FACULTY OF LIFE SCIENCES
UNIVERSITY OF COPENHAGEN

Molecular Structure and Functionality of Health Promoting Molecules in Food

-with emphasis on β-glucan

PhD Thesis

Niels Johan Christensen

Molecular Structure and Functionality of Health Promoting Molecules in Food - with emphasis on β-glucan

PhD thesis by
Niels Johan Christensen, 2009

Supervisors:

Professor Søren Balling Engelsen
Quality and Technology, Department of Food Science
Plant Food Science and Spectroscopy and Chemometrics
and
Associate Professor Lars Hemmingsen
Bioinorganic Chemistry, Department of Basic Sciences and Environment

Faculty of Life Sciences
University of Copenhagen

Title: Molecular Structure and Functionality of Health Promoting Molecules in Food – with emphasis on β-glucan

ISBN: 978-87-7611-327-8
Printed by Samfundslitteratur Grafik, Frederiksberg C, Denmark

Preface

This dissertation is submitted to the Department of Food Science, Faculty of Life Sciences, University of Copenhagen in partial fulfillment of the requirement for the PhD degree in Food Science and Technology. The Danish Food Industry Agency (DFFE) is gratefully acknowledged for sponsoring the "Byg din mad" ("Build your food") project and this PhD project.
I would like to thank my supervisors professor Søren Balling Engelsen and associate professor Lars Hemmingsen for their guidance, help and inspiration.
Likewise, I am indebted to my mentors in carbohydrate molecular dynamics, NMR spectroscopy and SPME-GC/MS, Peter, Flemming and Mikael & Mehdi, respectively.
Thanks to everyone at Quality and Technology for contributing to a great atmosphere. In particular, I would like to thank my office mates Michael and Minah for the time.
I am grateful to all my collaborators through the years and my coauthors on the papers leading up to this thesis; thanks everyone.
I thank Åsmund for the QSAR sparring sessions, and Mette, Susana and Birthe for the β-glucan adventures.
Thanks to Mehdi, Lisbeth, Karina and Abdel for invaluable assistance with the lab work.
Gilda and Kirsten are thanked for assistance with many tasks.
Thanks to Frans for computational support. I thank Lars and Rasmus for chemometric hints and inspiration.
Thanks to Saddik for providing a steady supply of interesting compounds.
Thanks to Bente Steffansen and the lab staff at the Faculty of Pharmaceutical Sciences for introducing me to the GLpKa instrument and for granting me unlimited access to the instrument. Lars Olsen is thanked for introducing me to QSAR at the Faculty of Pharmaceutical Sciences.
Thanks to Peter W for assistance with MacroModel and for providing the Quantum Cube.
Knud and Peter are thanked for their encouragement and patience.
Finally to my parents, Karina and my family and friends: Thank you for all the support and understanding.

Summary

Numerous epidemiological and experimental studies have shown beyond doubt that dietary fibre from vegetables, fruit and cereals protects against a range of diseases. Mixed linkage (1→3),(1→4)-β-D-glucan (referred to as BG in the following) is an example of a dietary fibre from oat and barley. This polysaccharide combines unique texture qualities with a range of health promoting factors such as the ability to lower postprandial blood glucose and insulin levels. Yet the mechanisms responsible for aspects such as immunopotentiation and the hypocholesterolemic effect pertaining to this dietary fiber are unknown. Similarly, a fundamental understanding of the influence of BG on the release of aroma compounds is lacking.

The main purpose of this PhD project was to apply computer based models and methods in the exploration and understanding of measured data from BG model systems. Aspects of particular importance were (1) the understanding of the structural and dynamical properties in solution of the fundamental BG motifs and (2) explanation of measurements of specific molecular interactions between aroma compounds, bile salts and BG. As a side project the combination of spectroscopy and molecular modeling was applied to the study of ginkgolide B, which is an active constituent of popular dietary supplements. Another side project focused on an important technical aspect pertaining to one of the computational methods, QSAR (quantitative structure-activity relationships), employed in the project.

The disaccharides methyl β-cellobioside and methyl β-laminarabioside are the fundamental structural motifs of BG, since they represent respectively the $\beta(1\rightarrow4)$ and the $\beta(1\rightarrow3)$ glycosidic linkage of the polymer. After a thorough molecular mechanics investigation of their conformational preferences, explicit solvent (water) molecular dynamics simulations were carried out for these disaccharides (Paper I) using a second generation carbohydrate force field. The study was supported by NMR (nuclear magnetic resonance) spectroscopy, and the combined computational and experimental approach showed that the conformational preferences of the two sugars in solution are dictated by a combination of intramolecular hydrogen bonds and hydration. The average values of the glycosidic dihedral angles Φ and Ψ calculated from the most relevant molecular dynamics trajectories were employed in the construction of plausible three-dimensional models of BG polymers. These models revealed symmetries that were indicative of polymer packing.

Carbohydrate-aromatic interactions may play a role in e.g. the binding of flavor compounds to BG. The interactions in solution between the smallest BG

motif methyl β-glucopyranoside and the important flavor compound vanillin and a reference molecule (phenol) were probed with 1D ^{1}H NMR experiments (Section 5.4.1). Comparison of the ^{1}H spectra obtained for the sugar with and without aromatic compound indicated deshielding of all sugar protons in the presence of vanillin and shielding in the presence of phenol. A set of very pragmatic quantum chemical calculations were employed in an attempt to explain the differences. Despite their qualitative nature, the results were compatible with a previously postulated stacked complex of the sugar and the aromatic ring in case of phenol. For vanillin, calculations on the stacked model system did not agree with the measured shifts, which may indicate that the stacking interaction does not occur for vanillin.

QSAR (in general QSPR - quantitative structure property relationships) is a well-established method in the pharmaceutical sciences. In Paper II it was employed for the first time in the study of interactions between BG matrices and aroma-related compounds. Robust QSPR models for the flavor release behavior of alcohols and esters from BG matrices were constructed, demonstrating the pertinence and potential of the method in this novel food relevant context.

The application of QSPR for the analysis of equilibrium dialysis of vanillin-related aromatic compounds in BG matrices (Paper III) turned out to be problematic. The fact that no reliable QSPR relationships could be established may be an indication of the problematic nature of the employed compounds, since detection problems for vanillin have been noted in other contexts. Nonetheless, the example may serve as inspiration for the investigation of the dialysis behavior of compounds with less structural resemblance to vanillin.

Investigations of a technical aspect of QSAR were undertaken in a study designed to reveal the influence of geometry optimization of molecular sets on the final QSAR/QSPR model (Paper IV). The results from the analysis of three diverse datasets showed that the QSAR models were not particularly sensitive to the method used for geometry optimization, even when based on molecular descriptors strongly influenced by the molecular geometry. This suggests that the computational time requirements in connection with certain types of QSAR can be reduced by the application of pragmatic geometry optimization methods.

Combined vibrational circular dichroism spectroscopy (VCD) and density functional theory (DFT) studies of the natural product ginkgolide B, which is an active constituent of widely used dietary supplements, allowed for the assignment of the absolute configuration of the compound (Paper V). This reinforces the view of the VCD/DFT combination as a powerful method for the determina-

tion of the absolute configuration of natural products with several chiral centers.

Together, the projects in this thesis emphasize the significant potential that methods from molecular modeling and chemometrics offer in the interpretation of advanced measurements of structure and functionality in food relevant model systems. It is expected that the employed computational methods will assume increasingly important roles in meeting the future demands for characterization of foods at the molecular level.

Sammendrag

Talrige epidemiologiske og eksperimentelle studier har utvetydigt demonstreret at kostfibre fra grøntsager, frugter og cerealier beskytter mod en række sygdomme. β-glukaner med vekslende $\beta(1\rightarrow3)$ og $\beta(1\rightarrow4)$ glykosidiske bindinger (omtales i det følgende som BG) er et eksempel på kostfibre fra havre og byg, som kombinerer unikke teksturmæssige kvaliteter med helsefremmende egenskaber som f.eks. dæmpning af postprandialt udsving i blodsukker og insulinniveau. Således har BG et stort potentiale for anvendelse i fremtidens fødevarer. Imidlertid er endnu uforklarede mekanismer ansvarlige for eksempelvis immunpotentierende og kolesterolsænkende virkninger ved indtaget af denne kostfiber. Ligeledes mangler en grundlæggende indsigt i kostfiberens betydning for frigivelsen af aromastoffer.

Hovedformålet med dette PhD projekt var at anvende computerbaserede modeller og metoder i analysen og forståelsen af eksperimentelle data fra BG modelsystemer. Vigtige emner som søgtes belyst var (1) en gennemgribende forståelse af de strukturelle og dynamiske egenskaber for BG polymerens mindste enheder og (2) forklaring af målte molekylspecifikke interaktioner mellem aromastoffer, galdesalte og BG. Som et sideprojekt, blev kombinationen af spektroskopi og molekylmodellering anvendt på naturstoffet ginkgolid B der er en aktiv bestanddel af et populært kosttilskud. Et andet sideprojekt fokuserede på et vigtigt teknisk aspekt ved en af de computerbaserede metoder, QSAR (kvantitative struktur aktivitet relationer), som blev anvendt i projektet.

Disakkariderne methyl β-cellobiosid og methyl β-laminarabiosid er de fundamentale strukturelle enheder i BG polymerer. De repræsenterer henholdsvis $\beta(1\rightarrow3)$ og $\beta(1\rightarrow4)$ bindingen i BG polymeren. Efter en gennemgribende molekylmekanisk undersøgelse af deres konformationelle præferencer blev disse molekylers dynamiske opførsel studeret med molekylær dynamik simulationer i eksplicit solvent (vand) under anvendelse af at andengenerations kulhydrat-kraftfelt (Artikel I). Studiet blev understøttet af NMR (kernemagnetisk resonans) målinger på stofferne, og de kombinerede eksperimentelle og beregnede resultater viste at de to sukres foretrukne konformationer i opløsning bliver dikteret af et samspil af intramolekylære hydrogenbindinger og hydrering. Gennemsnitsværdierne for de glykosidiske dihedrale vinkler Φ og Ψ blev beregnet for de mest relevante molekylær dynamik trajektorier og anvendt til konstruktionen af plausible tredimensionelle modeller af BG polymerer. Disse modeller viste symmetrier som antydede muligheden for pakning af polymererne.

Kulhydrat-aromat interaktioner kan have betydning for eksempelvis binding af aromastoffer til BG. Interaktionerne i opløsning mellem methyl β-glukopyranosid, som er den mindste sukkerenhed i BG, det vigtige aromastof vanillin og ref-

erencestoffet phenol blev undersøgt med en-dimensionelle NMR eksperimenter (Sektion 5.4.1). Sammenligning af målte ^{1}H NMR spektre viste at alle kulhydratprotoner i blandinger af methyl β-glukopyranosid og aromat blev afskærmet mere af phenol og mindre af vanillin i forhold til den rene opløsning af methyl β-glukopyranosid. I et forsøg på at forklare disse observationer blev en række pragmatiske kvantekemiske beregninger udført på modeller af de formodede aromat-kulhydrat komplekser. På trods af deres kvalitative natur, understøttede resultaterne den hidtidige antagelse om et plant kompleks med interaktion mellem phenol og pyranoseringens α-side. I modsætning hertil understøttede sammenligningen mellem beregnede og målte ændringer i kemisk skift ikke et plant kompleks for methyl β-glukopyranosid og vanillin.

