher reaction rate and the yield of oxoMAT was higher for RRS. Eklund *et al.* [9] concluded in the oxidation of HMR with DDQ, that the RRS isomer gave oxoMAT, but only small amounts were obtained when the RRR isomer was used. Earlier it was concluded that RRS reacted with a higher rate also in the hydrogenolysis reaction to matairesinol (MAT) [10, 11].

In the work behind Papers III–V, the aim was to synthesize a new class of 1,4-diols, *i.e.* TADDOL-like conidendrin-based chiral ligands (LIGNOLs). One objective of the computational study was to observe how the angle between the hydroxyl groups in these LIGNOLs behaves in the optimized structures. If the OH groups furthermore point to the same direction, an intramolecular hydrogen

bond forms between them, and the bridging hydrogen atom falls at the same place a chelate-bonded metal ion would be situated.

One factor that often affects the molecular global minimum energy configuration is the solvent. Traditional quantum chemical calculations do not account for the solvent, *i.e.* these are made assuming that the molecules are isolated. With continuum models the solvent can be taken into account as a bulk medium. However, in molecular dynamics simulations, the solvent molecules are explicitly present, thus defining a more accurate description. In Papers I and V, the aim was to explore how the solvent molecules affect the conformation of HMR and the LIGNOLs, respectively.

Chapter 2

Materials

The history of hydroxymatairesinol (HMR) is a series of coincidences and forgetfulness. Karl Freudenberg and Leo Knof were the first to report HMR in Norway spruce in 1957. [12] It was not until 20 years later, in 1976, [13,14] that Dr. Rainer Ekman at the Laboratory of Wood and Paper Chemistry at Åbo Akademi University reported HMR to be the most abundant lignan in Norway spruce. However, a small detail, the knot, was still missing in Ekman's study. HMR became an topic again in 1996 as Hormos Medical Ltd. found interest in natural lignans, because of their possible antioxidative and anticarcinogenic effects.

Prof. Emer. Bjarne Holmbom at the Laboratory of Wood and Paper Chemistry is the man behind many coincident innovations. In 1998, he brought a knot from a spruce tree to the laboratory for analysis. Dr. Lari Vähäsalo analysed it and found huge amounts of lignans in the knot, dominated by HMR. Prof. Stefan Willför continued the research on knots in more detail and showed that the Norway spruce knots could contain up to 25 % lignans, of which approximately 80 % was HMR. [4] Beginning in 2001, Docent Rainer Sjöholm at the Laboratory of Organic Chemistry started a research project on synthetic modification of HMR, which developed into in Docent Patrik Eklund's doctoral

thesis in 2005, [15] which covers most of the chemistry of lignans related to the 2,3-dibenzylbutyrolactone structure. A complete quantum chemical structural analysis on HMR was made by Dr. Antti Taskinen [16] at the same time.

The knots, *i.e.* the parts of the branches embedded in the wood stem, were for a long time an unprofitable problem in the pulping processes as they caused so-called over-sized chips. However, the group headed by Prof. Emer. Bjarne Holmbom at the Laboratory of Wood and Paper Chemistry developed a separation method, named ChipSep [17–19], in which over-sized wood chips are ground to splinters and the knotwood material is separated by sedimentation in water. To produce pure HMR, the knotwood is then extracted with ethanol and finally HMR is precipitated by adding potassium acetate to the solution. Theoretically it would be possible to produce 600 kg of pure HMR per day in a Finnish pulping mill.

HMR is already sold by the Food and Drug Administration in the USA as health nutrition, and as HMR was already available in kg-scale it could be used as a source for the synthesis of other lignans. In 2008, Docent Patrik Eklund started the project "Lignocats" with the aim of synthesizing a new class of 1,4-diols, *i.e.* TADDOL-like conidendrin-based chiral ligands (LIGNOLs), based on natural products with the same catalytic functionality as TADDOLs.

Chapter 3

Methods

Molecular modelling is a method of combining computational chemistry techniques with graphical visualization in order to simulate and predict three-dimensional structures, chemical processes and physiochemical properties of molecules and ensembles. The strength of molecular modelling is that one can predict properties before one has to conduct the experiments in the laboratory.

J.-R. Hill *et al.* [20] and A. R. Leach [21] give an outline of the network of theoretical chemistry which is easy to understand.

3.1 Structural optimizations

A non-linear molecule with N atoms has $3N-6$ different structural parameters or internal degrees of freedom. The optimal structure is at the minimum on a potential energy surface in $3N-6$ dimensions. The optimization algorithms can deal with multi-dimensional surfaces but they usually find the closest minimum, which is not necessarily the lowest one, the global minimum.

Not having to rely merely on one's intuition, a number of starting structures can be chosen in a systematic way, then optimized and finally their energies can be compared. Such is the way I have proceeded in the three Papers; S1, S2 and

IV.

In practice, I have started by drawing the molecule using the program Sybyl [22], and performed a torsional analysis on the two or three most relevant single bonds (degrees of freedom) using the step size of 30–60°. I have used the rough molecular mechanics (MM) method, which is totally based on classical physics, and applied Tripos force field [23]. Among these structures, I have then selected the energetically most favourable structures (with MM energies within 20–30 kJ/mol) and optimized them using the Gamess program version 22 Feb 2006 [24] at the Hartree-Fock (HF) level [25] and the 6-31G* basis set [26–28].

When needed, I have then changed to the program package Turbomole [29–31] with the density functional method (DFT) [32], the exchange-correlation functional BP86 [33, 34] or the hybrid functional B3LYP [33, 35, 36] and the basis set TZVP [37], to include electron correlation without the calculations becoming much heavier than the HF calculations.

3.1.1 Basis set notations

In my master's thesis [38], I have carefully explained the HF method and basis sets in general. As I want to focus on molecular dynamics in this PhD thesis, I refer to a fundamental text book by Cramer [39] for more details on the DFT method. A short explanation about the different ways of denoting basis sets is not excluded.

For basis sets developed by Pople and co-workers [40] the notation a-bc $\cdots$G, where a, b, c etc. are integers and G stands for Gaussian functions [41], is used. The integer a denotes the number of primitive Gaussian functions for the core atomic orbitals needed to obtain good total energy, while b, c etc. are the number of valence orbitals to obtain a balanced chemical behaviour.

Nowadays two or more contracted atomic orbitals with different radii are exclusively used, and their weight is balanced so that the actual bond length can be expressed more properly. These types of basis sets are hence called

split-valence basis sets.

The notations can be further expanded with asterisks (∗) for polarization functions and plus signs (+) for diffuse functions. The polarization functions give more flexibility to the orbital shape and enhance the angular behaviour, and the diffuse functions describe long-range electrons.

Other basis sets are usually classified by the number of basis functions used for each atomic orbital. A basis set with two basis functions for each (inner) orbital is called a double-ζ-basis set (DZ), three basis functions give a triple-ζ-basis set (TZ) etc. If the large number of basis functions is limited to the valence field, the basis set is called valence double-ζ-basis set, which Pople's 6-31G basis set, though, is. P in the basis set TZVP that I used for the DFT calculations, stands for polarization functions.

3.2 Solvent effects

In chapter 3.1, I described how the structures could be optimized. In Papers I and II, pre-optimized structures [16] were used as initial geometries for subsequent structural optimizations at different levels of theory. The optimized structures were then used in studies of molecular properties.

The main method in Papers I and V was molecular dynamics which I discuss in chapter 3.3. The greatest benefit of molecular dynamics for me is that one can see directly how the solvent affects the conformations of the molecules, as the solvent molecules are explicitly present.

In some systems, however, the solvent only works as a bulk medium, which certainly affects the behaviour of the solute, but in which it is unnecessary to contemplate each separate solvent molecule. The solvent works as a perturbation on the gas phase behaviour of the system. This is the idea behind the continuum solvent models (CSM) [42]. There is a great amount of such models to be used together with both quantum chemical and empirical models.

Two important contributions in the study of solvent effects were presented

by Born in 1920 [43] and Onsager in 1936 [44]. Born derived the electrostatic component of Gibb's free energy of solvatization, when a charge (ion) is placed in a spherical solvent cavity, and Onsager extended this to also hold for a dipole in a spherical cavity.

The model by Onsager is a special case of the results derived by Kirkwood in 1934 [45] regarding an arbitrary distribution of charges in a spherical cavity. The solvated dipole in the cavity induces a dipole in the surrounding medium, which further induces an electric field in the cavity (reaction field). The reaction field then interacts with the solvated dipole by that stabilizing the system further.

The reaction field method can be used in quantum chemical calculations by including a reaction-field term in the Hamiltonian and treating this as a perturbation. The method is then called the Self-Consistent Reaction Field method (SCRF).

A drawback of the SCRF method is that a spherical cavity is used. Molecules are seldom spherical. A much more realistic cavity shape is used in the PCM method (Polarizable Continuum Model) [46–48], where a cavity is formed in a polarizable dielectric medium with the electric permittivity ε depending on the van der Waals radius of the molecule.

The PCM method is implemented in a large number of quantum chemical program packages and has been used, *e.g.* by Taskinen *et al.* [16], the results of which I compared my results in Paper I with.

The method I have used in Paper I and V is called COSMO (COnductorlike Screening MOdel) [49,50] and it resembles PCM. It is based on the fact that the solvated molecule forms a cavity in a dielectric continuum with the permittivity ε corresponding to the solvent. The charge distribution of the solute polarizes the dielectric medium, and the response from the medium is described by the formation of screening charges on the cavity surface [51].

Continuum models for solvents usually imply that quite complicated boundary conditions (see chapter 3.3.2) have to be employed for a dielectric (non-conductive) material to get the screening charges. In COSMO, the much easier

boundary conditions with vanishing electrostatic potential for a conductor,

$$\mathbf{\Phi}^{tot} = 0\,, \tag{3.1}$$

are used instead. This corresponds to an electrostatically ideal solvent with $\varepsilon = \infty$. The total electrostatic potential vector on the surface element of the cavity is determined by the potential of the solute, $\mathbf{\Phi}^{sol}$, consisting of the electronic part and the core part, and the screening charge vector $\mathbf{q}$,

$$\mathbf{\Phi}^{tot} = \mathbf{\Phi}^{sol} + \mathbf{A}\mathbf{q} = 0\,. \tag{3.2}$$

$\mathbf{A}$ is the Coulomb matrix for the screening charge vector interaction. The boundary condition $\mathbf{\Phi}^{tot} = 0$ entails that the screening charges for a conductor is defined by

$$\mathbf{q} = -\mathbf{A}^{-1}\mathbf{\Phi}^{sol}\,. \tag{3.3}$$

To include the finite permittivity for real solvents, the screening charges are weighed by a suitable factor,

$$f(\varepsilon) = \frac{\varepsilon - 1}{\varepsilon + \frac{1}{2}}\,, \tag{3.4}$$

thus

$$\mathbf{q}^{\star} = f(\varepsilon)\mathbf{q}\,. \tag{3.5}$$

3.3 Molecular dynamics simulations

As mentioned in the preface, the focus of my research has since its beginning in autumn 2005 been on molecular dynamics. The sources, which have been most useful for me in this area are references [1, 52, 53]. A quite common procedure is that experienced researchers make their own molecular dynamics programs, but in the work behind Papers I and V, I have chosen to use the well-tested

and reliable Gromacs program package version 3.3 (Paper I) and version 4.5.3 (Paper V) [54–58] with an open source code adjustable to one's own needs.

In molecular dynamics, successive configurations of a system with N particles are generated by solving the Newtonian equations of motion. The result is a trajectory which describes in detail how the positions and velocities of the particles change with time.

The trajectory is received by solving the differential equations in Newton's second law ($\mathbf{F} = m\mathbf{a}$):

$$\frac{\partial^2 x_i}{\partial t^2} = \frac{\mathbf{F}_{x_i}}{m_i},\ i = 1 \ldots N. \tag{3.6}$$

This equation describes the motion of a particle with the mass m along a coordinate (x_i), where $\mathbf{F}_{x_i}$ is the force affecting the particle in that direction.

The force is calculated as the negative partial derivative of a potential function, $V(x_1, x_2, \ldots, x_N)$:

$$\mathbf{F}_{x_i} = -\frac{\partial V}{\partial x_i}. \tag{3.7}$$

3.3.1 Potential models

In a system consisting of N atoms, the potential energy can be divided into terms dependent on the coordinates of individual atoms, pairs, triplets etc. The contribution of the three-body term is fairly large, but is seldom included in computer simulations, as the calculation of whichever quantity that contains a sum over triplets of molecules is extremely time consuming. Fortunately, the pairwise approximation gives an accurate description of the properties of liquids, as the average three-body effects can be included by defining an effective pair potential by

$$\hat{V} \approx \sum_i \nu_1(\mathbf{r}_i) + \sum_i \sum_{j>i} \nu_2^{\text{eff}}(r_{ij})\,. \tag{3.8}$$

Effective pair potentials are also implemented in Gromacs for non-additive pair interactions.

An easier, and more idealized pair potential often used in computer simula-

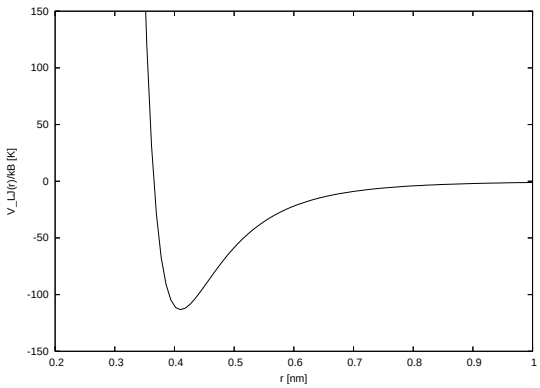


Figure 3.1: The Lennard-Jones 12-6 pair potential for O_2.

tions is the Lennard-Jones potential (LJ),

$$\nu^{\mathrm{LJ}}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right], \tag{3.9}$$

shown in Figure 3.1. It gives an adequate description of the properties of a molecule in a computer simulation if the parameters ε and σ are chosen suitably. The potential assumes an asymptotic long-ranged form of $-\frac{1}{r^6}$, which explains long-range attraction, caused mainly by correlation between the electron clouds around the atoms (van der Waals- or London dispersion). Additionally, Coulomb terms are included for charged particles. The potential has a negative well with the depth ε corresponding to cohesion in the condensed phase. Finally, the potential has a sharp repulsive wall at distances shorter than $r \sim \sigma$ caused by non-bonded overlapping of the electron clouds. The depth of the well is often given in temperature units as $\frac{\varepsilon}{k_{\mathrm{B}}}$, where k_{B} is the Boltzmann constant. In the potential in Figure 3.1, the parameter values $\frac{\varepsilon}{k_{\mathrm{B}}} = 113.3$ K and $\sigma = 365.4$ pm for O_2 have been used [59]. It should, though, be stressed that the values corresponding with experimental properties are not exactly the same values,

which would hold for an isolated pair of O_2 molecules. By deriving (3.9) with respect to r, it is easily seen that the depth of the well, *i.e.* the minimum, is located at $r_{\min} = 2^{1/6}\sigma$.