QSAR (generelt QSPR - kvantitativ struktur egenskabs relation) er en veletableret metode i strukturbaseret lægemiddelforskning. I Artikel II blev QSPR anvendt for første gang i studiet af interaktioner mellem BG matricer og aroma-relaterede stoffer. Der blev etableret robuste QSPR modeller for aroma frigivelse af alkoholer og estere fra BG matricer, hvilket demonstrerede brugbarheden af metoden i denne nye fødevarerelaterede sammenhæng.

Anvendelsen af QSPR til analyse af ligevægtsdialyse af vanillin-relaterede aromatiske stoffer i BG matricer (Artikel III) viste sig at være problematisk. Det faktum at en brugbar QSPR model ikke kunne etableres, afspejler muligvis den problematiske natur som vanillin har udvist i andre eksperimentelle sammenhænge. Eksemplet kan tjene som inspiration til afprøvning af forsøgsopsætninger hvor stoffer med fjernt slægtskab til vanillin dialyseres.

Undersøgelsen af et teknisk aspekt af QSAR blev foretaget i et studie designet til at afsløre indflydelsen af valg af geometrioptimeringsmetode på den endelige QSAR model (Artikel IV). Resultaterne fra analyser af tre forskelligartede datasæt viste at QSAR modellerne ikke var udpræget følsomme over for valget af geometrioptimeringsmetode, selv når molekylære deskriptorer med en stærk geometriafhængighed blev anvendt. Dette antyder at beregningsstiden i forbindelse med udviklingen af visse typer QSAR modeller kan nedsættes ved anvendelse af mere pragmatiske geometrioptimeringsmetoder.

Udtræk fra tempeltræ (*Ginkgo biloba*) er blandt de mest anvendte kosttilskud i verden. Stoffet ginkgolid B udgør en naturlig bestanddel af sådanne produkter og er en potent antagonist for pladeaktiverende faktor (PAF). Kombination af vibrationel cirkulær dikroisme spektroskopi (VCD) og tæthedsfunktionalteori (DFT) blev anvendt i et studie af ginkgolid B (Artikel V) som klarlagde stoffets absolutte stereokemiske konfiguration og derved understregede brugbarheden af

den kombinerede VCD og DFT tilgang til opklaring af absolut konfiguration i naturprodukter med adskillige kirale centre.

Tilsammen understreger projekterne i denne afhandling det betydelige potentiale som metoder fra molekylmodellering og kemometri har i forbindelse med tolkning af resultater fra avancerede målinger af struktur og funktionalitet i fødevarerelaterede systemer. Det forventes at de anvendte beregningsmæssige metoder i stigende grad vil indtage en fremtrædende rolle i fødevareforskningen, hvor de i samspil med især spektrometriske målemetoder vil bidrage til udviklingen af fremtidens fødevarer.

List of publications

Paper I

Niels Johan Christensen, Peter Ibsen Hansen, Flemming Hofmann Larsen, Tore Folkerman, Mohammed Saddik Motawia, and Søren Balling Engelsen. *A combined nuclear magnetic resonance and molecular dynamics study of the two structural motifs for mixed linkage β-glucans: Methyl β-cellobioside and methyl β-laminarabioside.* Carbohydrate Research. **Submitted.**

Paper II

Niels Johan Christensen, Susana Murtinheira da Trindade Leitão, Mikael Agerlin Petersen, Birthe Møller Jespersen, and Søren Balling Engelsen. *A quantitative structure-property relationship study of the release of some esters and alcohols from barley and oat β-glucan matrices.* Journal of Agricultural and Food Chemistry, 2009, 57(11), 4924-4930. **Published.**

Paper III

Henrik Toft Simonsen, Mette Skau Nielsen, Niels Johan Christensen, Ulla Christensen, Thomas V. La Cour, Mohammed Saddik Motawia, Birthe Pontoppidan Jespersen, Søren Balling Engelsen, and Birger Lindberg Møller. *Molecular interactions between barley and oat beta-glucans and phenolic derivatives.* Journal of Agricultural and Food Chemistry, 2009, 11, 57(5), 2056-2064. **Published.**

Paper IV

Åsmund Rinnan, Niels Johan Christensen, and Søren Balling Engelsen. *How the energy evaluation method used in the geometry optimization step affects the quality of the subsequent QSAR/QSPR models.* Journal of Computer-Aided Molecular Design. **Accepted.**

Paper V

Niels H. Andersen, Niels Johan Christensen, Peter Rygaard Lassen, Teresa B. Freedman, Laurence A. Nafie, Kristian Strømgaard, and Lars Hemmingsen. *Structure and absolute configuration of ginkgolide B characterized by IR- and VCD spectroscopy.* Chirality, 2009. **Published (web).**

Abbreviations

A_{rel}	Flavor release from a BG matrix relative to the flavor release from water
B3LYP	The hybrid exchange-correlation functional due to Becke, Lee, Yang, and Parr
BG	Mixed-linkage(1→3),(1→4)-β-D-glucan
CSFF	Carbohydrate Solution Force Field
DFT	Density Functional Theory
fs	femtosecond (10^{-15} s)
GB	Ginkgolide B
GC/MS	Gas Chromatography/Mass Spectroscopy
GTO	Gaussian Type Orbital
HF	Hartree-Fock
MD	Molecular Dynamics
MM	Molecular Mechanics
MP2	Second order Møller-Plesset perturbation theory
MW	Molecular Weight
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
ns	nanosecond (10^{-9} s)
o.d.	Outer diameter of NMR tubes
PAF	Platelet Activating Factor
PC	Principal Component
PCA	Principal Components Analysis
PCR	Principal Components Regression
PES	Potential Energy Surface

PLS	Partial Least Squares regression
QSAR	Quantitative Structure Activity Relationship
QSPR	Quantitative Structure Property Relationship
RMSECV	Root Mean Square Error of Cross Validation
SPME	Solid Phase Micro-Extraction
STO	Slater Type Orbital
VCD	Vibrational Circular Dichroism

Contents

1 Introduction **1**

2 Molecular modeling **5**
2.1 Molecular mechanics 6
2.2 Electronic structure methods 13
2.3 Molecular descriptors 21

3 Chemometrics **25**
3.1 Linear regression 25
3.2 Principal Components Analysis 26
3.3 Partial Least Squares regression 27
3.4 Variable selection 28
3.5 Model validation 28
3.6 Quantitative Structure Activity Relationships 29

4 Experimental methods **31**
4.1 NMR spectroscopy 31
4.2 IR spectroscopy 34
4.3 VCD spectroscopy 37
4.4 GC/MS 38
4.5 Equilibrium dialysis 40

5 Applications **43**
5.1 Molecular dynamics studies of β-glucan motifs 43
5.2 QSPR models for flavor release 55
5.3 Dialysis 62
5.4 Carbohydrate-aromatic interactions 68
5.5 The influence of geometry optimization on QSAR 77
5.6 VCD and DFT studies of ginkgolide B 81

6 Thesis outlook **87**
6.1 BG building blocks 87
6.2 Carbohydrate-aromatic interactions 87
6.3 QSPR/QSAR 88
6.4 VCD and DFT studies of ginkgolide B 90

7 Conclusions **91**

Bibliography **95**

Benefits should be conferred gradually; and in that way they will taste better.

Niccolo Machiavelli

1 Introduction

In spite of the large potential for planning and rational thinking endowed by a particularly well-developed prefrontal cortex [1], primitive urges triggered by smell and taste play impressively determining roles in man's choice of food [2]. Thus paradoxically, in societies where food supply, safety, nutrition and nutritional advice are more than adequately provided for, hedonistic impulses turn consumers into patients. A well-known example is the obesity crisis now affecting most of the developed world [3, 4]. This crisis is still unfolding in Europe and Northern America, despite government initiatives like the American "five-a-day" campaign [5] aimed at increasing the intake of fruits and vegetables in the general population. How can such a situation be reversed if the appeal to consumer rationality is a battle lost in advance? Food science may hope to contribute to the resolution of the obesity crisis in several ways. For instance, understanding the flavor release profiles of health promoting foods or food ingredients will undoubtedly play a major role in improving their acceptance by consumers. Consider for example the substance shown in **Figure 1.1**. It is an aqueous solution of mixed linkage (1→3),(1→4)-β-D-glucan (BG), a dietary

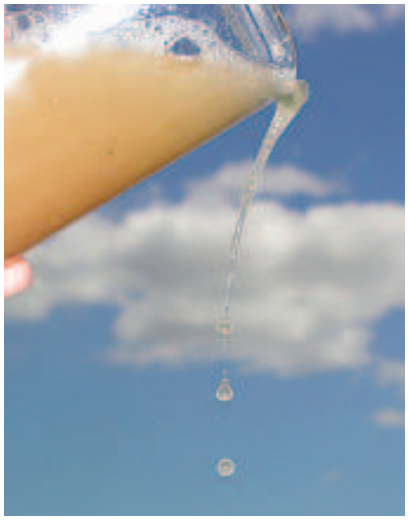


Figure 1.1: Demonstration of the viscosity and visual appeal of the solution of a commercial β-glucan product. Photo by Karina Topolianaki and Birthe Møller Jespersen.

fibre extracted from oat and barley. Its consumption offers many health benefits [6], some of which may reverse or at least soften the blows from the chronic abuse of hamburgers, salted meats and other staples [7, 8] of the Western diet. Nevertheless, the limited aesthetic appeal of the viscous, semitranslucent solution of this refined product would undoubtedly repel most consumers. More promising is the use of BG as a *component* of food; In Norway, for example, it has been included as a fat replacer in low-fat mayonnaise, where it controls viscosity and imparts a desirable mouthfeel to the product [9]. A recent study also reported excitement in a sensory panel following ingestion of sugar cookies and peanut spreads enriched with BG [10]. However the use of BG is not limited to mayonnaise and cookies; one may imagine a wide range of existing foods that could be improved both in terms of texture and healthfulness by the addition of BG. It is also possible that completely novel foods based primarily on BG may arise. In all cases, it is highly relevant to understand how the presence of BG influences the release (or retention) of flavor compounds and thereby modulates the crucial hedonic dimension of food [11]. Release/retention mechanisms may also be involved in a direct way with health, since noxious substances in the gut might be absorbed or engulfed by the BG fraction of the ingested meal [12]. The binding effect, whether affecting flavor molecules in foods or undesirable compounds in the gut, is poorly understood but can in both cases be viewed as a health promoting (or at least health *relevant*) mechanism. Therefore part of the work in this PhD project tried to provide a basis for the understanding of the affinity of various small molecules for BG. Two classes of approach, top-down and bottom-up, were employed. To the former class belong experiments with headspace analysis (Section 5.2) and equilibrium dialysis (Section 5.3) involving the whole BG polymer. In these cases, the formulation of hypotheses pertaining to the involved mechanisms was attempted based on the computational analysis of measurements of small molecule affinity. On the other hand, bottom-up approaches focus on the structure and dynamics of building blocks of the BG polymer using spectroscopy and molecular dynamics. The results from such procedures can be very detailed, as demonstrated for e.g. the BG building blocks in Section 5.1, and hold promise for establishing a link between basic molecular structures and dynamics and macroscopic quantities [13].