The interaction energy between the molecules depends not only on the distance between them, but also on their relative orientations and, where possible, their conformations. The interaction energy between two molecules is commonly calculated by a site model, where the interaction is determined by a sum of interactions between all pairs of sites in the two molecules. The sites are often identified by the position of the nuclei, but not necessarily.

Systems with many atoms imply inevitably that interactions between different kinds of atoms have to be calculated. Calculating the Lennard-Jones interaction energy between two carbon monoxide molecules with a two site model would require not only the parameters for the carbon–carbon interaction and the oxygen–oxygen interaction, but also for the carbon–oxygen interaction. A system consisting of N different kinds of atoms would require $\frac{1}{2}N(N-1)$ sets of parameters for the interaction between different atoms. Choosing the parameters can be difficult and time consuming, hence, it is common to assume that the parameters for the cross interactions can be received from the parameters for the separate atoms by combination rules. In the Lorentz–Berthelot combination rules, which are often used, the so-called collision diameter σ_{AB} for the interaction A–B is received as the arithmetical average of the values for two separate atoms, while the depth of the well ε_{AB} is given by the geometrical average:

$$\sigma_{AB} = \frac{1}{2}(\sigma_{AA} + \sigma_{BB}) \tag{3.10}$$

and

$$\varepsilon_{AB} = \sqrt{(\varepsilon_{AA}\varepsilon_{BB})}\,. \tag{3.11}$$

When considering larger molecules it can be necessary to model several atoms as a uniform site, such as CO_2, CH_4 or CCl_4 (see Table 4, p. 117 in reference [59]).

3.3.2 Periodic boundary conditions

Simulations are often performed on much smaller systems than the real ones, where the number of molecules is approximately Avogadro's number. In the MD simulations in Papers I and V, the number of water molecules was in each case < 6000. Despite this, such simulations are very time consuming. The system with N molecules could be limited by a potential representing a container, which keeps them from moving away from each other, but these arrangements are not enough to simulate a bulk liquid. A major hindrance for such a simulation is the large number of molecules occuring at the surface of each sample. Out of 1000 molecules arranged in a $10 \times 10 \times 10$ cube no fewer than 488 occur at the surface of the cube. Whether the cube is limited by a barrier or not, the molecules at the surface will feel very different forces than those in the bulk.

The problem with surface effects can be overcome by introducing periodic boundary conditions (PBC). This was done by Born and von Karman in 1912 [60]. The cubic box is replicated over the whole space to form an infinite lattice. When a molecule moves in the original box during the simulation, its periodic image in each neighbouring box will move in exactly the same way. When a molecule leaves the original box, one of its images will consequently come in from the opposite side. No barriers nor any surface molecules exist along the sides of the original box. This box forms a practical coordinate system to decide the coordinates of the N molecules. A two-dimensional version of a periodic system such as this is shown in Figure 3.2. The box copies are denoted arbitrarily A, B, C etc. When particle 1 moves through a side, its images $1_A, 1_B$ etc. (where the subscript denotes in which box the image is) move through their corresponding sides. The number density in the original box (and hence in the whole system) is preserved.

The coordinates for each image do not necessarily have to be saved in a simulation (an infinite number); the coordinates for the molecules in the original box is sufficient. When a molecule leaves the box by moving through a side, the

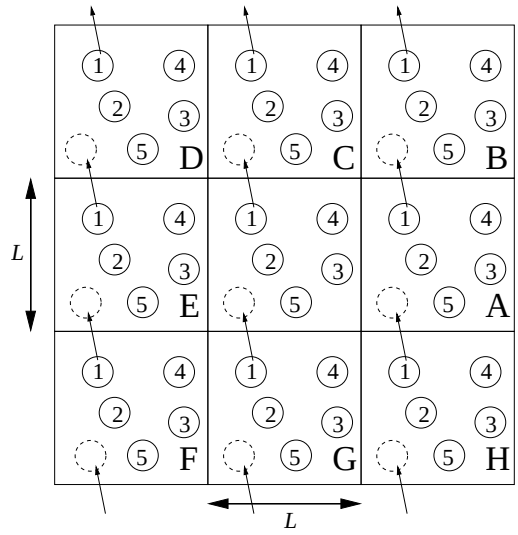


Figure 3.2: A two-dimensional periodic system.

focus can be moved to the incoming molecule instead.

It is important to question whether the properties of a small, infinite periodic system, and the macroscopic system it represents, are the same. This depends on how far the intermolecular potential reaches and the phenomenon which is studied. For a liquid consisting of Lennard-Jones atoms it should be possible to perform a simulation in a cubic box with the edge $L \approx 6\sigma$ without the particles feeling the symmetry of the periodic lattice. If the potential reaches far (*i.e.* $\nu(r) \sim r^{-v}$, where v is less than the dimension of the system), a considerable interaction will appear between a particle and its image in the neighbouring box, and consequently the symmetry of the cell structure will be forced on an isotropic liquid.

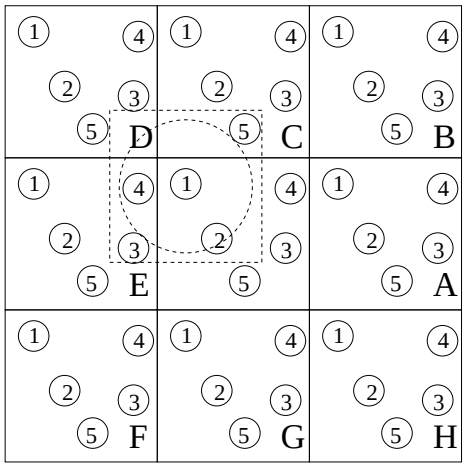


Figure 3.3: Minimum image conventionen in a two-dimensional system.

Truncation of the potential

The core of the MD programs constitutes the calculation of the potential energy for a certain configuration and the forces acting on each molecule. When calculating the force acting on molecule 1, or the contributions to the potential energy concerning molecule 1, assuming pairwise additivity, interactions between molecule 1 and all other molecules i in the simulation box have to be included. There are $N-1$ terms in this sum. In principle, all interactions between molecule 1 and the images i_a, i_b etc. in the surrounding boxes also have to be included, but then the number of terms become infinite, which is impossible to count in practice. For a potential energy function, which does not reach far, this summation can be limited by an approximation. In Figure 3.3, molecule 1 occurs in the middle of an area of the same size and shape as the original box. Molecule 1 interacts with all the molecules, whose centre is within this area, *i.e.* with the nearest periodic images of the other $N-1$ molecules. This procedure is called "minimum image convention". In Figure 3.3, molecule 1 interacts with

the molecules $2, 3_E, 4_E$ and 5_C. This technique, which follows directly from the periodic boundary conditions, was first used in simulations by Metropolis in 1953 [5].

In the minimum image convention, the calculation of the potential energy constitutes $\frac{1}{2}N(N-1)$ terms. This can still be a very heavy calculation for 1000 particles, for example. Another approximation improves the situation considerably. The largest contribution to the potential and the forces comes from the neighbours close to the molecule, and for short-range forces, spherical cut-off is usually used, which means that the pair-potential $\nu(r)$ is set to zero for $r \geq r_c$, where r_c is the cut-off distance. The dashed line in Figure 3.3 represents the cut-off, and in this case the molecules 2 and 4_E contribute to the force on 1 as their centres are inside the cut-off, while 3_E and 5_C do not contribute. In a cubic simulation box with the edge L, the number of neighbours explicitely regarded is decreased approximately by a factor of $\frac{4\pi r_c^3}{3L^3}$, and this can be a notable saving. Including spherical cut-off should be a small perturbation, and the cut-off distance should be set large enough to ensure this. An example of this is that the value of the pair-potential at the edge of a cut-off sphere with the radius $r_c = 2.5\ \sigma$, in a simulation with Lennard-Jones atoms, is just 1.6 per cent of the depth of the well.

3.3.3 Long-range forces

Spherical cut-off is especially important in molecular dynamics, as it makes the intermolecular forces finite. Spherical cut-off is always used for LJ interactions in the versions of Gromacs I used in Papers I and V. Using spherical cut-off results in the thermodynamic and the other properties of the model liquid not being exactly the same as for the non-truncated LJ liquid, but it is possible to correct for long-range forces to approximatively regain the desired information. The electrostatic algorithms for long-range forces implemented in Gromacs are the Ewald summation [61, 62], PME (particle-mesh Ewald) [63, 64] and PPPM

(Particle-Particle Particle-Mesh) [65, 66]. I will next shortly explain the Ewald summation and then PME, which were used in Papers I and V.

Ewald summation and PME

Ewald summation is a technique to effectively summarize the interaction between an ion and all its periodic images. It was originally developed in the study of ionic crystals [61,62]. In Figure 3.2, ion 1 interacts with the ions 2, 2_A, 2_B and all other images of 2. The potential energy can be written as

$$\mathcal{V}^{zz} = \frac{1}{2}\sum_{\mathbf{n}}{}' \left(\sum_{i=1}^{N}\sum_{j=1}^{N} z_i z_j |\mathbf{r}_{ij} + \mathbf{n}|^{-1} \right), \tag{3.12}$$

where z_i and z_j are the charges. All factors $4\pi\varepsilon_0$ are omitted to make the notation more simple. The sum over $\mathbf{n}$ is the sum over all cubic lattice points, $\mathbf{n} = (n_x L, n_y L, n_z L)$, where n_x, n_y, n_z are integers. The prime denotes that $i = j$ is omitted for $\mathbf{n} = 0$. For long-range potentials this sum is conditionally convergent, *i.e.* the result depends on in which order the terms are summed. A natural choice is to take the boxes in order of their relation to the main box. The unity cells are added in order: in the first term $|\mathbf{n}| = 0$, *i.e.* $\mathbf{n} = (0, 0, 0)$; the second term $|\mathbf{n}| = L$ consists of the six boxes, which have their centres at $\mathbf{n} = (\pm L, 0, 0), (0, \pm L, 0), (0, 0, \pm L)$ etc. When more terms are added to the sum, the infinite system is layered roughly spherically as shown in Figure 3.4 [1]. When this procedure is applied, the character of the surrounding medium should be specified, especially its relative permittivity (dielectric constant) ε_r. The results for a sphere surrounded by a conductor such as a metal ($\varepsilon_r = \infty$) differs from the results for a sphere surrounded by vacuum ($\varepsilon_r = 1$) [67]:

$$\mathcal{V}^{zz}(\varepsilon_r = \infty) = \mathcal{V}^{zz}(\varepsilon_r = 1) - \frac{2\pi}{3L^3}\left| \sum_i z_i \mathbf{r}_i \right|^2 . \tag{3.13}$$

This equation can be applied within the limit of a large sphere of boxes. When

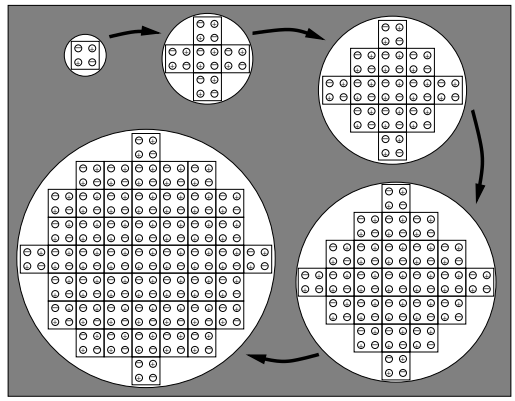


Figure 3.4: Construction of the sphere of simulation boxes. For simplicity just a small system with two ion pairs is illustrated. The shaded area represents an outer dielectric continuum with the relative permittivity ε_r.

all factors are taken into account, the final potential energy will include a sum for the real space plus a sum for the reciprocal space minus a self term plus the surface term in equation 3.13. The final result is the following:

$$\begin{aligned}\mathcal{V}^{zz}(\varepsilon_r = 1) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Bigg(\sum_{|\mathbf{n}|=0}^{\infty}{}' z_i z_j \frac{\operatorname{erfc}(\kappa|\mathbf{r}_{ij}+\mathbf{n}|)}{|\mathbf{r}_{ij}+\mathbf{n}|} \\ + \frac{1}{\pi L^3}\sum_{\mathbf{k}\neq\mathbf{0}} z_i z_j \left(\frac{4\pi^2}{k^2}\right) e^{\frac{-k^2}{4\kappa^2}} \cos(\mathbf{k}\cdot\mathbf{r}_{ij})\Bigg) \\ - \left(\frac{\kappa}{\sqrt{\pi}}\right)\sum_{i=1}^{N} z_i^2 + \left(\frac{2\pi}{3L^3}\right)\left|\sum_{i=1}^{N} z_i \mathbf{r}_i\right|^2 ,\end{aligned} \tag{3.14}$$

where $\operatorname{erfc}(x)$ is the complementary error function

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^{\infty} e^{-t^2}\mathrm{d}t \,, \tag{3.15}$$

which goes to zero as x increases. If κ is large enough, the only term contributing to the sum in the real space will be the one with $n = 0$, which reduces the first term to the minimum image method. The second term is a sum over reciprocal vectors $\mathbf{k} = \frac{2\pi\mathbf{n}}{L^2}$. The idea with the Ewald summation is merely to reshape a simple, slowly converging sum into two rapidly converging sums and a constant term.

To perform the reciprocal sum in an improved way, the PME method was presented by Tom Darden [63,64]. Instead of directly summing up wave vectors, the charges are ascribed a grid by using a cardinal B-spline interpolation [68, 69]. This grid is then Fourier transformed with a 3D FFT algorithm, and the reciprocal energy term is received as a simple sum over the grid in k-space. The potential at the mesh points is calculated by invers transformation, and by using the interpolation factors, the forces on each atom are received. The PME algorithm is considerably faster than the regular Ewald summation on slightly larger systems, but for very small systems, it might still be worth using the Ewald summation to avoid the expenses of setting up the grid and transforming.

3.3.4 Ensembles

In computer simulations, the properties of a system consisting of a number of particles are usually of interest. These particles can be distributed on different states, and all possible states which a quantum mechanical system can be in, form an ensemble. Each representative in the ensemble has its own energy, and the distribution within the ensemble follows the Boltzmann distribution. In the following paragraph, the four most common ensembles are presented.

The ensemble with the number of particles N, the volume V and the energy E fixed for each state is called microcanonical. Molecular dynamics are usually run under these conditions. In a traditional Monte Carlo simulation, samples are taken in the canonical ensemble, where N, V and the temperature, T, are kept constant. In the isothermic-isobaric ensemble (with N, T and the pressure, P,

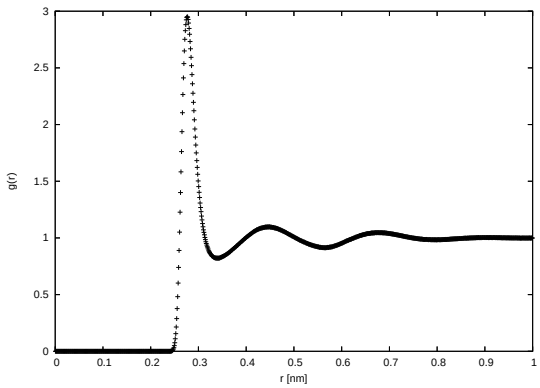


Figure 3.5: Pair distribution function for the oxygen atoms of water.

fixed) the number of particles are constant, but in a grand-canonical ensemble, the composition can be changed, *i.e.* the number of particles can increase or decrease.

3.3.5 Structural properties

The structure of a simple, monoatomic liquid is characterized by a set of distribution functions for the positions of the atoms, the simplest of which is the pair distribution function $g_2(\mathbf{r}_i, \mathbf{r}_j)$ or $g_2(r_{ij})$ or simply $g(r)$. This function gives the probability to find a pair of atoms at the distance r from each other, correlated to the probability expected for a totally random distribution at the same density: hence $g(r)$ is dimensionless. Figure 3.5 shows the pair distribution function for the oxygen atoms of water in a 500 ps MD simulation of a 0.023 M NaCl solution at the temperature 300 K. The pair distribution function is useful, not just because it gives insight to the structure of a liquid, but also because it can be used to express the ensemble average, the energy or even the chemical potential, although a direct evaluation of these properties are usually more accurate in practice.

3.3.6 Neighbour lists

When performing an MD simulation, the system consisting of molecules consisting of particles with well-defined interaction functions, has to be defined. The topologi of the molecule has to be specified, and the force field (*cf.* Chapter 3.3.10) chosen. From the beginning, Gromacs was developed for biopolymers, so the automatic topology constructors are not much of a help for other molecules. Constructing the topology for HMR in Paper I (with 51 bonds, 87 angles and 124 torsions) or for the LIGNOLs in Paper V, as well as finding out suitable atom types for each atom, was the task, which demanded the greatest deal of work.

To get the dynamics running, the internal forces between non-bonded particles for the pairs i, j, whose distance r_{ij} (between i and its closest image j) is less than the cut-off radius r_c, have to be generated. Some of the pairs of particles falling within this criteria are excluded, while their interaction is already taken into account in the bonded interactions. Gromacs uses a pair list, which contains the pairs of particles for which the non-bonded forces need to be calculated. Several ways of constructing, maintaining and updating these lists occur in the literature [65, 70].

3.3.7 Integrating the equations of motion

When the starting conditions are ready, new configurations of the system should be generated by integrating the Newtonian equations of motion (3.6).

Perhaps the most used method for integrating the equations of motion is the one first applied by Verlet [70] in 1967 and later assigned to Störmer [71]. This method is a direct solution of the second order equations (3.6). In the method, the positions $\mathbf{r}(t)$, the accelerations $\mathbf{a}(t)$ and the positions $\mathbf{r}(t - \delta t)$ from the previous step are assumed. The equation to move the positions forward is:

$$\mathbf{r}(t + \delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \delta t) + \delta t^2 \mathbf{a}(t) \,. \tag{3.16}$$

Several things need to be pointed out in equation (3.16). The velocities do not occur at all. They have been eliminated in the addition of those equations, which have been received by Taylor expansion around $\mathbf{r}(t)$:

$$\begin{aligned} \mathbf{r}(t+\delta t) &= \mathbf{r}(t) + \delta t \mathbf{v}(t) + \frac{1}{2}\delta t^2 \mathbf{a}(t) + \dots \\ \mathbf{r}(t-\delta t) &= \mathbf{r}(t) - \delta t \mathbf{v}(t) + \frac{1}{2}\delta t^2 \mathbf{a}(t) - \dots \end{aligned} \tag{3.17}$$

The velocities are not needed in the calculation of the trajectories, but they are useful in the calculation of the values of the kinetic energy (and hence the total energy). They can be obtained from

$$\mathbf{v}(t) = \frac{\mathbf{r}(t+\delta t) - \mathbf{r}(t-\delta t)}{2\delta t} \,. \tag{3.18}$$

While equation (3.16) is correct with exception of error of the order δt^4 (the local error), the velocities in equation (3.18) depends on error of the order δt^2. More accurate estimations of $\mathbf{v}(t)$ can be done if more variables are stored, but this increases the perturbation already included in (3.18), *i.e.* $\mathbf{v}(t)$ can only be calculated if one knows $\mathbf{r}(t+\delta t)$. It is also worth noting that the Verlet algorithm is well-centered (*i.e.* $\mathbf{r}(t-\delta t)$ and $\mathbf{r}(t+\delta t)$ are symmetric in equation (3.16)), which makes it time-reversible. A drawback to the Verlet algorithm is, though, that the velocities are treated in a quite clumsy and unpractical way, and according to Dahlquist and Björk [72], the shape of the algorithm can initiate unnecessary lack of exactness. This is because a small term is added to the difference between large terms in equation (3.16) to generate the trajectory.

Leap-frog

Modifications of the original Verlet scheme have been suggested in order to tackle these limitations. One of these is the so-called half-step leap-frog scheme [65, 73, 74]. The origin of the name becomes obvious when the algorithm is

printed:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t\,\mathbf{v}(t+\frac{1}{2}\delta t) \tag{3.19}$$

$$\mathbf{v}(t+\frac{1}{2}\delta t) = \mathbf{v}(t-\frac{1}{2}\delta t) + \delta t\,\mathbf{a}(t)\,. \tag{3.20}$$

The properties which are stored are the present positions $\mathbf{r}(t)$ and the accelerations $\mathbf{a}(t)$ together with the middle-step velocities $\mathbf{v}(t-\frac{1}{2}\delta t)$. The velocity equation (3.20) is first executed and the velocities jumps over the coordinates to give the next middle-step velocity $\mathbf{v}(t+\frac{1}{2}\delta t)$. During this step the present velocities can be calculated as

$$\mathbf{v}(t) = \frac{1}{2}\left(\mathbf{v}(t+\frac{1}{2}\delta t) + \mathbf{v}(t-\frac{1}{2}\delta t)\right)\,. \tag{3.21}$$

This is necessary to be able to calculate the energy at time t, as well as which properties ever, which need the positions and the velocities at the same moment. If this is followed, equation (3.19) is used to once more force the positions in front of the velocities. After this the new accelerations can be evaluated to be used in the next step, as shown in Figure 3.6. If the velocities are eliminated from these equations, it shows that the method is algebraically equivalent to the Verlet algorithm. There are, though, a few advantages in programming the equations (3.19–3.21), as the velocities (indeed not at time t) occur explicitly. To adjust the simulation energy is often achieved by scaling the velocities suitably. Numerical advantages derive from the fact that differences between two large properties are not calculated at any step to receive a small property. This minimizes the loss of accuracy on a computer. If storage place has to be preserved, the accelerations can be directly stored upon the velocities, hence decreasing the total need to the order of $6N$ words [75]. The cost is that equation (3.21) cannot be used anymore, and that it becomes necessary to calculate the value of the kinetic energy at time t from the known middle-step values. Finally it can be noted that the leap-frog procedure can be applied also to algorithms other than Verlet's [75].

Velocity Verlet

As can be seen in equation (3.21), the velocities are still not treated in a satisfying way in the leap-frog method. An algorithm, equivalent to Verlet's, that saves positions, velocities and accelerations, all at the same time t and minimizing rounding errors, was presented in 1982 by Swope, Andersen, Berens and Wilson [76]. This "Velocity Verlet algorithm" has the shape

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t\,\mathbf{v}(t) + \frac{1}{2}\delta t^2\,\mathbf{a}(t) \tag{3.22}$$

$$\mathbf{v}(t+\delta t) = \mathbf{v}(t) + \frac{1}{2}\delta t[\mathbf{a}(t) + \mathbf{a}(t+\delta t)]\,. \tag{3.23}$$

The Verlet algorithm can again be received by eliminating the velocities. The algorithm requires that only $\mathbf{r}$, $\mathbf{v}$ and $\mathbf{a}$ are stored. It contains two steps, with a calculation of the forces in between. First the new positions at time $t+\delta t$ are calculated with equation (3.22) and the middle-step velocities are calculated with

$$\mathbf{v}(t+\frac{1}{2}\delta t) = \mathbf{v}(t) + \frac{1}{2}\delta t\mathbf{a}(t)\,. \tag{3.24}$$

Then the forces and accelerations at time $t+\delta t$ are calculated, and the velocities are received readily as

$$\mathbf{v}(t+\delta t) = \mathbf{v}(t+\frac{1}{2}\delta t) + \frac{1}{2}\delta t\,\mathbf{a}(t+\delta t)\,. \tag{3.25}$$

At this point, the kinetic energy at time $t+\delta t$ is available. The potential energy at this time has been evaluated from the force loop. The whole process is shown in Figure 3.6 [1]. The method needs $9N$ words of memory, and its numerical stability, comfort and simplicity makes it perhaps the most attractive algorithm.

Furthermore, Beeman's [77] investigations of several algorithms can be mentioned, of which one is reduced to equation (3.16) when the velocities are elim-

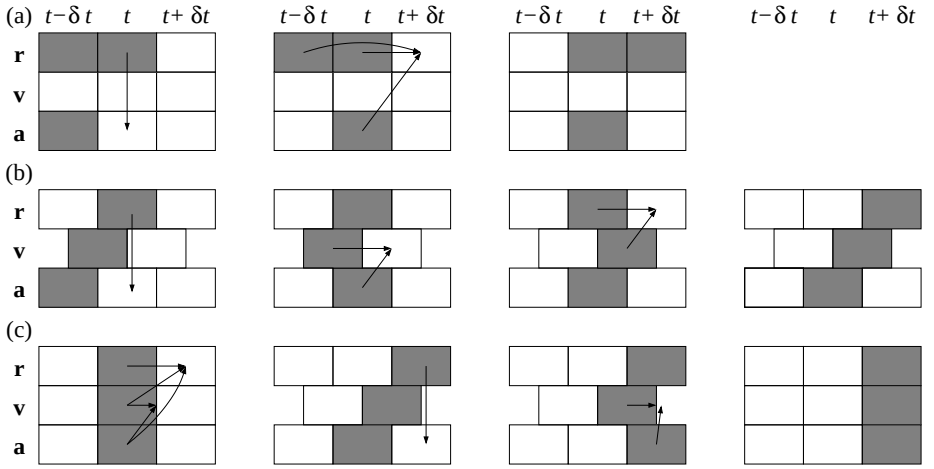


Figure 3.6: Different kinds of the Verlet algorithm. (a) The original method by Verlet. (b) The leap-frog variety. (c) Velocity variety. The successive steps in the execution of the algorithms are shown. For each case the variables are stored in the gray boxes.

inated [65]. The algorithm is as follows:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t\,\mathbf{v}(t) + \frac{2}{3}\delta t^2\,\mathbf{a}(t) - \frac{1}{6}\delta t^2\,\mathbf{a}(t-\delta t) \tag{3.26}$$

$$\mathbf{v}(t+\delta t) = \mathbf{v}(t) + \frac{1}{3}\delta t\,\mathbf{a}(t+\delta t) + \frac{5}{6}\delta t\,\mathbf{a}(t) - \frac{1}{6}\delta t\,\mathbf{a}(t-\delta t)\,. \tag{3.27}$$

In this method $\mathbf{r}(t)$, $\mathbf{v}(t)$, $\mathbf{a}(t)$ and $\mathbf{a}(t-\delta t)$ are stored. The counterweight to the complexity of these formulas and the need for storing the "old" accelerations is a more accurate velocity equation than (3.18) and consequently a notable improvement of the conservation of energy. However, all the methods, which are described here, are generally equivalent in that they have identical global errors, and *de facto* generate identical position trajectories.

In the literature there are still alternative algorithms [71, 78], which have been reviewed and compared in the references [79, 80].

3.3.8 Temperature and pressure

Controlling the temperature and pressure of a system is often necessary. The equations of motion are modified for the coupling of the temperature, and possibly the pressure. In Gromacs, temperature coupling can be performed either by Berendsen's scheme for weak coupling [81], which has been applied in Papers I and V, or with Nosé-Hoover's scheme [82, 83]. Both have been compared in reference [84]. In Gromacs, pressure coupling can be performed either by Berendsen's scheme [81], where the coordinates and the simulation box vectors are changed to scale for each step, or by Parrinello-Rahman's procedure [85, 86]. In Papers I and V, Berendsen's scheme for the pressure coupling was used.

With Berendsen's temperature algorithm, an attempt was made to imitate a weak coupling with first order kinetics to an external heat bath with the given temperature, T_0. The result of this algorithm is that a small deviation in the temperature of the system from T_0 is slowly corrected by

$$\frac{dT}{dt} = \frac{T_o - T}{\tau} \,, \tag{3.28}$$

which means that a temperature deviation decreases exponentially with a time constant τ.

In the Berendsen pressure coupling algorithm, the coordinates and the simulation box vectors are changed to scale for each step, by a matrix, μ, which has the consequence of a first order kinetic relaxation of the pressure against a given reference pressure, $\mathbf{P}_0$:

$$\frac{d\mathbf{P}}{dt} = \frac{\mathbf{P}_o - \mathbf{P}}{\tau_p} \,. \tag{3.29}$$

The matrix, μ, is given by

$$\mu_{ij} = \delta_{ij} - \frac{\Delta t}{3\tau_p}\beta_{ij}\big(P_{oij} - P_{ij}(t)\big) \,, \tag{3.30}$$

where β is the isotermic compressibility of the system.

3.3.9 Energy minimization

In chapter 3.1, structure optimizations of molecules are discussed. When an optimized structure is placed in an MD simulation box filled with some model of a solvent, *e.g.* TIP4P for water (*cf.* Chapter 3.3.10), the system might be not close to the equilibrium. The forces on the atoms can be extremly large, and the MD simulation can be interrupted before it even starts. In such a case, a rough energy minimization is needed for the system. In Gromacs, different methods for energy minimization are offered before starting a simulation: 1) steepest descent, which uses information from the (first) derivative, 2) conjugate gradient (see for example [87]), which uses the information of the gradient from the previous step and 3) L-BFGS [88, 89], which is based on the quasi-Newton minimization method BFGS (from the creators Broyden, Fletcher, Goldfarb and Shanno) [90–93], but which requires much less memory than the original method [94].

To minimize the starting structures of the systems in Papers I and V, steepest descent has been used, where a step is taken in the direction of the negative gradient, *i.e.* the force, without caring about what has been built up during previous steps. The step length is adjusted so that the minimization is rapid, but the direction always downwards. In general, the local minimum is reached quickly with steepest descent, while the conjugate gradient comes very close to the local minimum, but works more poorly further away from the minimum.

3.3.10 Force fields

As mentioned in chapter 3.3.6, a force field must be chosen before the simulations can be started. A force field consists mainly of two parts: 1) a set of equations (potential functions, *cf.* chapter 3.3.1), which are used to generate the potential energies and their derivatives (forces) and 2) the parameters needed

for constructing the equations. In Gromacs, several different force fields can be used, but one of the most common is OPLS-AA [95], which was used in the MD simulations in Papers I and V. Water as a solvent was described by the TIP4P model [96], which is rigid to save computational resources, but still interacts with its neighbours in a "correct" way. In TIP4P, the point charge is moved from the oxygen atom to the mass centre of the molecule. The edges of the simulation boxes were between 5.1–5.6 nm (the volume 131–174 nm^3), and the number of water molecules varied between 4265 and 5795.