While dietary fibres like BG could constitute the artillery in the battle against the spiral of hedonistic self-destruction, smaller molecules may assume the role of the infantry. An example studied in this project is ginkgolide B (GB), a molecule found in popular dietary supplements from the ancient *Ginkgo biloba* tree. GB is a potent inhibitor of the platelet-activating factor (PAF). The latter molecule is involved in exacerbating the effects of ischemic cerebral stroke [14] which is one particularly debilitating effect connected with the Western diet [15].

The contents of this thesis can be divided into three parts.

The first part serves as an introduction to the methods used, and is comprised by **Chapter 2**, **Chapter 3** and **Chapter 4** covering molecular modeling, chemometrics, and experimental methods, respectively.

The second part, applications, is contained in **Chapter 5** and covers the 5 manuscripts that have been submitted for publication in peer-reviewed journals. Additional unpublished work is also discussed in that chapter, in particular the combined NMR and molecular modeling study of interactions between methyl β-glucopyranoside and vanillin.

Finally, **Chapter 6** contains some thoughts on future extensions of the presented work and **Chapter 7** summarizes the main conclusions obtained.

2 Molecular modeling

Molecular modeling in the broadest sense refers to any computational approach employed to gain insights into molecular systems. Inspired by Labanowski [16] it may be instructive to employ a rough subdivision into the following three groups:

1. **Comparative approaches** focusing on initial generation of molecular models, visualization, shape analysis, superposition, topological descriptors, molecular fingerprints, traditional QSAR, rigid conformational search, etc. No quantitative consideration of energy.
2. **Empirical approaches** such as molecular dynamics and molecular mechanics, based on simple parametrizations of the energy in inter- and intramolecular interactions. Allow for the simulation of very large molecules, and can provide useful insights into structure, conformation, hydration, relative energies and other properties. Computational time scales roughly with N (with electrostatic cutoff) or N^2 (full electrostatics), where N is a measure of the system size. However, the lack of explicit electron treatment implies that these methods are very sensitive to parametrization and precludes for example the modeling of chemical reactions.
3. **Quantum approaches** based on explicit consideration of the electronic structure. These methods allow for highly accurate determinations of electron densities, energies, and geometries and allow for the modeling of chemical reactions, accurate calculation of spectra, and virtually any other imaginable property. Computational time scales roughly as N^3 (typical DFT calculation) to N^7 (QCISD(T)), where N is the number of orbitals [17].

Methods from group (1) and (2) are well-suited for many molecular modeling tasks in food science, since they allow the treatment of molecular systems of very large dimensions. The most generally relevant methodologies are probably molecular visualization, molecular mechanics and molecular dynamics. Previously molecular mechanics and dynamics have been used in the characterization

of explicitly hydrated starch motifs [18], for example, and were employed in this thesis for the explicit solvation studies of BG motifs (Section 5.1.1). Due to their often massive demand for CPU power, methods from group (3) are not suitable for application to typically sized food macromolecules. Quantum methods can however provide useful insights into small building blocks of food polymers. For instance, French et al. [19] calculated adiabatic maps for the cellulose and BG structural motif cellobiose with a quantum mechanical approach. The quantum approaches can also be used in the study of small-molecule food constituents when highly accurate results are needed, such as in the combined DFT/VCD study of Ginkgolide B in Section 5.6.

The following provides a brief introduction to the molecular modeling methods used in the project.

2.1 Molecular mechanics

2.1.1 The force field

Calculating the electronic energy a central theme in computational chemistry, and it is in general complex and computer intensive task. Molecular mechanics (MM) methods bypass the direct problem, by writing the electronic energy as a parametric function, the force field, of the nuclear coordinates, where the parameters are obtained from experimental data or from higher level calculations. A force field (FF) consists of contributions from bonds, angles dihedral angles, improper angles and non-bonded interactions. For instance the general FF expression used in the CHARMM program [20] is:

$$\begin{aligned} U(\boldsymbol{R}) &= \sum_{bonds} K(b-b_0)^2 + \sum_{angles} K_\theta(\theta-\theta_0)^2 \\ &+ \sum_{dihedrals} K_\phi[1+\cos(n\phi-\delta)] + \sum_{impropers} K_\omega(\omega-\omega_0)^2 \\ &+ \sum_{Urey-Bradley} K_{UB}(S-S_0)^2 + \sum_{nonbond} \frac{q_i q_j}{\epsilon_1 r_{ij}} \\ &+ \sum_{nonbond} \epsilon \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - \left(\frac{R_{min,ij}}{r_{ij}} \right)^{6} \right], \end{aligned} \tag{2.1}$$

where $\boldsymbol{R}$ is the vector of atomic coordinates for the system and the K's are force constants for the various type of molecular deformation (bond stretching, bending, and so on). Furthermore the FF expression is seen to include terms for well-known phenomena such as the van der Waals interaction, and electrostatics.

The parameters, K, and the particular number of terms included, define the force field. Because of the simple energy evaluation, MM methods are much faster than quantum mechanical methods, and can therefore be applied to huge systems such as the complete tobacco mosaic virus [21] or a cube of 1 trillion atoms [22]. The methodology is thus well suited for studies in the life sciences, not least food science, where molecules often are numerous, large, and flexible. The speed, however, does come at a price; as a consequence of the intrinsic lack of direct electron treatment a FF parameterized for one group of compounds rarely performs well for other compounds. This explains why a myriad of FFs are available.

Carbohydrate force fields

Given their flexible nature, it is not surprising that a multitude of force fields parameterizations are available for carbohydrates. GLYCAM [23], 45A4 [24], and OPLS [25] are some well-known examples. Recent additions include an improved hexopyranose force field [26] and a force field for acyclic carbohydrates [27]. In Paper I of this project, we employed the Carbohydrate Solution Force Field (CSFF) [28]. This force field works with the CHARMM [29, 30] program and is a development of the previous PHLB [31] and HGFB [32] carbohydrate force fields. Although attention has been paid to improving hydroxymethyl rotations in CSFF, our study did not focus in particular on this aspect. Our choice of CSFF was rather motivated by its generally good performance in earlier hydration studies of methyl α-D-maltoside and methyl α-D-isomaltoside [33] and amylopectin trisaccharide building blocks [18].

Water force field

Due to the importance of water in determining the solution structure of carbohydrates, a good representation of water is required in simulations. Because CSFF was parameterized using the TIP3P [34] water model implemented in CHARMM [20], this choice was also employed in the simulations in Paper I. In the CHARMM version of TIP3P, each of the three atoms in the water molecule is represented by a point charge and a Lennard-Jones potential energy term. The SHAKE [35] algorithm ensures that the geometry of the water molecule does not deviate from the known structure of water throughout the simulation.

2.1.2 Geometry optimization

The starting point for any kind of molecular modeling is the creation of a set of atomic coordinates representing a molecule of interest. Simply drawing the molecule in a suitable molecular editing program (Arguslab [36], for example) is

a common way of generating an initial molecular model. The direct use of such a coarsely estimated starting geometry may result in problems in e.g. molecular dynamics simulations, which may fail due to unrealistically large forces on atoms. Therefore some form of geometry optimization [37, 38] is typically performed on the initial structure to find the arrangement of the nuclear coordinates that minimizes the potential energy of a molecule. More formally, one wants to find a molecular configuration satisfying $dE(\boldsymbol{R})/d\boldsymbol{R} = 0$, where $\boldsymbol{R}$ is the vector of nuclear coordinates. The energy function of nuclear coordinates, $E(\boldsymbol{R})$, may originate from a molecular mechanical or quantum mechanical method, and the search for a minimum can be performed with a variety of minimization algorithms, where popular implementations are of the steepest descent or conjugate gradient type. Clearly, the quality of the structure resulting from geometry optimization is only as good as the potential energy function employed allows. The influence of the choice of geometry optimization in the context of QSAR/QSPR was investigated in Paper IV.

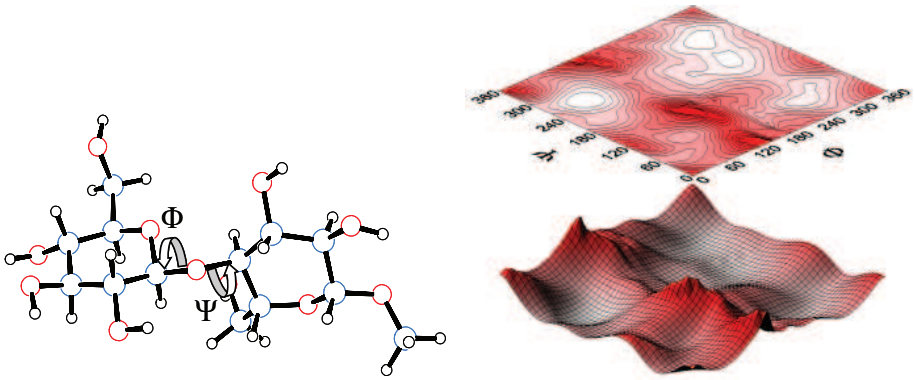
2.1.3 Conformational search

The potential energy surface (PES) of all but the simplest molecules have multiple minima and geometry optimization will always produce the structure corresponding to the minimum closest to the (often arbitrarily chosen) initial structure. This may not be desired; one often wishes to find the *global minimum* which is to say the lowest minimum of the PES. A conformational search tries to find multiple minima on the PES by performing geometry optimizations from many different initial geometries. This is problematic for most systems, because the many molecular degrees of freedom offer a staggering number of possible starting geometries. Therefore, rather than exhaustive sampling, an energy guided stochastic search of the PES is employed in methods such as the Monte Carlo Multiple Minimum (MCMM) method [39, 40], which was used for all conformational searches in this project.

2.1.4 Adiabatic map

The adiabatic map provides a convenient graphical representation of the conformational preferences of a saccharide of interest. An example is shown in **Figure 2.1(b)** for methyl β-cellobioside. The map is established by selecting a starting conformation for the exocyclic groups and systematically varying the dihedral angles Φ and Ψ (**Figure 2.1(a)**) for this conformation. At each fixed (Φ, Ψ)-pair the geometry of the sugar residues is relaxed and the energy is recorded for the optimized structure. The final adiabatic map arises by combining a number of sub-maps, each corresponding to a particular configuration of the exocyclic groups. In practice it is only possible to consider a limited subset of all possible

combinations of exocyclic conformations. In accordance with previous studies [41], the adiabatic maps in Paper I were produced by combining the lowest energy points from the 36 individual maps corresponding to all combinations of the following exocyclic arrangements: Clockwise and anticlockwise orientation of all OH groups (2×2 possibilities), and *gg*, *gt*, and *tg* rotamers for the hydroxymethyl groups (3×3 possibilities). The values of Φ and Ψ were varied from 0° to 360° in 10° increments.



(a) (b)

Figure 2.1: (a) Definition of the glycosidic dihedral angles Φ and Ψ, illustrated for methyl β-cellobioside. (b) The adiabatic map for methyl β-cellobioside (top) is a contour plot of a potential energy surface in (Φ, Ψ)-space (bottom) where all other degrees of freedom are relaxed, see text. Areas of high and low energy are colored red and white, respectively.

2.1.5 Molecular Dynamics

The time evolution of a system of N atoms can be simulated by solving Newton's second law for each atom i:

$$m_i \frac{d^2 \boldsymbol{r}_i}{dt^2} = \boldsymbol{f}_i, \tag{2.2}$$

where m_i and $\boldsymbol{r}_i$ are the mass and coordinate of atom i and $\boldsymbol{f}_i$ is the total force on this atom due to interactions with the remaining $(N-1)$ atoms. The force is calculated as the negative gradient of the potential energy function defined by a force field:

$$\boldsymbol{f}_i = -\boldsymbol{\nabla}_{\boldsymbol{r}_i} U(\boldsymbol{r}) \tag{2.3}$$

Since there are three force components for each of the N atoms, $3N$ second-order differential equations must be solved in order to predict the time evolution of the system.

The velocity Verlet algorithm

Given the dynamical variables for all particles at time t the problem is to predict these variables after a small but finite timestep δt. In this project the Velocity Verlet algorithm [42] was employed. In this scheme the position and velocity of particle i are determined by:

$$\boldsymbol{r}_i(t+\delta t) = \boldsymbol{r}_i(t) + \boldsymbol{v}_i(t)\delta t + \frac{1}{2}\boldsymbol{a}_i(t)\delta t^2 \tag{2.4}$$

$$\boldsymbol{v}_i(t+\delta t) = \boldsymbol{v}_i(t) + \frac{1}{2}[\boldsymbol{a}_i(t) + \boldsymbol{a}_i(t+\delta t)]\delta t \tag{2.5}$$

Evaluation of (2.5) is broken into two step when it is used in MD calculations. The velocity at time $t + \frac{1}{2}\delta t$ is evaluated first with the forces corresponding to time t

$$\boldsymbol{v}_i(t+\frac{1}{2}\delta t) = \boldsymbol{v}_i(t) + \frac{1}{2}\boldsymbol{a}_i(t) \tag{2.6}$$

Next the forces are updated corresponding to positions $\boldsymbol{r}_i(t+\delta t)$ and the velocity at time $t + \delta t$ is evaluated:

$$\boldsymbol{v}_i(t+\delta t) = \boldsymbol{v}_i(t+\frac{1}{2}\delta t) + \frac{1}{2}\boldsymbol{a}_i(t+\delta t) \tag{2.7}$$

A molecular dynamics simulation using the velocity Verlet algorithm consists of the following steps:

1. $\boldsymbol{r}_i$ and $\boldsymbol{v}_i$ are assigned initial values
2. Energies and forces are calculated from the potential
3. New positions $\boldsymbol{r}_i + \delta t$ are calculated from (2.4)
4. Velocities are advanced according to (2.6)
5. Potential energy and forces are calculated at time $t + \delta t$
6. The velocity at $t + \delta t$ is calculated from (2.7)
7. Kinetic and total energies are calculated
8. The procedure is repeated from step 2

The time step is always chosen as large as possible since this reduces the amount of computing time required. The fastest vibrational motion of molecular systems corresponds to periods of approximately 10 fs. To represent such vibrations realistically, the maximum time step must be significantly smaller. In practice, it is often chosen to be of the order of one fs.

Periodic boundary conditions

The finite size of a system studied by molecular dynamics simulations, implies that surface effects cannot be neglected. Unless these effects are the subject of interest itself, they represent artifacts. Imposing periodic boundary conditions (PBC) on the system being studied effectively eliminates surface effects. This is achieved by placing the system at the center of a simulation box surrounded by all sides by identical copies of itself, see **Figure 2.2** for a two-dimensional example. When a molecule moves in the central box, its periodic image in each

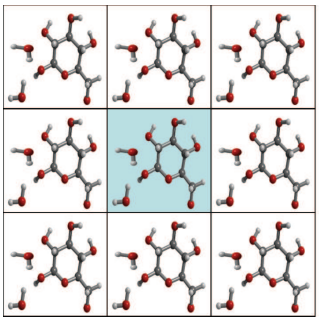


Figure 2.2: Periodic boundary conditons (PBC) demonstrated in two dimensions. The molecular system (center) is placed in a box surrounded by identical images of itself.

of the remaining boxes moves in exactly the same way and if a molecule leaves the central box, one of its images will enter through the opposite face. Since the central box is thus effectively unbounded, the system is virtually infinite. PBCs were used in the explicit water simulations in Paper I of methyl β-cellobiose and methyl β-laminarabiose.

2.1.6 Analysis of MD trajectories

A MD trajectory is a collection of regularly saved snapshots or "frames" of the positions (and sometimes momenta) of the atoms in a molecular dynamics simulation. As such, it provides a recorded history of the simulation, and it is through the analysis of the trajectory that insights from the simulation are made. When only average properties are of interest, as is the case in this thesis, the MD frames are usually not saved at every time step of the simulation since that would lead to excessive storage requirements. In the simulations in Paper I frames were saved at every 20th time step, which provides adequate sampling for our purposes. The main types of analysis performed on the molecular dynamics trajectories for the disaccharides in Paper I are briefly discussed below.

Time series

A time series is a plot of some variable of interest for each frame in the MD trajectory versus the simulation time. Through plots of this type, it is possible to achieve insights at a glance in important dynamical developments, such as changes in distances or dihedral angles. As an example, time series for the glycosidic dihedral angles Φ and Ψ for 10 ns simulations of methyl β-cellobioside and methyl β-laminarabioside are shown in **Figure 5.7(a)** and **Figure 5.7(b)**.

Hydration map

A description of the local water environment between two atoms of a solute is provided by the two-dimensional pair-distribution function [43]:

$$g(r_1, r_2) = \frac{N(r_1, r_2)}{\rho_{\text{water}} V_{\text{intersection}}(r_1, r_2, \Delta r)} \tag{2.8}$$

In the case of a saccharide, a water molecule may bridge two solute oxygen atoms and it is this situation one typically wishes to test for in the MD trajectories. Application of the function (2.8) to the MD trajectories, results in a two-dimensional probability density histogram for each solute oxygen pair combination, which is contoured and visually inspected for significant water densities. An example of the contour-plot is provided in **Figure 2.3** showing the highest density of water observed in this project, occurring between atoms 1O6 and 2O4 for methyl β-laminarabioside corresponding to six times the bulk density of water.

Heteronuclear scalar couplings

Determination of conformational preferences around the glycosidic linkages in sugars is of particular relevance for understanding structure and dynamics. One experimental measure for the glycosidic dihedral angles Φ and Ψ is the heteronuclear scalar coupling $^3J_{\text{CH}}$ from NMR spectroscopy. The Karplus relationship [44, 45] (2.9) establishes a relationship between the scalar coupling constant and the dihedral angle:

$$^3J_{\text{CH}}(\theta) = A + B\cos\theta + C\cos^2\theta \tag{2.9}$$

Various choices for the parameters A, B, and C in eq. (2.9) have been proposed. Although the parametrization by Tvaroska et al. [46] is well-known, we employed somewhat different parameters, see Paper I for details. Equation (2.9) is not invertible on its domain, implying that values of the dihedral angle cannot be obtained easily from the heteronuclear couplings. Hence in this project, rather than trying to determine dihedrals from NMR measurements via (2.9), the coupling constants were estimated from the molecular dynamics trajectories by evaluating

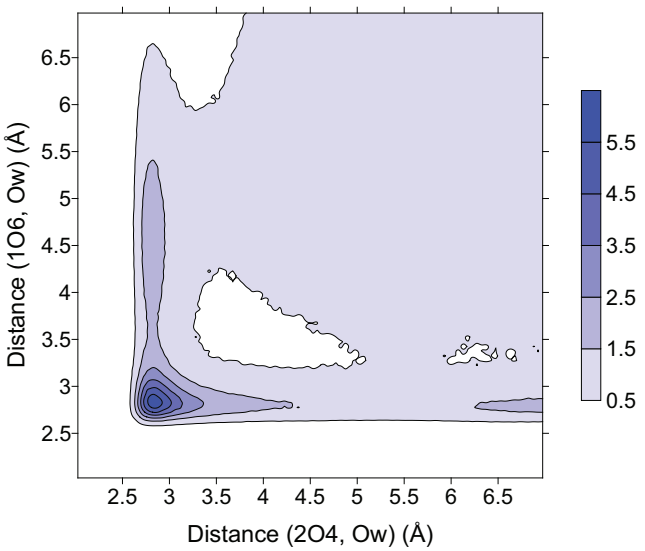


Figure 2.3: Contour plot of the probability distribution for finding a water molecule at a distance r from the two oxygen atoms 1O6 and 2O4 in methyl β-laminarabioside. Ow represents the oxygen atom in a water molecule. The color bar shows the probability density relative to bulk water.

(2.9) for all frames and averaging with the total number of frames. This approach is identical to that of Hansen [47].

2.2 Electronic structure methods

Only quantum mechanics can provide an accurate description of the electronic structure. All electronic structure methods are concerned with solving a time-independent Schrödinger equation which in the general form can be written as the eigenvalue problem:

$$\hat{H}\Psi = E\Psi, \tag{2.10}$$

where $\hat{H}$ is the Hamiltonian operator, E is the total energy, and Ψ is the electronic wave function to be determined. In most electronic structure methods, the Born-Oppenheimer approximation (BOA) is employed [48] which assumes that the movement of nuclei is insignificantly slow on the timescale of electron movement. This approximation has been shown to introduce insignificant errors in most applications [49].

Importance of the total energy

Virtually every quantity of interest can be derived from the total energy for an atomistic system [50]. Thus the primary goal of a electronic structure method calculation is to obtain the ground state energy. Given a nuclear configuration, the ground state energy can be estimated by fixing nuclei and minimizing the energy of the electrons in the constant potential of the nuclei.

The wave function

Formally, the properties of an interacting system of N electrons is described by the many-electron wave-function:

$$\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N), \tag{2.11}$$

where $\boldsymbol{r}_i$ denotes both spin and spatial degrees of freedom. Hartree [51] approximated the many-electron wave function as a product of single electron wave functions:

$$\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N) = \prod_{i=1}^{N} \psi_i(\boldsymbol{r}_i), \tag{2.12}$$

where each of the single-electron wave functions $\psi_i(\boldsymbol{r}_i)$ is the solution to a single-electron Schrödinger equation for a system where electron i experiences the mean electrical field of the remaining $N-1$ electrons.

$$\left(\frac{-\hbar^2}{2m}\nabla + V_{\mathbf{ext}} + \Phi_i\right)\psi_i(\boldsymbol{r}_i) = e_i\psi_i(\boldsymbol{r}), \tag{2.13}$$

where Φ_i and $V_{\mathbf{ext}}$ is the Coloumb and nuclear potential, respectively. The product wave function (2.12) does not obey the Pauli exclusion principle, however. To observe the Pauli exclusion principle, a wave function must be antisymmetric with respect to the interchange of arguments [52], that is to say it must satisfy $\psi(\boldsymbol{r}_1, \boldsymbol{r}_2) = -\psi(\boldsymbol{r}_2, \boldsymbol{r}_1)$. The simplest antisymmetric wave function is the Slater determinant, which in the general case has the form:

$$\Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(\boldsymbol{r}_1) & \psi_2(\boldsymbol{r}_1) & \cdots & \psi_N(\boldsymbol{r}_1) \\ \psi_1(\boldsymbol{r}_2) & \psi_2(\boldsymbol{r}_2) & \cdots & \psi_N(\boldsymbol{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\boldsymbol{r}_N) & \psi_2(\boldsymbol{r}_N) & \cdots & \psi_N(\boldsymbol{r}_N) \end{vmatrix} \tag{2.14}$$

By introducing the Slater determinant as a wave function, an additional energy term due to electron indistinguishability occurs in the Schrödinger equation. This term is known as the *exchange* term, and the formalism including it is known as

Hartree Fock (HF) theory. The problem with HF theory is that it omits another important contribution, namely electron *correlation*, and thus predicts that the probability of finding two electrons in the same region of space is equal to the product of their individual probabilities; this does not hold for electrons with parallel spins. Sometimes the effects of electron correlation are small and can be neglected without significant loss of accuracy. In the cases where electron correlation is important, a more accurate description of the many-electron wave function must be employed. On way to accomplish this is through the linear combination of Slater determinants, known as Configuration Interaction (CI), which provides an essentially complete description of the many electron wave function. Unfortunately, the precision comes at a high price: CI computations present a formidable computational barrier for all but the smallest systems.