In the force field OPLS-AA (all-atom), the atoms are defined separately instead of uniting the atoms in a methyl group to one unity. Gromacs also has to be aware of which atoms the different contributions to the potential functions should act on and which parameters should be used in the different functions. OPLS-AA is a non-reactive force field, where no bonds can be broken or formed. The interactions are divided into bonded and non-bonded, and a number of the non-bonded interactions (the first two neighbours) have to be omitted because they have already been taken into account among the bonded interactions.

As mentioned in chapter 3.3.6, suitable atom types have to be chosen for the system, which entails a number of statistical properties such as mass, charge and the parameters ε and σ in equation (3.9) being assigned to the concerned atom type from different topology parameter files in Gromacs. The bonded parameters (bonds, angles and torsions) are located in a separate file and are combined with the non-bonded (van der Waals parameters), according to the topology file one has constructed. Additional Lennard-Jones and electrostatic interactions can be added further between pair of atoms if necessary. In Gromacs, pair interactions are used to modify 1–4 interactions, but in OPLS-AA, these are calculated by scaling and are therefore unnecessary. In Tables 5.3 and 5.4 in [53], all possibilities to introduce interactions in the topology file are listed.

Chapter 4

Results

In the results part of this thesis, I will first present the results related to HMR (Papers I and II) and then the LIGNOL related studies (Papers III–V).

4.1 HMR

4.1.1 MD simulations of HMR

In Paper I, the aim was to be able to run MD simulations on HMR with the program package Gromacs, which is tailor-made for proteins and amino acids. The basis of this work was the quantum chemical conformational analysis on HMR made by Dr. Antti Taskinen five years earlier [16]. In this connection, it was also appropriate to try the solvent model COSMO for the quantum chemical optimizations.

Notations

The code for the conformations is directly adopted from [16]. The torsional angles are described in detail in Figure 4.1, and the symbols of the conformations

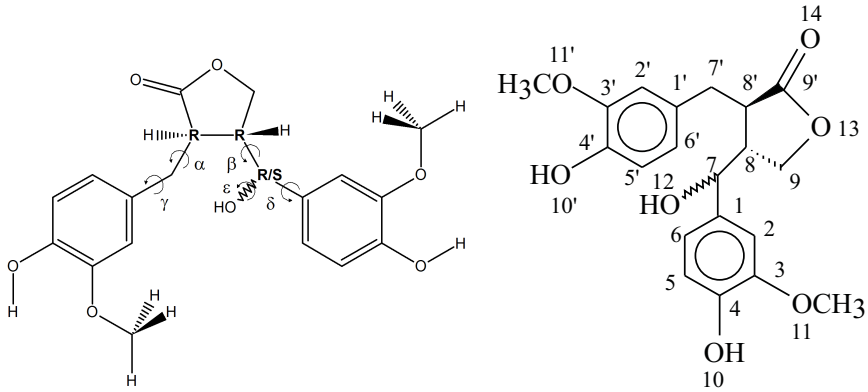


Figure 4.1: The five torsional angles and the three chiral centra in HMR as well as the old lignan way of numbering of atoms in HMR.

Table 4.1: The code for the angles α and β (left) and the angles γ and δ (right)

Angle	60°	180°	300°
α	**e**xtended	**d**own	**u**p
β	**e**xtended	**d**own	**u**p

Angle	$\gamma = 100°$	$\gamma = 270°$
$\delta = 100°$	**4**	**3**
$\delta = 270°$	**1**	**2**

are defined in Table 4.1. The conformation of the butyrolactone ring has the symbol A or B as described by Taskinen *et al.* in ref. [16].

Solvent effects

In Tables 4.2 and 4.3, the electronic energies and the dipole moments, respectively, are expressed in relation to the most stable conformer in the gas phase, *i.e.* RRR_ue_4_B. The first two columns show the electronic energies or dipole moments in the gas phase. In the second column, the level of theory used was BP86/SV(P). In the columns 4–6, the values from the studies in the presence of water ($\varepsilon_r = 78.39$) using the COSMO model are reported. The methods

are single point calculations (sp), reoptimization at the RHF/SV(P) level, and reoptimization at the DFT/B3LYP/TZVP level of theory.

As can be seen in Table 4.2 the energies from the COSMO calculations follow the same order as the conformers in the gas phase, but when looking at the values from the PCM calculations in [16] one can see that the order is almost opposite.

Table 4.2: Relative energies in kJ/mol including solvation effects

Conformation	$E_{HF}^{[16]}$	E_{DFT}	$E_{PCM}^{[16]}$	$E_{COS,sp}$	$E_{COS,HF}$	$E_{COS,DFT}$
RRR_ee_2_B	12.9	6.3	-1.4	6.6	5.8	1.5
RRR_ue_1_A	9.9	9.8	-1.8	3.6	3.4	2.5
RRR_ue_4_B	0	0	0	0	0	0
RRS_ee_3_B	7.8	5.7	-2.5	3.8	3.3	2.4
RRS_ue_3_B	5.3	5.4	9.9	1.4	1.0	0.3
RRS_uu_4_B	9.5	6.7	-0.4	5.9	6.1	3.1

When considering the dipole moments in Table 4.3, one can observe that the order between the conformers is preserved almost entirely for each method used, also for the PCM method, which gave somewhat different relative energies in [16]. Generally, the dipole moment always increases when a polar molecule is solvated. This is also true for the solvation models, as can be seen in Table 4.3. The reason for this is the fact that a solvated dipole induces an electric field on the solvent, which in turn causes a reaction field upon the solute within the cavity.

Molecular dynamics

The initial values of each of the five torsional angles mentioned in Figure 4.1 and Table 4.1 are shown in Table 4.4. One can see that they vary some degrees from the values indicated by the code. In Table 4.4, the torsional angles in the DFT optimized structures (column 2 in Tables 4.2 and 4.3) are shown as well, but the HF structures are those used as starting structures in the MD simulations.

Table 4.3: Dipole moments in Debye including solvation effects

Conformation	$\mu_{HF}^{[16]}$	μ_{DFT}	$\mu_{PCM}^{[16]}$	$\mu_{COS,sp}$	$\mu_{COS,HF}$	$\mu_{COS,DFT}$
RRR_ee_2_B	5.4	4.9	6.3	6.3	5.9	5.2
RRR_ue_1_A	3.2	3.3	4.0	3.9	3.6	3.5
RRR_ue_4_B	3.4	3.5	4.6	4.3	4.7	5.1
RRS_ee_3_B	4.8	4.6	5.4	5.3	6.2	5.3
RRS_ue_3_B	7.6	7.7	9.7	9.5	10.3	10.5
RRS_uu_4_B	4.4	4.2	5.3	5.5	5.3	5.2

Table 4.4: The initial (HF) and the DFT optimized torsional angles

Conformation	*Method*	α	β	γ	δ	ε
RRR_ee_2_B	HF	84.7°	64.9°	271.1°	267.2°	186.1°
	DFT	90.6°	65.4°	294.3°	265.8°	183.3°
RRR_ue_1_A	HF	297.2°	67.9°	78.2°	261.0°	188.9°
	DFT	296.6°	65.9°	76.6°	261.0°	182.0°
RRR_ue_4_B	HF	286.1°	64.8°	74.5°	78.0°	183.7°
	DFT	287.3°	62.4°	74.0°	78.0°	178.4°
RRS_ee_3_B	HF	93.3°	58.4°	279.9°	73.2°	181.2°
	DFT	91.1°	57.1°	286.4°	75.5°	184.7°
RRS_ue_3_B	HF	287.9°	58.9°	260.4°	67.7°	182.4°
	DFT	288.3°	55.6°	261.7°	67.3°	184.8°
RRS_uu_4_B	HF	291.2°	295.0°	79.7°	88.4°	174.3°
	DFT	290.4°	295.6°	78.1°	89.9°	179.7°

The Figures 4.2–4.3 show the changes in the five torsional angles during 5 ns simulations of the different conformations of HMR. The figures are plotted by *gnuplot*, *version* 4.0, and they include a smoothing using the Bezier algorithm, *i.e.* an approximation of the data with a Bezier curve of degree n (where n equals the number of data points) that connects the endpoints.

For *RRR* it was observed that all simulations ended in the conformation **de**. The reason for this was not found in the relative energies on any level of theory. Table 3 in [16], however, shows that **de** has by far the highest dipole moment of the *RRR* conformers studied, thus the reason is really a solvation effect. This

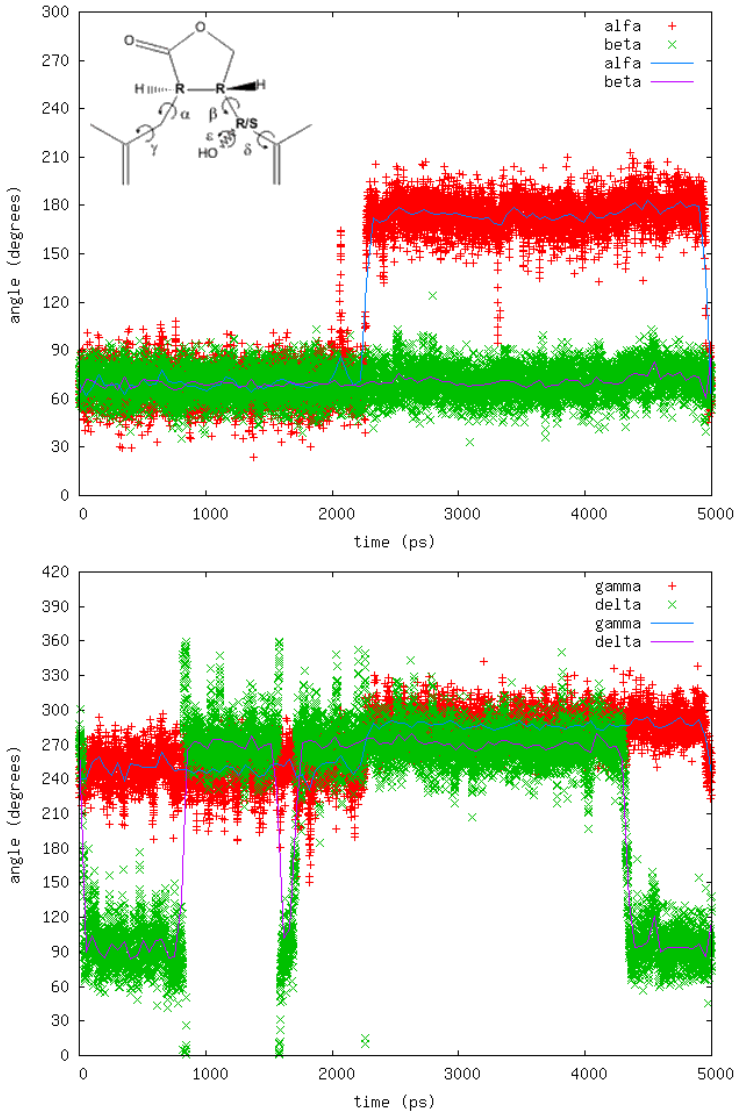


Figure 4.2: Representative torsional angles α and β (above) and γ and δ (below) in RRR_ee_2_B as a function of time, extracted from 5 ns of NVE simulation at 298 K.

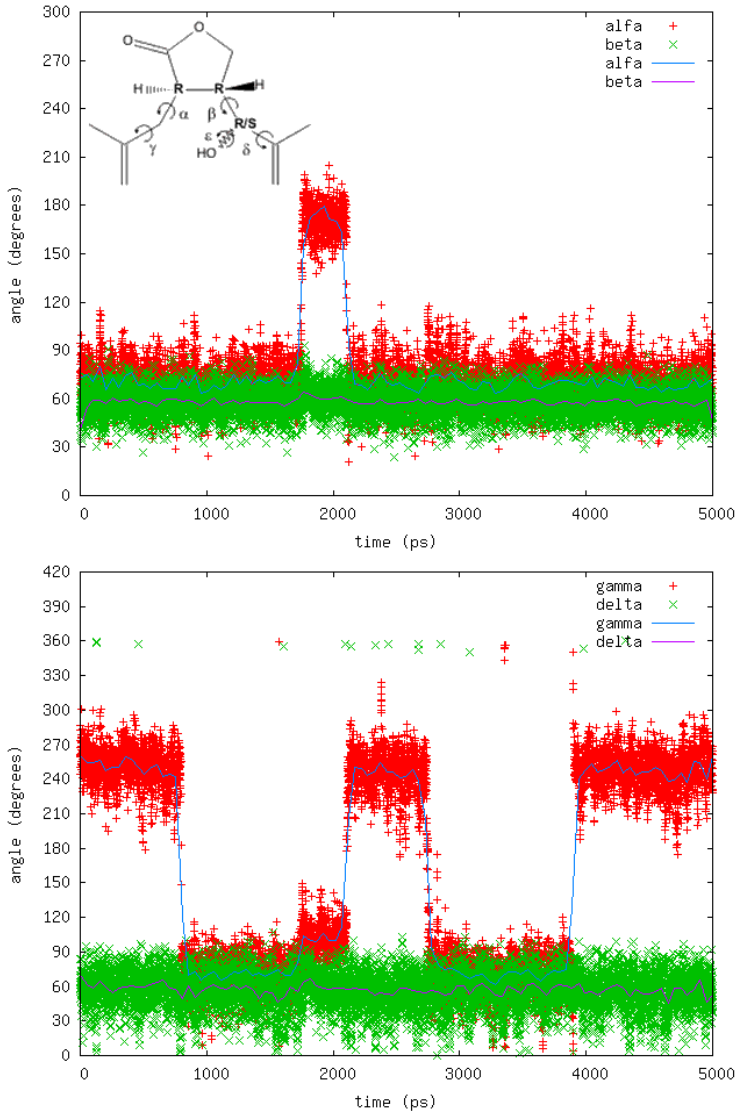


Figure 4.3: Representative torsional angles α and β (above) and γ and δ (below) in RRS_ee_3_B as a function of time, extracted from 5 ns of NVE simulation at 298 K

proves that quantum chemical calculations (in gas phase) do not always predict reliable trends for how molecules act in solvents.

This conclusion would also entail that the most favourable conformation of *RRS* in aqueous solution would be **ue_3_B**. Instead **ee** seems to be preferred, and in the last simulation, **eu** is the most frequently occuring conformation.

However, it should also be noted that not only the torsional angles α and β determine the structure. From Table 3 in [16] one can see that the difference in dipole moment between RRS_ue_3_B and RRS_ue_4_A is almost 4 D in the gas phase calculations.

Finally it can be stated that:

- The torsional angle β defined in Figure 4.1 seems to be preferred at 60°, while the other could vary considerably.
- The torsional angles α and γ were strongly correlated by the following pattern: $\alpha \rightarrow 70° \Rightarrow \gamma \rightarrow 70°$, $\alpha \rightarrow 180° \Rightarrow \gamma \rightarrow 100°$ and $\alpha \rightarrow 170° \Rightarrow \gamma \rightarrow 280°$.
- When γ and δ change conformation from **4** $\rightarrow$ **2** most often this seems to go via conformation **1**.

For the torsional angle ε, Taskinen *et al.* [16] stated that it is 180°, and the same holds also in the MD simulations on *RRR*. However, in the right-hand picture in Figure 9 in Paper I, it is obvious that the value 270° for ε in *RRS* is almost as probable as 180°. This is probably due to the interaction between the lone pairs of oxygen in the reactive hydroxyl group and the hydrogens on the other side of the butyrolactone ring. The distances between these atoms are about 240 pm as shortest, so hydrogen bonds can definitely appear especially via a water molecule. However, because of the fact that β stays almost constantly at 60°, this interaction cannot appear in the *RRR* isomer.

In Paper I, the hydration details around HMR were studied as well. In HMR there are three hydrogen bonding donors and altogether seven oxygen atoms, which work as hydrogen bonding acceptors. The number of hydrogen bonds per timeframe between HMR and the TIP4P solvent, averaged over the whole

trajectories, was found to be 7.6 for the *RRR* conformers and 7.7 for *RRS* conformers of about 10^7 possible hydrogen bonds in the whole system. The mean value of the hydrogen bond lengths is approximately 0.28 nm, which is just as expected, as the distance between O and H in hydrogen bonds is usually about 0.18 nm and the O–H bond length is just below 0.10 nm. The average lifetime of the uninterrupted hydrogen bonds was calculated to 1.2 ps.

4.1.2 Catalysis reactions with HMR

In Paper II, my part of the study was to investigate by quantum chemical calculations whether there are any noticeable differences between the two diastereomers of HMR, *i.e.* *RRR* (called HMR **1** in Paper II) and *RRS* (HMR **2**) (*cf.* Figure 4.4). As stated previously, other lignans can be synthesized using

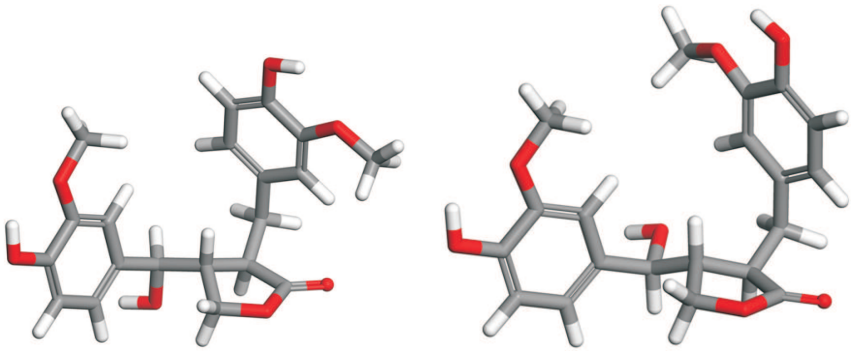


Figure 4.4: The two diastereomers of HMR: *RRR* (left) and *RRS* (right).

HMR as starting material (*cf.* Paper II), and both in the dehydrogenation to oxomatairesinol (oxoMAT) and hydrogenolysis to matairesinol (MAT) (see Figure 4.5) HMR **2** reacted with a higher reaction rate [10, 11]. Also the yield of oxoMAT was higher in the reaction with HMR **2**.

If one of the diastereomers has weaker bond strength or considerably different

charge in the reactive hydroxyl group, it could explain why the two diastereomers have different reaction rates. Therefore, Mulliken and Löwdin charges, electrostatic potential fit (ESP) and bond orders as well as bond distances were calculated for the optimized structures at HF/6-31G* level. All relevant parameters were essentially the same for HMR **1** and HMR **2**, as can be seen in Table 2 in Paper II.

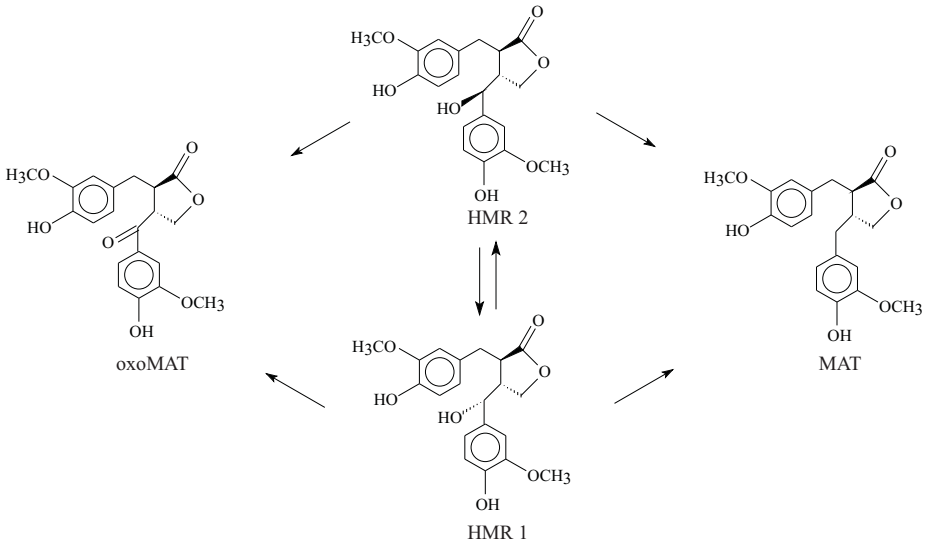


Figure 4.5: Dehydrogenation of HMR to oxomatairesinol (oxoMAT) and hydrogenolysis to matairesinol (MAT).

Protonation och deprotonation

As the role of the Brønsted acid sites present on the catalyst support is apparent [10, 11], protonation of HMR **1** and HMR **2** was studied next. Protonation has been proposed as the first step in the formation of MAT [10].

The HMR molecule has seven potential proton acceptors, namely the oxygen

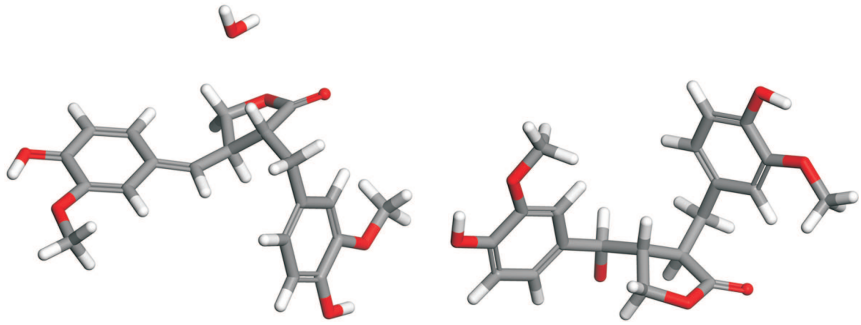


Figure 4.6: Protonated and deprotonated form of HMR **1**.

atoms. Besides, there is a possibility to form bifurcated bonding between the oxygen atoms outside the phenyl rings. The protonation energies varied between 800 and 909 kJ/mol for HMR **1** and between 785 and 909 kJ/mol for HMR **2**. The most stable structures were formed when the proton was attached to the reactive hydroxyl oxygen. In these cases, the protonation energy, here defined as the negative electronic energy change for the reaction:

$$\mathrm{HMR} + \mathrm{H}^+ \rightarrow \mathrm{HMR} - \mathrm{H}^+ \,, \tag{4.1}$$

was 911 and 915 kJ/mol for HMR **1** and HMR **2**, respectively. However, when attaching the proton to the reactive hydroxyl oxygen, the optimization did not result in protonated HMR **1** or HMR **2** molecules, instead a water abstraction and consequently a complex of water molecule and carbenium ion was observed (see Figure 4.6). In other words, protonation of the hydroxyl group leads to formation of a carbenium ion without any activation barrier. This is in line with the experiments as the corresponding hydroxyl group is the reactive one. When also taking into account thermodynamic contributions, *i.e.* calculating the Gibbs' free energy at 25 °C at the DFT/B3LYP/TZVP level, the formed complex of HMR **2** was 11.0 kJ/mol more stable than that of HMR **1**. If water

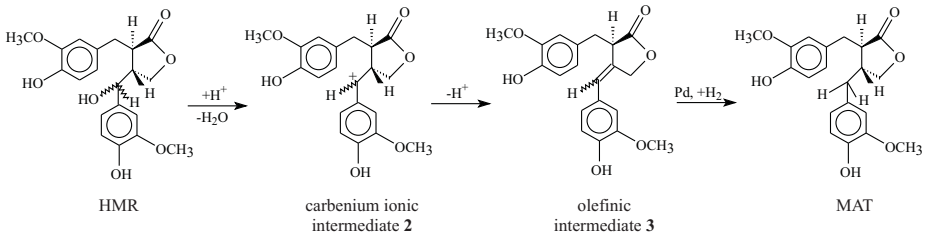


Figure 4.7: Reaction mechanism for hydrogenolysis of HMR to MAT.

is removed, *i.e.* only carbenium ions are considered, the carbenium ion formed from HMR **2** is 20.3 kJ/mol more stable than that formed from HMR **1**. As water is removed, the hydroxyl carbon becomes achiral, which means that the formed carbenium ions are only different conformations not different molecules as HMR **1** and HMR **2** are.

Deprotonation followed by hydride abstraction is a possible mechanism for the formation of oxoMAT. The deprotonated structures of HMR **1** and HMR **2** were optimized at the DFT/B3LYP/TZVP level, and both anions were observed to be equal in stability at 25 °C, *i.e.* the anion formed from HMR **2** is only 0.2 kJ/mol more stable than the anion formed from HMR **1**. The protonated form of HMR **1** is shown in Figure 4.6.

Reaction mechanisms

As proposed, protonation can be the initial step in formation of MAT [10]. The formed carbenium ionic intermediate is then attacked by a hydride resulting in formation of MAT. This mechanism is shown in Figure 4.7.

The possible mechanism for formation of oxoMAT is shown in Figure 4.8. The initial step is deprotonation of the hydroxyl group and formation of an anionic intermediate. Though, this mechanism is dubious, as the reaction works better in acidic conditions, which does not endorse any kind of deprotonation, but protonation. In an alternative mechanism, which is known in the presence

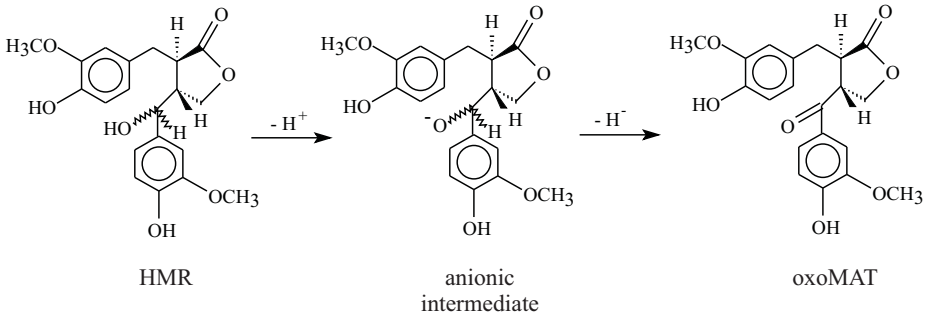


Figure 4.8: Reaction mechanism for dehydrogenation of HMR to oxoMAT.

of DDQ as a catalyst [16], the hydrogen atom leaves as a negative hydride ion, *i.e.* opposite to a proton, and has nothing to do with acidity. In the dehydrogenation reactions in Paper II, hydrogen gas is abstracted from the benzylic hydroxyl position (the reactive hydroxyl group). This can occur by abstraction of a hydride and a proton, alternatively two hydrogen radicals. The proton is probably accepted by the solvent (2-propanol) and the hydride by the catalyst (palladium). The free hydrogen from the dehydrogenation reaction can consequently be used in the formation of MAT.

To sum up, it can be stated that the theoretical calculations showed that the differences in reaction rate in hydrogenolysis and dehydrogenation were not due to differences in the charges or bond strength. The reason was either the intermediates or the interaction between HMR and the catalyst surface. As protonation of HMR **2**, though, caused a considerably more stable complex than protonation of HMR **1**, this could explain why HMR **2** reacts to MAT with higher rate than HMR **1**.

4.2 LIGNOLs

As mentioned in the materials chapter, the follow-up of the HMR studies was the project "Lignocats", started by Docent Patrik Eklund in 2008, with the aim of synthesizing a new class of 1,4-diols, *i.e.* TADDOL-like conidendrin-based chiral ligands (LIGNOLs), based on natural products with the same catalytic functionality as TADDOLs. The aim of this work was to synthesize conidendrin by protonation (a Friedel-Crafts type ring formation) and methylation of HMR, followed by oxidation and Grignard reaction to receive the new chiral ligand, as shown in Figure 4.9.

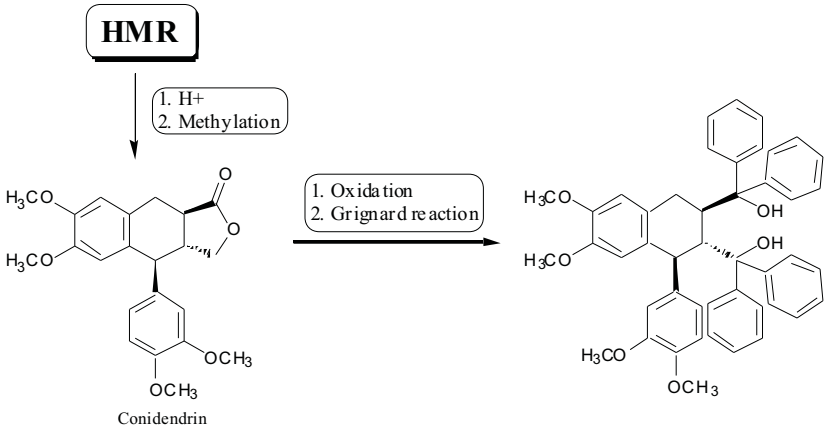
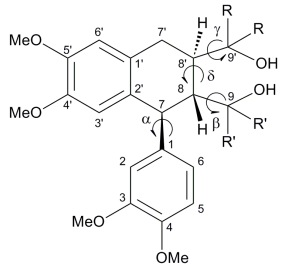


Figure 4.9: Reaction scheme for synthesis of a chiral ligand based on conidendrin with HMR as starting material.

HMR can occur in two diastereomeres (RRR and RRS), but both react to the same diastereomer of conidendrin, then resulting in fixed chirality on atoms 8' (R), 8 (R) and 7 (S) (see Figure 4.10). Figure 4.10 also shows the new way of numbering of atoms in lignans.



Angle	*Atoms*			
α	2	1	7	2'
β	7	8	9	O9
γ	7'	8'	9'	O9'
δ	9'	8'	8	9

Figure 4.10: The four most relevant torsional angles and the numbering of atoms. R, R' = phenyl, methyl or hydrogen.