2.2.1 Density Functional Theory

While configuration interaction and other so-called post Hartree-Fock methods encounter an exponential barrier when the number of electrons increases, density functional theory (DFT) presents an alternative approach allowing for the treatment of systems of much larger dimensions. As the name suggests, this theory focuses on the electron *density* as the principal function, rather than the high-dimensional many-electron wave function. Important quantities, such as the total energy E, are expressed as *functionals* (functions of functions). More exactly, the quantities are functionals of the electronic density. As pointed out by Kohn [53], the dimensionality reduction as well as the move from a abstract wave function to the more intuitive concept of electron density are important advantages of the method.

The ground state density

The fundamental assumption in the Hohenberg-Kohn formulation of DFT is that the ground state density $n(\mathbf{r})$ of a bound system of interacting electrons in some external potential $v(\mathbf{r})$ determines this potential uniquely. The simple proof can be found in e.g. [53].

The variational principle and DFT

Here it is discussed how to use the electronic density to find the electronic ground state energy E. It is assumed that a set of trial wavefunctions $\{\tilde{\Psi}^i_{\tilde{n}}\}$ exist, where each member of the set corresponds to the same electronic density $\tilde{n}(\mathbf{r})$. The energy minimum at this density can be found:

$$E_v[\tilde{n}(\mathbf{r})] \equiv \min_i \langle \tilde{\Psi}^i_{\tilde{n}} | \hat{H} | \tilde{\Psi}^i_{\tilde{n}} \rangle = \int v(\mathbf{r}) \tilde{n}(\mathbf{r}) d\mathbf{r} + F[\tilde{n}(\mathbf{r})], \qquad (2.15)$$

where $F[\tilde{n}(\mathbf{r})] = \min_i \langle \tilde{\Psi}^i_{\tilde{n}} | \hat{T} + \hat{U} | \tilde{\Psi}^i_{\tilde{n}} \rangle$ and $\hat{T}$ and $\hat{U}$ are kinetic energy and interaction potential energy operators, respectively. To get to the electronic ground state, minimization is carried out with respect to $\tilde{n}(\mathbf{r})$:

$$E = \min_{\tilde{n}(\mathbf{r})} E_v[\tilde{n}(\mathbf{r})] \tag{2.16}$$

The important point is that determination of E is formally achieved by minimization with respect to a three-dimensional function $\tilde{n}(\mathbf{r})$ instead of the $3N$ dimensional wavefunction. Using (2.15) and (2.16) the variational principle in terms of the electronic trial density $\tilde{n}(\mathbf{r})$ can be written:

$$E \leq E[\tilde{n}(\mathbf{r})] = \int v(\mathbf{r})\tilde{n}(\mathbf{r})d\mathbf{r} + F[\tilde{n}(\mathbf{r})] \tag{2.17}$$

Equation (2.17) is referred to as the Hohenberg-Kohn variational principle. The functional $F[\tilde{n}(\mathbf{r})]$ can be written more explicitly as a sum of electron energy contributions:

$$F[\tilde{n}(\mathbf{r})] \equiv T_s[\tilde{n}(\mathbf{r})] + E_{\text{Coulomb}}[\tilde{n}(\mathbf{r})] + E_{\text{xc}}[\tilde{n}(\mathbf{r})], \tag{2.18}$$

with

$$E_{\text{Coulomb}}[\tilde{n}(\mathbf{r})] = \frac{1}{2}\int \frac{\tilde{n}(\mathbf{r})\tilde{n}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}'. \tag{2.19}$$

Here $T_s[\tilde{n}(\mathbf{r})]$ is the kinetic energy functional for non-interacting electrons and $E_{\text{xc}}[\tilde{n}(\mathbf{r})]$ is the still undetermined exchange-correlation energy.

The self-consistent Kohn-Sham equations

Using Lagrangian formalism and the Hohenberg-Kohn variational principle, Kohn and Sham showed [53] that the minimizing electronic density $n(\mathbf{r})$ for a system of *interacting* electrons can be found by solving a set of single particle equations:

$$\left(-\frac{1}{2}\nabla^2 + v_{\text{eff}}(\mathbf{r})\right)\phi_j(\mathbf{r}) = \epsilon_j \phi_j(\mathbf{r}), \tag{2.20}$$

where

$$v_{\text{eff}}(\mathbf{r}) = v(\mathbf{r}) + \int \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + v_{\text{xc}}(\mathbf{r}) \tag{2.21}$$

The first two terms in this effective potential are the external potential and the Coulomb potential. The last term is defined as the functional derivative of the exchange-correlation energy:

$$v_{\text{xc}}(\mathbf{r}) \equiv \frac{\delta}{\delta\tilde{n}(\mathbf{r})} E_{\text{xc}}[\tilde{n}(\mathbf{r})]|_{\tilde{n}(\mathbf{r})=n(\mathbf{r})} \tag{2.22}$$

To reach the ground state, (2.20) must be solved *self-consistently* with respect to the electronic density. Thus one inserts a trial $n(\mathbf{r})$ into (2.21) and evaluates (2.20). Next, one evaluates:

$$\tilde{n}(\mathbf{r}) = \sum_{j=1}^{N} |\phi_j(\mathbf{r})|^2, \tag{2.23}$$

where the sum runs over all occupied electronic states. If this is not equal to the initial guess, the calculation is repeated with $\tilde{n}(\mathbf{r})$ as the new initial value. Self-consistency is achieved when the electronic density used to calculate the Kohn-Sham wavefunctions is the same as the density calculated from these wavefunctions. At this point, the electronic ground state density is reached and the ground state energy can be calculated:

$$E = \sum_j \epsilon_j + E_{\mathrm{xc}}[n(\mathbf{r})] - \int v_{\mathrm{xc}}(\mathbf{r}) n(\mathbf{r}) d\mathbf{r} - \frac{1}{2} \int \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} \tag{2.24}$$

The exchange-correlation functional

Employment of the Kohn-Sham DFT formulation requires an expression for the exchange-correlation functional $E_{\mathrm{xc}}[n(\mathbf{r})]$. Setting this energy contribution to zero yields the Hartree self-consistent equations, which completely neglect exchange-correlation effects [54]. On the other hand, an exact expression for the exchange-correlation energy would give an exact solution to the many-electron ground state problem [50]. In general however, there is no analytical form for $E_{\mathrm{xc}}[n(\mathbf{r})]$ so at this point an approximate expression for the quantity is introduced.

The B3LYP functional

In quantum chemistry, the B3LYP functional is arguably one of the most widely used approximations for the exchange-correlation functional in DFT. It is a hybrid functional, meaning that it linearly combines exact exchange from Hartree-Fock theory, $E_{\mathrm{x}}^{\mathrm{HF}}$, with correlation functionals and other types of exchange functionals. B3LYP has the form:

$$E_{\mathrm{XC}}^{\mathrm{B3LYP}} = (1-a_0)E_{\mathrm{x}}^{\mathrm{LSDA}} + a_0 E_{\mathrm{x}}^{\mathrm{HF}} + a_{\mathrm{x}} \Delta E_{\mathrm{x}}^{\mathrm{B88}} + a_{\mathrm{c}} E_{\mathrm{c}}^{\mathrm{LYP}} + (1-a_{\mathrm{c}}) E_{\mathrm{c}}^{\mathrm{VWN}}, \tag{2.25}$$

where $E_{\mathrm{x}}^{\mathrm{LSDA}}$ is the standard local exchange functional [55], $\Delta E_{\mathrm{x}}^{\mathrm{B88}}$ is Becke's gradient correction [55, 56] to the exchange functional, $E_{\mathrm{c}}^{\mathrm{LYP}}$ is the correlation functional by Lee et al. [57], and $E_{\mathrm{c}}^{\mathrm{VWN}}$ is the local correlation functional by Vosko et al. [58]. The coefficients $a_0 = 0.2$, $a_{\mathrm{x}} = 0.72$, and $a_{\mathrm{c}} = 0.81$ were determined

by Becke by fitting to heats of formation for a set of small molecules [55]. B3LYP often gives good energies in general applications (see e.g. [59]), although recent studies by Csonka et al. [60] showed better performance of other functionals in calculations on simple carbohydrates. In this project B3LYP was employed in the calculation of geometries and VCD spectra for ginkgolide B (Section 5.6) and in the study of influence of geometry optimization on QSAR (Section 5.5).

2.2.2 Basis sets

The molecular wave function obtained from an electronic structure method calculation is a numerical solution to a relevant Schrödinger-type equation. In calculations, the numerical solution is constructed by superposition of a finite number of simple basis functions [59, 61]. A molecular orbital can be constructed, for instance, by the linear combination of atomic orbitals. However, while conceptually appealing, this representation is not suitable for the purpose of computational quantum chemistry where more computationally efficient basis functions are required. Slater [62] introduced a type of orbital, aptly referred to as the Slater Type Orbital (STO), which is a product of an orbital-type determining spherical harmonic function $Y_l^m(\theta, \phi)$ with angular momentum quantum numbers l and m and a radial function of the form:

$$\frac{\zeta^3}{\pi^{0.5}} e^{-\zeta r}, \tag{2.26}$$

where the pre-exponential factor is a normalization constant, ζ is a measure of the size of the orbital, and r is the distance from the nucleus. STOs provide a favorable description of orbital behavior, but their direct use in quantum chemical calculations is complicated by the fact that they offer no analytical solution for three- and four-center two-electron integrals (see e.g. [61]). On the other hand, these integrals can be solved analytically by employing the Gaussian Type Orbital (GTO) introduced by Boys [63]. The radial part of a GTO has the form:

$$\frac{2\chi}{\pi^{0.5}} e^{-\chi \boldsymbol{r}^2} \tag{2.27}$$

The quantum chemical program Gaussian [64] used in this project employs GTOs as primitive basis functions. By combining GTOs, the desirable features of the STOs are approximated, while allowing for more efficient calculations [65].

Pople basis sets

A plethora of basis sets are available, but for the quantum calculations in this project, only basis sets of the Pople type [66] were employed. A *minimal basis set* assigns one basis function for the description of each orbital in the neutral atom.