4.2.1 Conformational analysis

Chiral 1,4-diols such as TADDOLs [97, 98] or BINOLs [99] have been shown to work well in the production of chiral catalysts for several years. They have hindered structures containing two adjacent stereocenters, resulting in a fixed angle between the metal-complexing hydroxyl groups. The objective of the computational chemistry studies in Papers IV and V was also to investigate how the angle between the two hydroxyl groups was behaving in the optimized structures. Crystallographic data on these structures have been published by Weber *et al.* [100]. Relatively few theoretical works in the literature have dealt with these structures. A systematic investigation of the rotational behaviour of aryl substituents in TADDOLs has been conducted by Beck *et al.* [101–103], and the isomerization of BINOL has been theoretically studied in detail by Sahnoun *et al.* [104] at the DFT/B3LYP level of theory.

In Paper IV, quantum chemical calculations was used in order to investigate the minimum energy structures of the following chiral 1,4-diols: 1,1-diphenyl (**2Ph**), two diastereomers of 1,1,4-triphenyl (**3PhR**, **3PhS**), 1,1,4,4-tetraphenyl (**4Ph**) and 1,1,4,4-tetramethyl (**4Met**) 1,4-diol. The minimum energy structure of each of the LIGNOLs are presented in Figures 4.11–4.13. The code for the

conformations is adopted from Paper IV.

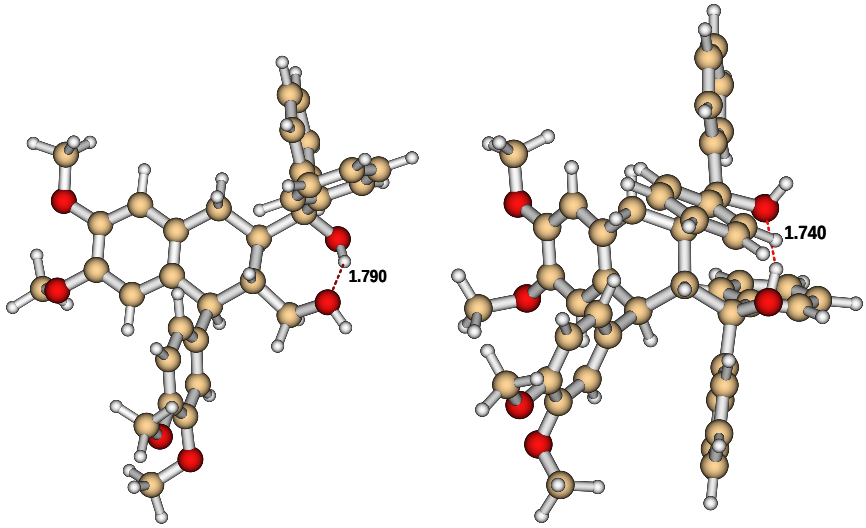


Figure 4.11: The minimum energy structure for the LIGNOLs diphenyl (left) and tetraphenyl (right).

The torsional angles in diphenyl 1,4-diol did not vary much from the molecular mechanically optimized angles. These could be denoted in the same way as the torsional angles for HMR in the study of Taskinen *et al.* [16]. However, the torsional angles for the triphenyl and tetraphenyl 1,4-diols, partly follow the same systematics, but with several exceptions, so the exact values for the angles had to be analysed.

One object of this study was to observe how the angle between the hydroxyl groups in these LIGNOLs, behaves in the optimized structures. The angle δ was analysed and compared to the crystallographic data of TADDOLs by Weber *et al.* [100]. Accordingly, δ should be close to 270°. Another property which made the structures more stable, seemed to be the boat conformation of the aliphatic six-membered ring, which is formed perfectly if δ is 240°. The

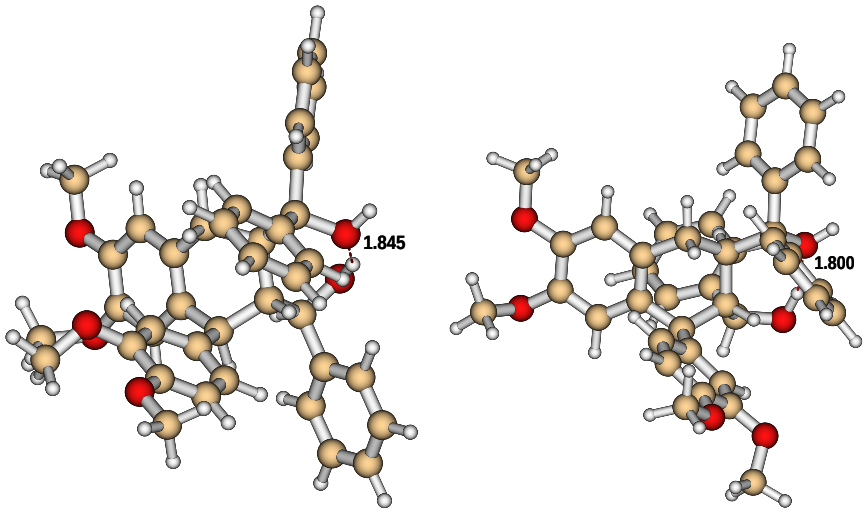


Figure 4.12: The minimum energy structure for the LIGNOLs triphenyl(R) (left) and triphenyl(S) (right).

value of δ should, consequently, be between 240° and 270°. If the hydroxyl groups furthermore point to the same direction, an intramolecular hydrogen bond forms between them, and the bridging hydrogen atom falls at the same place as a chelate-bonded metal ion would be situated. These factors are fulfilled in the triphenyl conformers: **3PhS3**, **3PhR3** and **3PhR7**.

To conclude the quantum chemical conformational analysis of the LIGNOLs, it can be stated that: The diphenyl 1,4-diol is hardly hindered at all resulting in a wide range of almost equally stable conformers. The stability of the other phenylated 1,4-diols is mainly determined by the ability to form $\pi - \pi$ interactions between phenyl rings and the possibility for the aliphatic six-membered ring to be in boat conformation, *i.e.* the torsional angle δ to be between 240° and 270°. The most stable triphenyl 1,4-diols according to the DFT calculations: **3PhS3**, **3PhR3** and **3PhR7** are also those which could work as catalysts. The

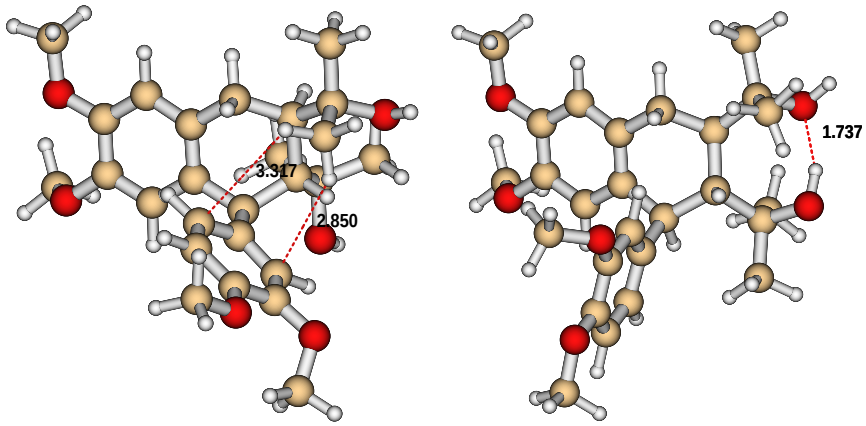


Figure 4.13: The minimum energy structure for the LIGNOLs tetramethyl (left) and tetramethyl which could work as catalyst (right).

most stable tetraphenyl 1,4-diols according to the DFT calculations: **4Ph3**–**4Ph4** and **4Ph7**–**4Ph8** also seems to be possible catalysts. Those conformers of the tetramethyl 1,4-diol, which have the hydroxyl groups pointing to the same direction, are almost 12 kJ/mol less stable than the most favourable one, when the entropy contributions are taken into account. Moreover, they do not have the aliphatic six-membered ring in the prefered boat conformation.

4.2.2 Experimental work

In Paper III, the readily available natural lignan hydroxymatairesinol (*cf.* Chapter 2) was transformed into sterically hindered and optically pure diphenyl, di-2-naphthyl, and tetramethyl 1,4-diols. In addition, the diastereoselective formation of stable hemiketals from the highly substituted butyrolactone was studied in detail. The stereochemical configurations of the hemiketals were determined by NMR and assigned as 8R,8'R,7S,9R for the phenyl substituted hemiketals and 8R,8'R,7S,9S for the methyl substituted, thus the order of priority for de-

termination of the stereochemistry is changed depending on the substituents.

Computational analysis of the methyl substituted hemiketals supported the fact that the thermodynamically favoured isomer should be the *9S* isomer. Gibbs' free energy of the most stable hemiketal isomer was lower by 24–28 kJ/mol compared to the other diastereomer.

The difference in energy, although significant, does not fully explain the unusual stability of these structures. Based on the optimized geometry, the torsion angle C9'–O9'–C9–O9 (Figure 4.14) was determined. The value for the most stable isomer of the methyl substituted *9S* hemiketal was 92°, while for the corresponding diastereomer, it was about 122°. A torsion angle of about 90° indicates a possible anomeric effect, and is due to the delocalization of orbital lone pair at O9' to antibonding orbital C9-O9, which leads to a slight increase in O9'–C9-bond strength (Figure 4.14).

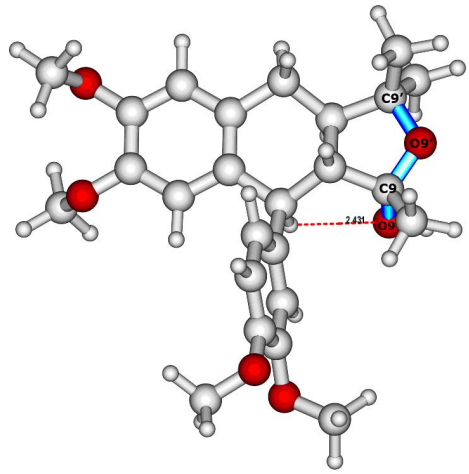


Figure 4.14: The most stable isomer of the methyl substituted *9S* hemiketal.

Mulliken and Löwdin charges, electrostatic potential fit (ESP) and bond orders were also calculated for both isomers of the methyl substituted hemike-

tal. The most stable isomer had less negative charge on O9 and O9', which means that these oxygens are less nucleophilic than O9 and O9' of the other diastereomer. Likewise, the ESP fitting showed that C9 was less positively charged (atomic charge ≈ 0.70), and thus less electrophilic, in the most stable isomer of the methyl substituted *9S* hemiketal compared to the *9R* hemiketal (≈ 0.95). This supports the experimental result that the methyl substituted *9R* hemiketal was not susceptible for nucleophilic attack or in equilibrium with the ketone. However, a complete explanation for the lack of equilibrium with the corresponding ketone as well as the extremely poor reactivity would require additional studies.

Furthermore, the addition reaction to the highly substituted lactones was studied in detail. This showed that substituted butyrolactones form stable hemiketals, which has not been previously discussed in the literature. The steric hindrance and the formation of a stable hemiketal structure, prevented the synthesis of the tetrasubstituted 1,4-diol derivatives, unless a methyl was used as the alkyl group.

4.2.3 MD simulations of LIGNOLs

In the last work included in this thesis, Paper V, molecular dynamics simulations were performed on the LIGNOLs, including torsional angle analysis, solvation effect study quantum chemically by COSMO and hydrogen bond analysis. The three quantum chemically most stable conformers in Paper IV, of each of the LIGNOLs (three per stereoisomer for triphenyl) were chosen for these studies.

Solvation

Solvation effects are important to include in quantum chemical calculations, as solvents play a role in the structures of many molecules. This was conducted for the LIGNOLs using the implicit solvation model COSMO [49, 50]. In the HMR study in Paper I, solvent effects calculated by COSMO were compared

to the corresponding values obtained using the Polarized Continuum Model (PCM) [46–48], and COSMO was found to give more credible relative energies than PCM.

As can be seen in Table 4.5, the energies from the COSMO calculations follow the same order as the conformers in the gas phase for diphenyl and tetraphenyl 1,4-diol. For triphenyl and tetramethyl 1,4-diol the case is different.

Table 4.5: Relative energies in kJ/mol including solvation effects

Conformation	E_{HF}	E_{DFT}	$E_{COS,HF}$	$E_{COS,DFT}$
2Ph1	6.1	7.3	5.7	9.4
2Ph2	5.2	6.7	4.2	8.1
2Ph9	0	0	0	0
3PhR3	21.2	0	0	0
3PhR4	20.5	24.5	28.6	22.5
3PhR5	27.8	36.1	46.2	44.5
3PhS3	35.5	11.6	22.0	12.7
3PhS7	19.6	11.7	25.2	12.6
3PhS10	0	12.1	15.9	17.3
4Met2	8.9	4.5	4.7	7.6
4Met3	0	4.2	0	9.2
4Met6	39.9	0	27.9	0
4Ph3	0	0	0	0
4Ph4	2.0	2.0	1.7	3.9
4Ph8	0.9	1.5	0.9	2.9

When considering the dipole moments in Table 4.6, calculated with GAMESS at HF/6-31G* level, it can be observed that the order between the conformers is preserved almost entirely for each method used. Generally, the dipole moment always increases when a polar molecule is solvated. This is also true for the solvation models, as can be seen in Table 4.6.

When examining the values in Table 4.6 more carefully, it can be seen that the diphenyl 1,4-diols are those which follow the energy trend the closest. The

Table 4.6: Dipole moments in Debye including solvation effects

Conformation	μ_{HF}	μ_{DFT}	$\mu_{COS,HF}$	$\mu_{COS,DFT}$
2Ph1	2.19	2.11	3.88	4.07
2Ph2	1.93	1.56	3.36	3.38
2Ph9	1.79	2.48	1.87	2.54
3PhR3	6.54	4.40	6.08	6.44
3PhR4	3.46	3.92	4.22	5.22
3PhR5	2.80	2.09	3.85	3.77
3PhS3	6.61	6.71	9.29	9.51
3PhS7	6.64	6.67	8.90	9.50
3PhS10	2.22	2.07	3.57	3.72
4Met2	2.03	2.27	2.50	3.42
4Met3	1.90	2.09	3.04	3.55
4Met6	5.33	3.34	7.48	5.48
4Ph3	6.45	6.50	8.40	9.35
4Ph4	3.91	3.85	5.73	6.24
4Ph8	3.48	3.13	4.24	4.12

only exception is μ_{DFT} in gas phase for **2Ph9**. For triphenyl and tetraphenyl 1,4-diol, a trend can be seen that conformers with higher dipole moment are more stable than the other, especially according to the DFT calculations. This also holds for tetramethyl 1,4-diol, for which conformer **4Met6** has by far the highest dipole moment and is most stable in DFT.

Molecular dynamics

The initial values of each of the four torsional angles mentioned in Figure 4.10 are shown in Table 4.7. In Table 4.7, the torsional angles in the DFT optimized structures in Paper IV are shown as well, but the HF structures are those used as starting structures in the MD simulations. The torsional angles which change significantly during the DFT optimizations are marked in bold.