The designation for a minimal Pople style basis set is STO-nG [65], meaning that n GTOs are fitted to approximate one STO. This approximate STO, consisting of a *contraction* of GTOs, is the basis function used for the description of each orbital. For hydrogen, a single basis function is employed while five functions are used in case of carbon, corresponding to the 1s, 2s, 2px, 2py, and 2pz orbitals. A minimal basis set cannot adequately describe different bonding in different directions [59]. The remedy is to employ an *extended basis set* which assigns more than the minimum number of required basis functions to each orbital. For example, a double-zeta orbital is a linear combination of two approximate STOs (again formed by contraction of GTOs) each characterized by a different value of ζ and therefore different "tightness":

$$\phi(\boldsymbol{r}) = \phi^{STO}(\boldsymbol{r}, \zeta_1) + d \cdot \phi^{STO}(\boldsymbol{r}, \zeta_2), \tag{2.28}$$

where d varies the amount of $\phi^{STO}(\boldsymbol{r}, \zeta_2)$ and hence allows for better description of orbital flexibility. Double-zeta (and multiple-zeta, in the general case) orbitals are expensive in terms of computational time. Therefore, instead of extravagantly applying them for the description of all orbitals, they are often only assigned to the valence orbitals in what is known as a *split valence* approach. Core electrons are then described with the less expensive GTOs. Split-valence basis sets of the Pople type have the general form k-nlmG, where k is the number of GTOs used for description of the core orbitals and the integers n, l, and m describe the number of functions the valence basis functions are split into as well as the number of GTOs used for each function. As an example, the 6-31G basis set employs a contraction of 6 GTOs for describing the core electrons and a double-zeta orbital for the valence electrons. The double-zeta orbital is the linear combination of two approximate STOs, the first being composed of $n = 3$ GTOs and the second consisting of $l = 1$ GTO.

Diffuse functions and polarization functions

Additional flexibility in the description of valence electron density can be achieved by introducing diffuse functions [67] and polarization [68] functions. Diffuse functions are GTOs with very small exponents (ζ) and their inclusion enables a better representation of the tail behavior of the orbitals. They improve the description of phenomena such as hydrogen bonding and van der Waals interactions [69]. In the Pople type basis sets the use of diffuse functions is indicated with '+' or '++' immediately in front of the symbol G, indicating diffuse functions on heavy atoms only or on all atoms, respectively.

The use polarization functions implies that the shape of basis functions is allowed to distort. This is important in order to provide an accurate description of electron density distributions of low symmetry. For instance, perturbations

in electron density induced by an electrical field call for the use of polarization functions. An illustration of polarization is provided in **Figure 2.4** where a perfect p orbital is distorted by the addition of a d function. The use of polarization functions is indicated by one or more symbols immediately after the G in Pople style basis sets. For example, in 6-31G(d,p) the (d,p) indicates polarization d functions on heavy atoms and p functions on hydrogens.

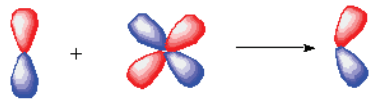


Figure 2.4: Example of polarization: A p orbital is distorted by mixing in a d polarization function. Image taken from [70].

2.2.3 Nuclear magnetic shielding

The diagonal elements of the nuclear magnetic shielding tensor can be written as [59, 71]:

$$\sigma_{ii} = \frac{\alpha^2}{2}\langle 0|\frac{r^2 - r_i^2}{r^3}|0\rangle - 2\alpha^2 \sum_{n_s \neq 0} \frac{\langle 0|\frac{l_i}{r^3}|n_s\rangle\langle n_s|l_i|0\rangle}{E_{n_s} - E_0}, \tag{2.29}$$

where $|0\rangle$ is the ground state, $|n_s\rangle$ is a singlet excited state, E_0 and E_n are the energies corresponding to these states, l_i is the orbital angular momentum, and α is the fine-structure constant. The first *diamagnetic* term of (2.29) depends on r^{-1} while the the second *paramagnetic* term depends on r^{-3}. In this project the GIAO [72] (Gauge Independent Atomic Orbital) approach was used to calculate ^{1}H chemical shielding constants. This method includes an exponential term containing the vector potential $\mathbf{A}$ with each atomic orbital. In this way, the arbitrary choice of origin and form (gauge) of the vector potential is removed . The following approximation is used to obtain the chemical shift from the isotropic shieldings:

$$\delta = \sigma_{ref} - \sigma_{iso}, \tag{2.30}$$

where δ is the chemical shift, σ_{iso} is the isotropic shielding and σ_{ref} is the isotropic shielding for a reference molecule. When only changes in chemical shifts are desired, σ_{ref} cancels out:

$$\Delta\delta = \delta_2 - \delta_1 = \sigma_1 - \sigma_2 \tag{2.31}$$

2.3 Molecular descriptors

A molecular descriptor is a number describing some aspect of a molecule. In special cases molecular descriptors are measured, but the term usually refers to computed quantities. Simple examples of molecular descriptors are the molecular weight, the number of a particular bond or atom type, or the van der Waals volume. Quantities such as HOMO/LUMO energies, polarizability, dipole moments and partial charges are readily calculated with quantum chemical methods and can also be used as molecular descriptors. In addition to molecular descriptors with a clear physicochemical connection, a large variety of more abstract molecular descriptors exist. Several specialized programs for the calculation of molecular descriptors can be purchased or downloaded freely from the web. In this thesis the commercial programs DRAGON [73] and the open-source descriptor calculator based on the Chemistry Development Kit [74] were employed for the calculation of a large selection of generic descriptors, while QikProp [75] was used for the calculation of a smaller set of pharmaceutically relevant descriptors. Molecular descriptors can be divided into several classes. Those with an obvious physicochemical connection are sometimes (e.g. in the DRAGON program [73] referred to as molecular properties. Lipophilicity, often denoted log p, is an example of a molecular property employed in classical studies, see e.g. [76].

2.3.1 Descriptors from molecular graphs

A large group of molecular descriptors is derived from a simple representation of the molecule as a molecular graph where vertices and edges represent atoms and bonds, respectively. This way of representing a molecule was pioneered by Cayley in 1874 [77]. Typically only non-hydrogen atoms are included, as illustrated for 2,2,3,5-tetramethylhexane in **Figure 2.5**. In Paper II, it was found that various descriptors derived from such graphs provided excellent QSPR relationships for the release of flavor compounds from β-glucan matrices. The following is a brief description of the origins of the descriptors ZM1V, PCR, GMTIV and DECC found to be important in the QSPR models for flavor release from five percent PromOat and Glucagel matrices in Paper II. A more exhaustive description is offered in e.g. [78].

Topological descriptors

Topological descriptors describe the chemical bonding pattern of atoms within molecules without consideration of the molecular geometry, and can encode features such as size, shape, symmetry, and branching and may also represent chemical information such as atom type and bond order [78, 79]. The topological descriptors are derived using appropriate algebraic operations on a matrix represen-

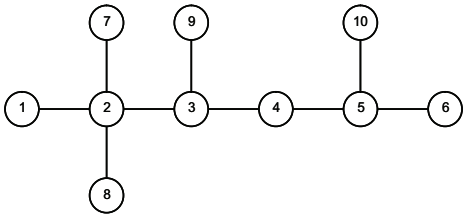


Figure 2.5: Hydrogen-depleted molecular graph representation of 2,2,3,5-tetramethylhexane. The numbering is arbitrary, but each number is an index to the adjacency matrix in (2.32).

tation of the molecular graph. As an example the adjacency matrix corresponding to **Figure 2.5** is shown in (2.32).

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{2.32}$$

Row (or column, since the adjacency matrix is symmetric) i corresponds to atom i in **Figure 2.5** and an element n in that row assumes the value 0 or 1 if atom i is unconnected or connected, respectively, to atom n. The *vertex degree*, δ_i, for an atom i is the number of atoms it is connected to, which is also the sum of a row in the adjacency matrix. The *valence* vertex degree, δ_i^v, for an atom i is defined as:

$$\delta_i^v = \delta_i + p_i + n_i = \sigma_i + p_i - h_i, \tag{2.33}$$

where p_i is the number of p orbital electrons, n_i is the number of lone-pair electrons, σ_i is the number of σ-bonds, and h_i is the number of hydrogen atoms bonded to the atom [80]. The first Zagreb index by valence vertex degrees, **ZM1V**, is defined as:

$$\mathrm{ZM1V} = \sum_{i=1}^{n} (\delta_i^v)^2, \tag{2.34}$$

where the summation runs over all non-hydrogen atoms [81]. **GMTIV** is the Gutman Molecular Topological Index by valence vertex degrees [82] which is

calculated as:

$$\text{GMTIV} = \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i^v \delta_j^v - d_{ij}, \tag{2.35}$$

where the summations run over all non-hydrogen atoms and d_{ij} is the topological distance between atoms i and j.

Other topological descriptors arise from the topological distance matrix, which differs from the adjacency matrix in that each element i, j holds the minimum topological distance (minimum number of bonds) between atoms i and j [78]. Analogous to the vertex degree of the adjacency matrix, a distance degree for the topological distance matrix is row (column) sum. The eccentricity ECC is defined as the sum over all non-hydrogen atoms of the maximum topological distance from an atom to any other atom (the atom eccentricity). The average eccentricity AECC is ECC divided by the number of non-hydrogen atoms. The descriptor **DECC** [83, 84] is a measure of the mean displacement of atom eccentricities from AECC.

Molecular walk and path counts

Molecular walk counts (MWC) [85] are also derived from hydrogen depleted molecular graph. They are related to molecular size and branching and are a measure of the complexity of the molecular graph. A walk in the molecular graph is a sequence adjacent edges (bonds) leading from one vertex (atom) to another, where any edge can be traversed several times [78]. The walk length is the number of edges in the walk. The molecular walk count MWCk of order k is the total number of walks of kth length in the molecular graph. The MWC of order zero is the number of graph vertices, and the MWC of order 1 is the number of graph edges. A molecular path [86, 87] is a special case of a walk; in the molecular path, each edge or vertex may only be traversed once. TPC is the total path count, which is the number of all possible paths of a defined maximum length. A path can be assigned a weight [88] evaluated by multiplication of the conventional bond order of all edges. Summing the weights of all paths of any length up to a chosen maximum, yields the conventional bond-order ID number (piID). Dividing piID by TPC yields **PCR**; the ratio of multiple path count over path count.

3 Chemometrics

3.1 Linear regression

A class of problems exists, where the relationship between a dependent variable y and a series of independent variables x_i can be approximated by a function of the form [89]:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \cdots + \beta_n \cdot x_n = \sum_{i=1}^{n} \beta_i \cdot x_i, \tag{3.1}$$

where β_i are unknown parameters, and x and y are known. In the case $n = 1$, equation 3.1 simplifies to a first order model, describing a straight line. For $n > 1$ it is a higher order polynomial. However, as long as the function is first order in all the parameters, the model is linear. By considering a finite number m of samples of the dependent and independent variables, equation 3.1 can be written in matrix form:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_m \end{bmatrix} \tag{3.2}$$

or more compactly:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}, \tag{3.3}$$

where variables are assumed to be mean centered and the residual vector $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$ has been introduced. The sum of squared residuals (SSR $= \boldsymbol{e} \cdot \boldsymbol{e}$) is a measure of the overall fit of the model. The problem solved by regression is the determination of a set of parameters β_i that minimizes $\boldsymbol{e} \cdot \boldsymbol{e}$. It can be shown, that the solution for this least squares problem is given by:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}, \tag{3.4}$$

where $(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}$ is the Moore-Penrose pseudoinverse [90, 91] of $\boldsymbol{X}$. This approach has been used liberally in various fields of science, including QSAR [92, 93]. The simplicity of the method and its easily interpretable models has undoubtedly contributed to its success. However, the matrix inversion in (3.4) is only well-conditioned as long as the variables in $\boldsymbol{X}$ are non-collinear. It is not uncommon, however, to encounter collinear variables in applications such as modern QSAR. For instance, collinearity always occurs when the system of linear equations is underdetermined ($m < n$). Furthermore, the variables (molecular descriptors) are often strongly correlated. These situations would result in division by a number close to zero in the matrix inversion. One solution is to use a variable selection method to reduce the dimensionality of X and hence the collinearity. This is not always attractive, since problem relevant information may be lost. One senses that progress could be made, if there was a way to retain only the non-collinear, problem relevant part of the information in X. The introduction of principal components (PCs) is a key to resolving the problem.

3.2 Principal Components Analysis

Each row (sample) of the mean-centered matrix $\boldsymbol{X}$ can be represented as a point in the space spanned by the p columns (variables). The first principal component (PC1) is defined as the direction in the p-dimensional sample space that maximizes the variance between the m samples. The second principal component (PC2) is a direction orthogonal to (PC1) that provides the next largest variance. This decomposition of the sample space into mutually orthogonal maximum variance directions can proceed until there are $min((m-1), p)$ PCs. In more mathematical terms, the direction of a PC in the sample space is described by a vector $\boldsymbol{p}$, termed the loading vector. $\boldsymbol{p}$ is a linear combination of the n original variables. The coordinates of the m samples along $\boldsymbol{p}$ are contained in the m-dimensional score vector, $\boldsymbol{t}$. In principal components analysis (PCA), the data matrix $\boldsymbol{X}$ is expanded into a sum of outer products of scores and loadings (structure part) plus an error matrix (noise part) [94]:

$$\boldsymbol{X} = \sum_{i}^{s} \boldsymbol{t}_i \boldsymbol{p}_i^{\mathrm{T}} + \boldsymbol{E} = \boldsymbol{T}\boldsymbol{P}^{\mathrm{T}} + \boldsymbol{E} \qquad (3.5)$$

The error matrix $\boldsymbol{E}$ is present if the product sum (3.5) is truncated after $s < n$ PCs and contains the difference $\boldsymbol{X} - \boldsymbol{T}\boldsymbol{P}^{\mathrm{T}}$. PCA offers tremendous data analytical power, as the essential features of a data matrix of high dimensionality can often be represented by a few PCs. In connection with QSAR, for instance, the major relationships between hundreds of descriptors can be achieved at a glance by observing two-dimensional score or loading plots. As indicated, the

core principles of PCA are also a valuable tool in the improvement of the regression methods of the previous section. By representing the samples as coordinates on orthogonal axes pointing in maximum variance directions, the PCA decomposition effectively eliminates the collinearity problem of the original data matrix $\boldsymbol{X}$. Thus by employing a matrix containing a suitable number of score vectors instead of X in , the matrix inversion problem never occurs. This approach is called principal components regression (PCR). While free from the matrix singularity problems of MLR, PCR may not be ideally suited for all applications. This is because the maximum variance directions in the $\boldsymbol{X}$ matrix are not necessarily those of importance in describing the variation in the dependent variable y. This shortcoming is addressed in partial least squares (PLS) regression.

3.3 Partial Least Squares regression

In contrast to the two-step procedure of PC regression (independent PCA decomposition on $\boldsymbol{X}$ followed by MLR using the scores), partial least squares (PLS) regression [95–97] employs $\boldsymbol{y}$ information already in the decomposition of $\boldsymbol{X}$. This guarantees that only information relevant to describing $\boldsymbol{y}$ is used in the regression. In other words PLS seeks to establish the maximum covariance between directions in a matrix of predictors (e.g. molecular descriptors) and a matrix of dependent variables (e.g. properties/activities). In the context of QSAR/QSPR, the activity or property is usually a single vector $\boldsymbol{y}$. For a univariate $\boldsymbol{y}$ block the NIPALS PLS algorithm [95] assumes the following simple form:

1. $\boldsymbol{w}_i = \boldsymbol{X_i}^{\mathrm{T}}\boldsymbol{y_i}/||\boldsymbol{X_i}^{\mathrm{T}}\boldsymbol{y_i}||$
2. $\boldsymbol{t}_i = \boldsymbol{X}_i\boldsymbol{w}_i$
3. $q_i = \boldsymbol{y}_i^{\mathrm{T}}\boldsymbol{t}_i/\boldsymbol{t}_i^{\mathrm{T}}\boldsymbol{t}_i$
4. $\boldsymbol{p}_i = \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{t}_i/\boldsymbol{t}_i^{\mathrm{T}}\boldsymbol{t}_i$
5. $\boldsymbol{X}_{i+1} = \boldsymbol{X}_i - \boldsymbol{t}_i\boldsymbol{p}_i^{\mathrm{T}} \qquad \boldsymbol{y}_{i+1} = \boldsymbol{y}_i - q_i\boldsymbol{t}_i$
6. $i = i + 1$

In the above, $\boldsymbol{w}_i$, $\boldsymbol{p}_i$, and $\boldsymbol{t}_i$ are known as loading weights, loadings, and scores, respectively. The regression coefficients $\boldsymbol{\beta}$ are obtained from the PLS model via:

$$\boldsymbol{\beta} = \boldsymbol{W}(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{W})^{-1}\boldsymbol{Q}^{\mathrm{T}}, \tag{3.6}$$

where $\boldsymbol{W}$ and $\boldsymbol{P}$ are matrices containing the p loading weights $\boldsymbol{w}_i$ and loadings $\boldsymbol{p}_i$ for a PLS model with p LVs. The quality of the PLS model is evaluated after each iteration of the above sequence of steps, and the algorithm continues until there is no improvement. At this point the model is based on an optimal number of PLS components or latent variables (LV). Typically the measure of

the quality of the model is the root mean square error of leave-one-out (LOO) cross validation (RMSECV). This parameter is evaluated as:

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}, \tag{3.7}$$

where y_i is the measured value for sample i and $\hat{y}_i$ is the value predicted from a PLS model built from all samples *except* sample i. The LOO type of cross-validation can be extended by leaving out more samples at a time and predicting their response values using the PLS model based on the remaining samples. This is leave-many-out or segmented cross-validation.
In QSAR, a widely used measure of model quality is q^2 [98]:

$$q^2 = 1 - \frac{\sum_{i}^{n}(y_i - \hat{y}_i)^2}{\sum_{i}^{n}(y_i - \bar{y})^2}, \tag{3.8}$$

where $\bar{y}$ is the mean of the measured $\boldsymbol{y}$.

3.4 Variable selection

Due to the large ratio of molecular descriptors to compounds encountered in many QSAR applications, it is often desirable to employ some sort of variable selection to improve the robustness and interpretability of the final model [99]. In Paper II, forward variable selection was employed to determine a suitable subset of molecular descriptors. When used on a block with n molecular descriptors, forward selection starts by building n models for the prediction of $\mathbf{y}$, each containing only the ith descriptor. The model with the lowest value of RMSECV is selected, and the method proceeds by trying to improve this model by inclusion of one of the remaining $n - 1$ descriptors. This stepwise inclusion of descriptors continues until there is no improvement in model quality or a predefined number of descriptors (or latent variables) is reached.

3.5 Model validation

It is important that any statistical model is subjected to validation, not least in multivariate regression, where the risk of chance correlation increases with the number of independent variables. As noted above, a form of validation is already incorporated in the construction of a PLS model, since the optimal model dimensionality is chosen at the minimum of a plot of a conveniently chosen error-of-prediction function versus the number of LVs. However, whether the model dimensionality is determined using using full (leave-one-out) cross-validation or

the arguably better alternative segmented cross-validation, the final PLS model should be tested for robustness. The test employed in the QSPR study in Paper II is discussed in Section 3.6.1.

3.6 Quantitative Structure Activity Relationships

Almost one and a half century ago Crum-Brown and Fraser suggested that the physiological action of a compound is a function of its chemical composition and constitution [100]. Roughly hundred years later Hansch [76] combined this idea with a suitable mathematical framework and thus laid the foundation for what is today known as quantitative structure activity relationships (QSAR).
The central idea in QSAR is to predict or explain the variation in some expensive or complex molecular activity or property via more easily obtainable or desirable variables. Toxicity, for instance, is an obvious example of a molecular property that is complex and for which there may be an interest in limiting direct measurement. The task of QSAR in that particular case would be to determine other molecular properties that predicts the toxicity quantitatively.
Although there are variants of the list (see e.g. [101, 102]) a QSAR study often employs several of the following steps, which could loosely be said to constitute a generalized "QSAR protocol":

1. Preparation of input data: Building initial molecular structures, obtaining activities (properties) for molecules
2. Improving the molecular structures through conformational analysis, geometry optimization, and/or alignment
3. Calculation of molecular descriptors
4. QSAR model building: Establishing relationship between molecular descriptors and target activity (property) using (multivariate) regression methods (e.g. MLR, PCR, PLS) or mapping (ANN).
5. Validation, interpretation and prediction

From the list above it is noted that the method combines concepts from molecular modeling (generation of molecular structures, molecular descriptors, etc.) and chemometrics (multivariate regression).

3.6.1 Robustness of QSPR models

Since QSAR models are multivariate in all but the simplest cases, their robustness must be scrutinized. The best way of doing this is a matter of debate, see e.g. [103] for a discussion. Since QSAR/QSPR models in general may serve

two different goals, explanation or prediction, the type of validation varies accordingly. The QSAR studies in this thesis were mainly explorative in nature, and the validation scheme employed was thus predominantly a test of internal robustness. In practice, we employed a response permutation test where many regression models were built from the same block of molecular descriptors against a permuted ("scrambled") response vector. In the literature, this is sometimes referred to as a $\boldsymbol{y}$-randomization test [104]. Plotting the quality of models built from permuted response vectors against the correlation coefficient between the original response and its permutation, gave an indication of the stability of the QSPR model. An example of such a test is shown in **Figure 3.1** for two of the QSPR models for flavor release discussed in Paper II. It is seen that the model error (segmented RMSECV) increases as the correlation between the original response vector and its permutated cousin decreases. Since the scrambled models thus perform significantly worse than the original model, the latter is taken to represent a real phenomenon.

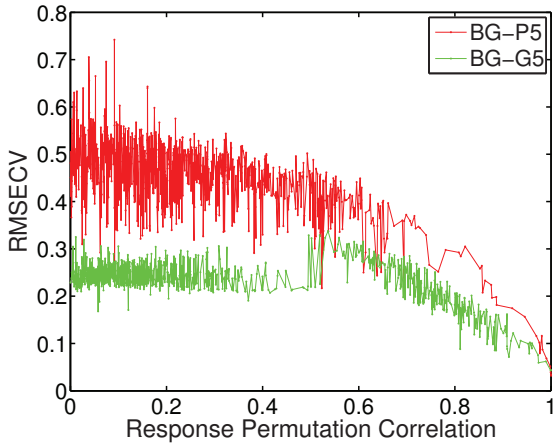


Figure 3.1: Example of response permutation tests for two QSPR models for flavor release from Paper II.