The four most relevant torsional angles in the LIGNOLs were properly anal-

Table 4.7: The initial (HF) and the DFT optimized torsional angles

Conformation	*Method*	α	β	γ	δ
2Ph1	HF	115.9°	49.6°	78.3°	314.5°
	DFT	115.6°	49.5°	78.3°	313.5°
2Ph2	HF	294.9°	48.8°	78.2°	315.0°
	DFT	294.2°	49.7°	78.1°	314.4°
2Ph9	HF	117.1°	198.3°	194.8°	311.8°
	DFT	117.3°	196.7°	195.3°	310.8°
3PhR3	HF	90.0°	190.4°	198.8°	285.0°
	DFT	**348.8°**	**160.3°**	180.2°	**256.8°**
3PhR4	HF	280.5°	190.8°	200.3°	285.8°
	DFT	270.9°	196.3°	198.3°	280.9°
3PhR5	HF	103.5°	206.4°	301.1°	302.4°
	DFT	101.0°	209.8°	303.5°	297.9°
3PhS3	HF	116.7°	224.4°	143.9°	305.1°
	DFT	**170.8°**	216.4°	**185.3°**	**267.6°**
3PhS7	HF	128.6°	208.1°	198.8°	299.6°
	DFT	**171.4°**	215.6°	186.2°	**268.3°**
3PhS10	HF	344.0°	64.8°	293.3°	260.9°
	DFT	344.4°	64.5°	292.2°	260.5°
4Met2	HF	166.0°	177.9°	66.6°	247.1°
	DFT	167.8°	180.6°	65.1°	245.7°
4Met3	HF	163.8°	305.8°	187.7°	248.2°
	DFT	165.6°	307.8°	187.3°	247.0°
4Met6	HF	117.6°	326.2°	196.8°	294.9°
	DFT	**166.3°**	307.1°	187.6°	**247.0°**
4Ph3	HF	87.7°	204.3°	193.2°	289.0°
	DFT	84.0°	208.6°	190.2°	283.3°
4Ph4	HF	280.2°	203.3°	196.5°	289.2°
	DFT	280.4°	205.4°	196.3°	286.0°
4Ph8	HF	275.3°	203.3°	195.1°	290.0°
	DFT	273.8°	205.1°	194.8°	287.1°

ysed during the simulations. Those are explained in detail in Figure 4.10. The Figures 4.15 and 4.16 show the changes in the four torsional angles during two examples of 10 ns simulations of the studied LIGNOLs. The Figures are plotted by *gnuplot*, *version* 4.0, including a smoothing (the thicker line) using the Bezier algorithm.

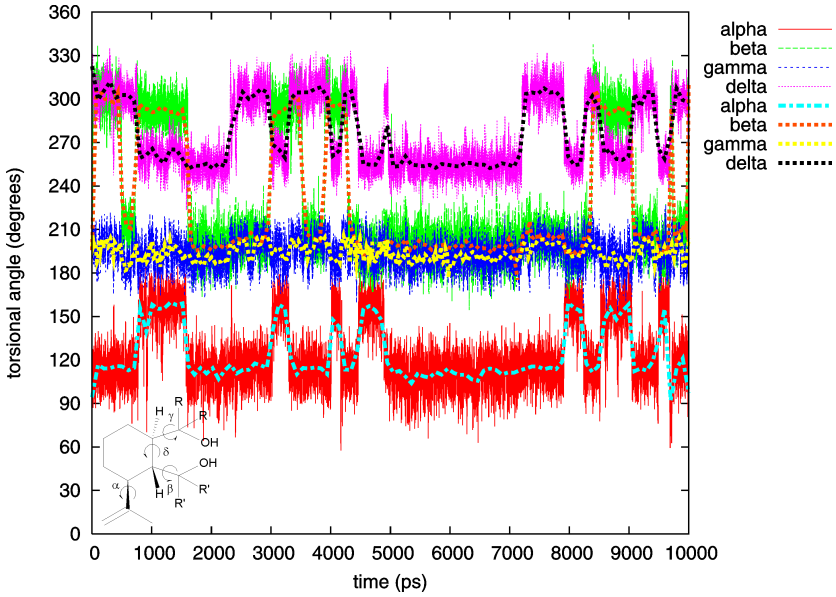


Figure 4.15: Representative torsional angles $\alpha - \delta$ in diphenyl 1,4-diol as a function of time, extracted from 10 ns of NVE simulation at 298 K.

For diphenyl 1,4-diol the highest populated conformations seemed to be **2Ph1** and **2Ph2**. Another highly populated conformation in the simulations in TIP4P water is **2Ph9**, which is the third one picked out as starting geometry for the simulations, but with the exception that the torsional angle δ is rather the favourable 255° than the starting value $\approx$ 315°. However, this change of δ seems to imply that α takes the uncommon value of 150°, also causing or

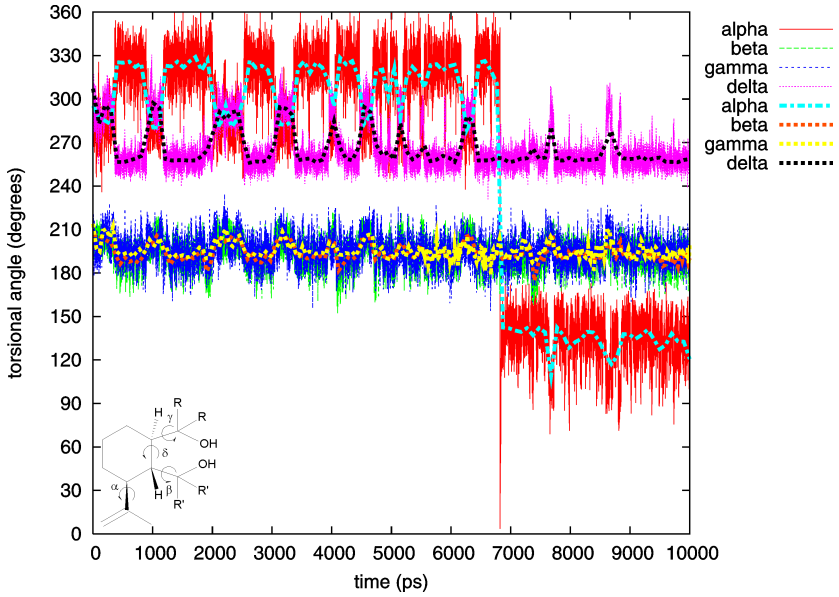


Figure 4.16: Representative torsional angles $\alpha - \delta$ in tetraphenyl 1,4-diol as a function of time, extracted from 10 ns of NVE simulation at 298 K.

implying β to be 300°.

Regarding triphenyl 1,4-diol, especially the R stereoisomer, it is more comprehensible that the value 150°/330° occurs for α as there is an additional phenyl ring substituted on C9. The highest populated conformers are $\approx$ the DFT optimized **3PhR3**, and **3PhS3** and **3PhS7**, respectively, even though the values for the S stereoisomers vary slightly from the optimized ones. The small variation in β and γ from the optimized ones might be explained by the sensitivity of them, when forming hydrogen bonds between the hydroxyl groups, O9–H and O9'–H, and the TIP4P water. The torsional angles in the highest populated conformer in the last triphenyl(S) simulation is close to **3PhS10**.

In tetramethyl 1,4-diol, the three most stable conformers in gas phase;

4Met2, **4Met3** and **4Met6** are dominating, the last two of which are actually the same conformer.

The last but not least molecule studied, tetraphenyl 1,4-diol, is more complicated to summarize in the system of the conformers optimized in gas phase. The most frequent value of δ in the simulations, 255°, only occurs for the conformers **4Ph5** and **4Ph6**, but for those the other torsional angles vary significantly. The only conformers of the simulations, fitting well with the gas phase values, are **4Ph4** and **4Ph8**, actually the same, which occur in the simulation shown in Figure 4.16. The large fluctuations compared to the other LIGNOLs might be a sign of various populations of the molecule.

Opposite to the conclusion for HMR in Paper I, the quantum chemical calculations (in gas phase) seemed to predict reliable trends for how these molecules act in solvents. Perhaps those were more reliable as the molecules were bigger than HMR.

Hydration

In the LIGNOLs, there are four hydrogen bonding acceptor oxygen atoms in methoxy groups. However, the most interesting oxygens from the reaction point of view, are those in the metal-complexing hydroxyl groups, O9–H and O9'–H. In those groups there are also two hydrogen bonding donors.

In order to understand the hydration effect more properly the g_hbond analysing program implemented in GROMACS was used to study the number of hydrogen bonds for the oxygen atoms O9 and O9', and totally for each LIGNOL conformer, as well as the average lifetime of the uninterrupted hydrogen bonds, which are all shown in Table 4.8.

A hydrogen bond can be defined simply as a stabilizing interaction between all donors D and acceptors A, *i.e.* D is an electronegative atom such as nitrogen, oxygen or the halogens, and A usually has a lone pair. In hydrogen bonds, the distance between D and A is less than the sum of their van der Waals radii.

The hydrogen bond can also be further simplified to an interaction between H(ydrogen) and A(cceptor). If a donor group is connected to two acceptors, this is called a bifurcated hydrogen bond.

The g_hbond program computes and analyses hydrogen bonds between all possible donors and acceptors. The existance of an hydrogen bond is determined by a geometrical criterion, *i.e.* $r_{DA} \leq 0.35\ nm$ and $\alpha_{HDA} \leq 30°$ [53].

Table 4.8: The average number of hydrogen bonds per timeframe and the average lifetime (ps) of the uninterrupted hydrogen bonds between the LIGNOLs and the TIP4P solvent

$Conformation$	num_{O9}	$num_{O9'}$	num_{Tot}	$life_{O9}$	$life_{O9'}$	$life_{Tot}$
2Ph1	1.24	0.98	6.23	3.35	**6.18**	1.16
2Ph2	1.23	0.98	6.26	3.26	**5.82**	1.15
2Ph9	1.41	0.80	6.08	2.29	2.24	1.06
3PhR3	1.16	0.71	5.54	2.03	1.67	0.97
3PhR4	1.11	0.78	5.53	1.97	1.72	0.98
3PhR5	1.27	1.15	6.08	2.88	3.15	1.11
3PhS3	1.25	0.65	5.58	2.08	1.58	0.99
3PhS7	1.12	0.80	5.60	1.96	1.78	0.98
3PhS10	1.10	1.14	6.51	3.53	3.35	1.19
4Met2	1.20	**1.33**	6.35	3.95	3.63	1.18
4Met3	1.37	**1.24**	**7.05**	3.57	3.83	1.27
4Met6	1.37	**1.24**	**7.04**	3.56	3.86	1.28
4Ph3	**0.62**	1.05	5.31	1.57	2.07	0.95
4Ph4	**0.68**	0.95	5.23	1.82	2.00	0.96
4Ph8	**0.79**	0.85	5.26	1.88	1.77	0.94

In Paper I, the average number of hydrogen bonds per timeframe for the oxygen atoms in the methoxy groups was ≈ 0.24, which means that the four methoxy oxygens in the LIGNOLs contribute with one hydrogen bond altogether. The rest consists mainly of contributions from the reactive centra, *i.e.* the hydroxyl groups, O9–H and O9'–H.

The first three columns in Table 4.8 show that O9 seems to be slightly more

likely to form hydrogen bonds to the solvent. This conclusion might be erroneous due to the fact that the hydroxyl groups O9–H and O9'–H form an internal hydrogen bond, which takes away one connection to the solvent, for those conformers which have the OH groups pointing in the same direction. However, it can be noted that tetramethyl 1,4-diol is more likely to form hydrogen bonds to TIP4P, and tetraphenyl less, mainly due to the small tendency of O9 to form hydrogen bonds to TIP4P. Figure 4.17 shows hydrogen bonding of tetramethyl 1,4-diol with TIP4P water.

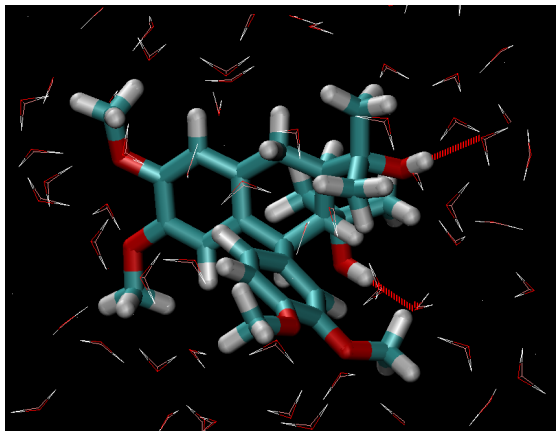


Figure 4.17: Hydrogen bonding of tetramethyl 1,4-diol with TIP4P water.

Considering the lifetimes, in Paper I, it was stated that the average lifetime of the uninterrupted hydrogen bonds was calculated to 1.2 ps. The last column shows that this holds for the LIGNOLs as well. A few are just below 1 ps. When looking at columns 4 and 5 more interesting values can be noted, especially for O9' (col. 5). However, a correlation can be seen to the number of hydrogen bonds, as the lifetimes are longer for tetramethyl 1,4-diol and shorter for tetraphenyl. A shorter lifetime for a large average number of hydrogen bonds may imply that they are slightly weak, meaning that the hydrogen bonds from

O9' in **2Ph1** and **2Ph2** might be strong. This again could be important for the application of these LIGNOLs as metal-binding agents, as their bonding to a metal-atom catalyst is comparable to the hydrogen bonding of the diol to TIP4P water. Diphenyl 1,4-diol is the only LIGNOL in this study with phenyls at C9' and not at C9, thus the reason for this phenomenon is probably the electronic effects of the phenyl rings at C9'.

The hydrogen bond lengths were also analysed using g_hbond, and the mean value of the hydrogen bond lengths is approximately 0.28 nm, exactly as in Paper I.

To sum up, it can be stated that in the MD simulations on the LIGNOLs, the conformations preferred were the energetically most favourable ones according to quantum chemical DFT calculations in gas phase, almost irrespective of the dipole moment. The four most relevant torsional angles, $\alpha - \delta$, defined in Figure 4.10, varied during dynamics in accordance with their symmetry. The torsional angle δ was generally more preferred at the stabilizing value 255° than what was seen in the gas phase optimizations in Paper IV. No strong correlation patterns were found, but in the simulation of **2Ph9**, α and δ changed simultaneously, while β either initialized a conformational change or lagged behind. In the hydration studies, **2Ph1** and **2Ph2** were found to have strong hydrogen bonds from O9', which could be very important for the application of these LIGNOLs as metal-binding agents. The hydration studies of the MD simulations show that several of these LIGNOLs, produced from a renewable source, have a great potential of acting as chiral catalysts.

Chapter 5

Conclusions and Prospects

The work behind this thesis has proved a "long and winding road". It has been uncomfortable (*i.e.* the method of development) and I have lost my way several times during the journey. Still, I have never lost faith in finishing the work. One contributory reason for this is the fact that I have climbed every rung of the ladder completing all the academic degrees, even though it has not been compulsory for me. The risk of dropping down and out is much smaller then.

Research is the kind of work which never finishes. The point of a thesis is to bring it to an end. Several methods have been used in this thesis, at the latter part with the aim of solving the problem in the synthesis of the tetraphenyl substituted 1,4-diol, which has not yet been attained.

$\pi - \pi$ interactions are essential for all the molecules included in this study. The next step in this work is to perform optimizations on the model structures using the DFT-D approaches proposed by Stefan Grimme [105–107], which is taking the dispersion effects much better into account than standard Hartree-Fock or Density Functional Theory used in Papers II, III and IV.

Also the possibility of having the structures stuck in local minima during the MD simulations in Papers I and V, should be investigated further. Sterical hindrance is affecting the rotational barriers a lot already in the starting

torsional analysis at MM level.

Another important issue regarding these structures is a more proper analysis of the solvation effects in the MD simulations. Associate Professor Aurora Clark at the Department of Chemistry at Washington State University in Pullman has developed a new method for analysis of statical mechanical data based on graph theory. The largest global application of graph theory is the development of the PageRank formalism for the organization of web search engine results. Ass. Prof. Clark's group has recently shown the similarities to statistical mechanical simulations, making PageRank a unique new tool for analysis of large chemical systems [108–111]. The future prospects from this point of view are to re-analyse the already produced simulation data on lignans and find an explanation of the problems. A new solution would allow Docent Eklund's group to synthesize the tetraphenyl substituted 1,4-diol structures within a reasonable time-frame.

Bibliography

[1] Allen, M. P.; Tildesley, D. J. *Computer simulations of liquids.* Clarendon Press, Oxford, 1987/2005.

[2] Sandberg, T. *Beräkningskemiska studier av lignaner.* PhL thesis, Åbo Akademi University, 2009.

[3] Ayres, D. C.; Loike, J. D. *Lignans - Chemical, Biological and Clinical Properties.* Cambridge University Press, Cambridge, 1990.

[4] Willför, S.; Hemming, J.; Reunanen, M.; Eckerman, C.; Holmbom, B. *Holzforschung*, 2003, *57*, 27–36.

[5] Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.*, 1953, *21*, 1087–1092.

[6] Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.*, 1957, *27*, 1208–1209.

[7] Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.*, 1959, *31*, 459–466.

[8] Rahman, A. *Phys. Rev.*, 1964, *136A*, 405–411.

[9] Eklund, P. C.; Sjöholm, R. E. *Tetrahedron*, 2003, *59*, 4515–4523.

[10] Markus, H.; Mäki-Arvela, P.; Kumar, N.; Kul'kova, N. V.; Eklund, P.; Sjöholm, R.; Holmbom, B.; Salmi, T.; Murzin, D. Y. *Catal. Lett.*, 2005, *103*, 125–131.

[11] Markus, H.; Mäki-Arvela, P.; Kumar, N.; Heikkilä, T.; Lehto, V.-P.; Sjöholm, R.; Holmbom, B.; Salmi, T.; Murzin, D. Y. *J. Catal.*, 2006, *238*, 301–308.

[12] Freudenberg, K.; Knof, L. *Chem. Ber.*, 1957, *90*, 2857.

[13] Ekman, R. *Holzforschung*, 1976, *30*, 79.

[14] Ekman, R. *Acta Academiae Aboensis, ser. B*, 1979, *39*, 1.

[15] Eklund, P. C. *Synthetic modification of the natural lignan hydroxymatairesinol.* PhD thesis, Åbo Akademi University, 2005.

[16] Taskinen, A.; Eklund, P.; Sjöholm, R.; Hotokka, M. *J. Mol. Struct.*, 2004, *677*, 113–124.

[17] Holmbom, B.; Eckerman, C.; Hemming, J.; Reunanen, M.; Sundberg, K.; Willför, S. PCT Int. Appl., 2002. WO2002098830 A1.

[18] Eckerman, C.; Holmbom, B. Pat appl. PCT/FI101/00691, WO02/09893 A1.

[19] Holmbom, B.; Eckerman, C.; Eklund, P.; Hemming, J.; Nisula, L.; Reunanen, M.; Sjöholm, R.; Sundberg, A.; Willför, S. *Phytochemistry Rev.*, 2003, *2*, 331–340.

[20] Hill, J.-R.; Subramanian, L.; Maiti, A. *Molecular Modeling Techniques in Materials Science*. Taylor & Francis Group, Boca Raton, FL, 2005.

[21] Leach, A. R. *Molecular modelling, principles and applications*. Pearson Education Limited, Singapore, 1996/1999.

[22] Tripos Inc., 1699 South Hanley Road, St. Louis, Missouri, 63144, USA. *SYBYL 7.3*.

[23] Clark, M.; Cramer III, R. D.; van Opdenbosch, N. *J. Comput. Chem.*, 1989, *10*, 982–1012.

[24] Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. J.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.*, 1993, *14*, 1347–1363.

[25] Hehre, W. J.; Radom, L.; v.R. Schleyer, P.; Pople, J. A. *Ab Initio Molecular Orbital Theory*. John Wiley & Sons, New York, 1986.

[26] Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.*, 1972, *56*, 2257–2261.

[27] Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta*, 1973, *28*, 213–222.

[28] Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; DeFrees, D. J.; Pople, J. A.; Gordon, M. S. *J. Chem. Phys.*, 1982, *77*, 3654–3665.

[29] Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.*, 1989, *162*, 165–169.

[30] Häser, M.; Ahlrichs, R. *J. Comput. Chem.*, 1989, *10*, 104–111.

[31] von Arnim, M.; Ahlrichs, R. *J. Comput. Chem.*, 1998, *19*, 1746–1757.

[32] Treutler, O.; Ahlrichs, R. *J. Chem. Phys.*, 1995, *102*, 346–354.

[33] Becke, A. D. *Phys. Rev. A*, 1988, *38*, 3098–3100.

[34] Perdew, J. P. *Phys. Rev. B*, 1986, *33*, 8822–8824.

[35] Lee, C.; Yang, W.; Parr, R. *Phys. Rev. B*, 1988, *37*, 785–789.

[36] Becke, A. D. *J. Chem. Phys.*, 1993, *98*, 5648–5652.

[37] Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.*, 1994, *100*, 5829–5835.

[38] Sandberg, T. Kvantkemiska beräkningar av syra-basegenskaper hos några toluenderivat. Master's thesis, Åbo Akademi University, 2005.

[39] Cramer, C. J. *Essentials of computational chemistry, Theories and Models*. John Wiley & Sons, Chichester, England, 2002.

[40] Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.*, 1984, *80*, 3265–3269.

[41] Boys, S. F. *Proc. Roy. Soc. (London)*, 1950, *A200*, 542–554.

[42] Smith, P. E.; Pettitt, B. M. *J. Phys. Chem.*, 1994, *98*, 9700–9711.

[43] Born, M. *Z. Phys.*, 1920, *1*, 45–48.

[44] Onsager, L. *J. Am. Chem. Soc.*, 1936, *58*, 1486–1493.

[45] Kirkwood, J. G. *J. Chem. Phys.*, 1934, *2*, 351–361.

[46] Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.*, 1981, *55*, 117–129.

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

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

[49] Klamt, A.; Schüürmann, G. *J. Chem. Soc. Perkin Trans.*, 1993, *2*, 799–805.

[50] Klamt, A.; Jonas, V. *J. Chem. Phys.*, 1996, *105*, 9972–9981.

[51] Ahlrichs, R.; Furche, F.; Hättig, C.; Klopper, W.; Sierka, M.; Wiegend, F. *Turbomole Tutorial V5-10.* www.cosmologic.de, 2008.

[52] Frenkel, D.; Smit, B. *Understanding molecular simulation, from algorithms to applications.* Academic Press, San Diego, 1996.

[53] van der Spoel, D.; Lindahl, E.; Hess, B.; van Buuren, A. R.; Apol, E.; Meulenhoff, P. J.; Tielemann, D. P.; Sijbers, A. L. T. M.; Feenstra, K. A.; van Drunen, R.; Berendsen, H. J. C. *Gromacs User Manual version 4.5.3.* www.gromacs.org, 2010.

[54] Bekker, H.; Berendsen, H. J. C.; Dijkstra, E. J.; Achterop, S.; van Drunen, R.; van der Spoel, D.; Sijbers, A.; Keegstra, H.; Reitsma, B.; Renardus, M. K. R. Gromacs: A parallel computer for molecular dynamics simulations. In *Physics Computing 92*, de Groot, R. A.; Nadrchal, J., Ed., Singapore, 1993. World Scientific.

[55] Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.*, 1995, *91*, 43–56.

[56] Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Mod.*, 2001, *7*, 306–317.

[57] van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.*, 2005, *26*, 1701–1718.

[58] Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.*, 2008, *4*, 435–447.

[59] Eriksson, F. *Fysikalisk kemi – formler och tabeller.* Åbo Akademi University, Åbo, 2002.

[60] Born, M.; von Karman, T. *Physik Z.*, 1912, *13*, 297–309.

[61] Ewald, P. *Ann. Phys.*, 1921, *64*, 253–287.

[62] Madelung, E. *Phys. Z.*, 1918, *19*, 524–532.

[63] Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.*, 1993, *98*, 10089–10092.

[64] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.*, 1995, *103*, 8577–8593.

[65] Hockney, R. W.; Eastwood, J. W. *Computer simulation using particles.* McGraw-Hill, New York, 1981.

[66] Luty, B. A.; Tironi, I. G.; van Gunsteren, W. F. *J. Chem. Phys.*, 1995, *103*, 3014–3021.

[67] de Leeuw, S. W.; Perram, J. W.; Smith, E. R. *Proc. R. Soc. Lond.*, 1980, *A373*, 27–56.

[68] de Boor, C. *A practical guide to splines.* Springer-Verlag, 1978.

[69] Brinks, R. *Comp. Appl. Math.*, 2008, *27*, 79–92.

[70] Verlet, L. *Phys. Rev.*, 1967, *159*, 98–103.

[71] Gear, C. W. *Numerical initial value problems in ordinary differential equations.* Prentice-Hall, Englewood Cliffs, NJ, 1971.

[72] Dahlquist, G.; Björk, A. *Numerical methods.* Prentice-Hall, Englewood Cliffs, NJ, 1974.

[73] Hockney, R. W. *Methods Comput. Phys.*, 1970, *9*, 136–211.

[74] Potter, D. *Computational physics.* Wiley, New York, 1972.

[75] Fincham, D.; Heyes, D. M. *CCP5 Quarterly*, 1982, *6*, 4–10.

[76] Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *J. Chem. Phys.*, 1982, *76*, 637–649.

[77] Beeman, D. *J. Comput. Phys.*, 1976, *20*, 130–139.

[78] Gear, C. W. *The numerical integration of ordinary differential equations of various orders.* Report ANL 7126, Argonne National Laboratory, 1966.

[79] van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.*, 1977, *34*, 1311–1327.

[80] Berendsen, H. J. C.; van Gunsteren, W. F. Practical algorithms for dynamic simulations. In *Molecular dynamics simulation of statistical mechanical systems*, Proceedings of the Enrico Fermi Summer School, s. 43–65, Varenna, 1985. Soc. Italiana di Fisica, Bologna.

[81] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.*, 1984, *81*, 3684–3690.

[82] Nosé, S. *Mol. Phys.*, 1984, *52*, 255–268.

[83] Hoover, W. G. *Phys. Rev. A*, 1985, *31*, 1695–1697.

[84] Berendsen, H. J. C. Transport properties computed by linear response through weak coupling to a bath. In *Computer Simulations in Material Science*, Meyer, M.; Pontikis, V., Ed., s. 139–155. Kluwer, 1991.

[85] Parrinello, M.; Rahman, A. *J. Appl. Phys.*, 1981, *52*, 7182–7190.

[86] Nosé, S.; Klein, M. L. *Mol. Phys.*, 1983, *50*, 1055–1076.

[87] Zimmerman, K. *J. Comput. Chem.*, 1991, *12*, 310–319.

[88] Nocedal, J. *Math. Comput.*, 1980, *35*, 773–782.

[89] Byrd, R. H.; Nocedal, J.; Schnabel, R. B. *Math. Prog.*, 1994, *63*, 129–156.

[90] Broyden, C. G. *Inst. Maths. Applics.*, 1970, *6*, 76–90.

[91] Fletcher, R. *Comput. J.*, 1970, *13*, 317–322.

[92] Goldfarb, D. *Math. Comput.*, 1970, *24*, 23–26.

[93] Shanno, D. F. *Math. Comput.*, 1970, *24*, 647–656.

[94] Nocedal, J.; Wright, S. J. *Numerical Optimization.* Springer-Verlag, New York, 1999.

[95] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.*, 1996, *118*, 11225–11236.

[96] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.*, 1983, *79*, 926–935.

[97] Seebach, D.; Weidmann, B.; Widler, L. In *Modern Synthetic Methods*, Scheffold, R., Ed., volume 3, s. 217–353. Salle+Sauerländer, Aarau, 1983.

[98] Seebach, D.; Beck, A. K.; Schiess, M.; Widler, L.; Wonnacott, A. *Pure Appl. Chem.*, 1983, *55*, 1807–1822.

[99] Brussee, J.; Jansen, A. C. A. *Tetrahedron Lett.*, 1983, *24*, 3261–3262.

[100] Weber, E.; Dörpinghaus, N.; Wimmer, C.; Stein, Z.; Krupitsky, H.; Goldberg, I. *J. Org. Chem.*, 1992, *57*, 6825–6833.

[101] Ito, Y. N.; Ariza, X.; Beck, A. K.; Boháč, A.; Ganter, C.; Gawley, R. E.; Kühnle, F. N. M.; Tuleja, J.; Wang, Y. M.; Seebach, D. *Helv. Chim. Acta*, 1994, *77*, 2071–2110.

[102] Beck, A. K.; Dobler, M.; Plattner, D. A. *Helv. Chim. Acta*, 1997, *80*, 2073–2083.

[103] Seebach, D.; Beck, A. K.; Heckel, A. *Angew. Chem. Int. Ed.*, 2001, *40*, 92–138.

[104] Sahnoun, R.; Koseki, S.; Fujimura, Y. *J. Mol. Struct.*, 2005, *735–736*, 315–324.

[105] Grimme, S. *J. Comput. Chem.*, 2004, *25*, 1463–1473.

[106] Grimme, S. *J. Comput. Chem.*, 2006, *27*, 1787–1799.

[107] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.*, 2010, *132*, 154104.

[108] Mooney, B. L.; Corrales, L.; Clark, A. E. *J. Comput. Chem.*, 2012, *33*(8), 853–860.

[109] Mooney, B. L.; Corrales, L. R.; Clark, A. E. *J. Phys. Chem. B*, 2012, *116*(14), 4263–4275.

[110] Hudelson, M.; Mooney, B. L.; Clark, A. E. *J. Math. Chem.*, 2012, *50*, 2342–2350.

[111] Mooney, B. L.; Corrales, L. R.; Clark, A. E. *J. Phys. Chem. B*, Submitted 2012.