4
Experimental methods

Liquid and gaseous phases are natural environments for food constituents. Thus probing the behavior of molecular food components in these environments is clearly imperative for gaining relevant insights. From the perspective of this thesis, the methods that generated the experimental data for the computational investigations can be divided conceptually into two groups. The first group deals mainly with *structure* of molecular food components and includes NMR spectroscopy (Section 4.1) and IR/VCD spectroscopy (Section 4.3). The second group of methods deals mainly with *transport* of molecular food components. This transport is in general of immense importance in food science. For instance, the transport (release) of aroma compounds from a food matrix to air determines the flavor of a food, and thereby its appeal to customers [105]. The flavor release can be investigated with dynamic headspace GC/MS (Section 4.4). Also movement of food molecules between miscible phases is of importance. For instance, solvated dietary fibres, such as BG, may partially exert their beneficial influence by absorbing noxious substances from the aqueous environment in the gut [106]. This type of transport can be modeled in an equilibrium dialysis assay (Section 4.5).

4.1 NMR spectroscopy

4.1.1 Nuclear magnetic resonance

Nuclear magnetic resonance (NMR) spectroscopy is the most widely used method for the elucidation of molecular structure in solution and is unique in providing dynamical information over a large range of timescales [107]. In the experimental set-up, a sample is placed in a strong uniform magnetic field in the z-direction which lifts the degeneracy of nuclear spin states due to the differences in energy when the associated nuclear magnetic moments are aligned parallel or antiparallel

with the magnetic field. This is a manifestation of the nuclear Zeeman effect [108]. Since a magnetic moment only exists for nuclei with a total spin I, only such nuclei can participate in the spin polarization. It is a convenient fact of nature that ^{1}H and ^{13}C both have $I = 1/2$ and that the isotopes are present in virtually all molecules of biological relevance. The relevant z component of the magnetic moments for these nuclei can be obtained from the relationship $\mu_z = g\mu_N m_I$, where $\mu_N = 3.1524516628 \cdot 10^{-8}$ eV/T is the nuclear magneton, g is the gyromagnetic ratio for the nucleus, and $m_I = \pm 1/2$ is the quantum number for the z-component of the spin. The energy of the magnetic moment in the magnetic field is the product of the z-components of the B-field and magnetic moment $E_i = Bg\mu_N m_i$ and the difference in energy between the two spin-states for the $I = 1/1$ situation in the field is:

$$\Delta E = E_{m=-1/2} - E_{m=1/2} = \gamma \hbar B \tag{4.1}$$

If an incoming photon with energy $\hbar\omega$ fulfills $\hbar\omega = \Delta E$, a transition between the two spin states occur. This is nuclear magnetic resonance. For an isolated proton, the required frequency of the incoming electromagnetic radiation is 42.5781 Mhz per Tesla [109]. Relaxation to the lower energy state is accompanied by the emission of a photon with the same energy. At equilibrium only a very small fraction of nuclei occupy the low energy state and are hence candidates for absorption of radiation. However, as evident from (4.1) the splitting of energy levels ΔE is directly proportional to the strength of the magnetic field. This is illustrated schematically in **Figure 4.1**.

4.1.2 NMR spectroscopy

NMR spectroscopy takes advantage of the fact that the electronic environment of a nucleus perturbs the magnetic field experienced by the nucleus and hence changes the energy difference (Equation 4.1) of the spin states. Thus identical nuclei, protons for example, will exhibit slightly different resonance frequencies depending on their chemical environment in a molecule [110]. The modulation of the local magnetic field at nucleus due to the electronic structure is described by the 3×3 shielding tensor $\boldsymbol{\sigma}$. The actual field at the nucleus arising from both the external field and the shielding can be expressed as $\boldsymbol{B}_{local} = \boldsymbol{B} - \boldsymbol{\sigma}\boldsymbol{B}$ and the measured variation in nuclear magnetic resonance frequency due to shielding is known as the chemical shift. For measurements in solution, only the average of the trace of the shielding tensor is relevant:

$$\sigma_{iso} = \frac{1}{3}(\sigma_{xx} + \sigma_{yy} + \sigma_{zz}) \tag{4.2}$$

As mentioned in Section 2.2.3 the shielding tensor can be calculated with quantum mechanical methods. This is employed in Section 5.4.1.

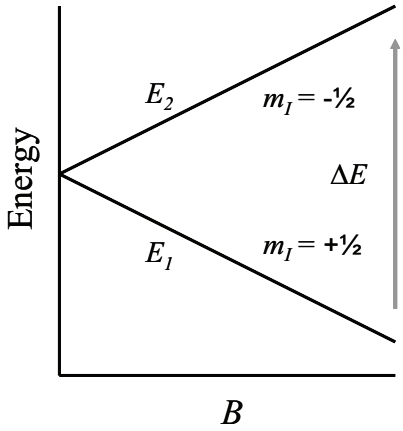


Figure 4.1: Energy difference between spin states as a function of the strength of the external magnetic field B.

NMR and disaccharide conformation

In connection with the experimental investigation of carbohydrate conformation and hydration, the most relevant information that can be obtained from NMR measurements is arguably (1) the heteronuclear scalar couplings, $^3J_{\mathrm{CH}}$, pertaining to the glycosidic dihedral angles Φ and Ψ and (2) the strength of the nuclear Overhauser effect, NOE [111, 112], particularly between protons on different glycosidic residues. The effects are illustrated in **Figure 4.2**. Two-dimensional

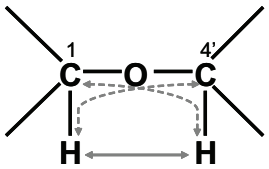


Figure 4.2: Illustration of the heteronuclear scalar couplings between C and H through the glycosidic linkage (dashed arrows) and NOEs between two hydrogens through space (solid arrows). Inspired by Figure 4.10 in [47], page 47.

NMR experiments are employed for the measurement of these effects, and the J-HMBC experiments by Meissner and Sørensen [113] were used in Paper I in the determination of the relevant heteronuclear scalar coupling constants $^3J_{\mathrm{H1-C4'}}$

and ${}^3J_{\mathrm{C1-H4'}}$ for methyl β-cellobioside and ${}^3J_{\mathrm{H1-C3'}}$ and ${}^3J_{\mathrm{C1-H3'}}$ for methyl β-laminarabioside. Comparison of these measurements with values calculated from molecular dynamics simulations is possible via the use of a Karplus-like [44, 45] equation as described in Section 2.1.6. The NOEs which are measured in ROESY [114] and NOESY [115] experiments, can sometimes be rationalized by comparison with average structural and dynamical trends observed in computations. For instance the hydration maps obtained from molecular dynamics trajectories (Section 2.1.6) may suggest conformational constraints compatible with NOEs.

Aromatic shielding

The shielding and de-shielding effects observed in aromatic systems have traditionally been attributed to ring-current magnetic anisotropy [116]. The view of circulating π-electrons as the sole cause of aromatic magnetic anisotropy has been challenged by ab inito calculations of the shielding contributions [117]. These calculations indicate a more complex picture where shielding and de-shielding zones in aromatic system can only partially be attributed to ring currents. In the light of these findings, we decided to abandon proton shift prediction using parameterizations based on the simple ring-current picture, such as the Johnson-Bovey equation [118]. Instead, a computational scheme similar to that used in [119] was used, where NMR shieldings are calculated at the HF/6-31G(d,p) level of theory.

4.2 IR spectroscopy

Infrared absorption spectroscopy determines transitions associated with changes in the vibrational states of molecules. The simplest classical picture of a molecular vibration emerges from the consideration of two atoms with masses m_1 and m_2 connected by an ideal spring with spring constant k. The displacement of the atoms from the center of mass is described by:

$$\mu \frac{d^2 r}{dt^2} = -k(r - r_{eq}), \tag{4.3}$$

where $\mu = m_1 m_2/(m_1 + m_2)$ is the reduced mass, r is the distance between the atoms and k is the spring force, which is zero at the equilibrium distance r_{eq}. The motion corresponding to (4.3) is that of a harmonic oscillator, i.e. the force is derived from the quadratic potential energy expression $V(r) = \frac{1}{2}k(r - r_{eq})^2$. This *harmonic approximation* is only valid for small fluctuations around r_{eq}. As the interatomic separation becomes larger, higher order terms in the potential

are required to account for anharmonicity. The classical solution for (4.3) is:

$$r(t) = Acos(\omega t), \tag{4.4}$$

where $\omega = \sqrt{(}k/\mu)$. Substituting the quantum mechanical operators for the classical kinetic and potential energy terms yields the Hamiltonian:

$$\hat{H} = \frac{1}{2}kr^2 - \frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial r^2} \tag{4.5}$$

The eigenvalues for the corresponding Schrödinger equation $\hat{H}\Psi = E_n\Psi$ are:

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega \qquad (n = 0, 1, 2, 3, \ldots) \tag{4.6}$$

In spectroscopy, it is customary to report energies divided by hc. The *vibrational terms* $G(n)$ for a molecule are (4.6) divided by hc:

$$G(n) = \left(n + \frac{1}{2}\right)\tilde{\nu} \qquad (n = 0, 1, 2, 3, \ldots), \tag{4.7}$$

where $\tilde{\nu} = \frac{\omega}{2\pi c}$ is the frequency of the oscillator in cm^{-1}. The ground state eigenfunction is given by:

$$\Psi_0(r) = \left[\frac{1}{\pi r_0^2}\right]^{\frac{1}{4}} e^{\frac{r^2}{2r_0^2}}, \tag{4.8}$$

where r_0 is the turning point for the classical harmonic oscillator:

$$r_0^2 = \frac{\hbar}{\sqrt{k\mu}}. \tag{4.9}$$

Higher states can be calculated from the recursion formula:

$$\Psi_{n+1}(r) = \frac{1}{\sqrt{n+1}}\left[-\sqrt{\frac{r_0^2}{2}}\frac{\partial}{\partial r} + \sqrt{\frac{1}{2r_0^2}}r\right]\Psi_n(r). \tag{4.10}$$

4.2.1 Selection rules for vibrational transitions

The extinction coefficient for the vibrational transition $|i\rangle \rightarrow |f\rangle$ is:

$$\varepsilon_{i\rightarrow f} = |\langle i|\hat{\boldsymbol{\mu}}|f\rangle|^2, \tag{4.11}$$

where $\hat{\boldsymbol{\mu}}$ is the transition dipole moment operator which measures the linear displacement of charge upon transition. The evaluation of (4.11) for a set of

eigenstates enables the identification of allowed transitions between these states. Only transitions with nonzero ε take place. Two important results follow from inserting the eigenstates for the harmonic oscillator into (4.11). First, a vibrational transition can only take place if the dipole moment of the molecule changes during the transition. This prohibits homonuclear diatomic molecules (e.g. N_2) from absorbing IR radiation. On the other hand, a heteronuclear diatomic molecule such as CO can further vibrational absorbtion since its electric dipole moment changes upon stretching. Second, the vibrational states can only change by one level.

4.2.2 Normal modes

For molecules consisting of more than two atoms the atomic displacements corresponding to IR absorption are superpositions of basic vibrations [120]. These composite atomic displacements are known as normal modes. The calculated normal mode coordinates for mode 131 (1873.81 cm^{-1}) of ginkgolide B are illustrated in **Figure 4.3** as vectors. Each of the $3N-6$ normal modes of a nonlinear

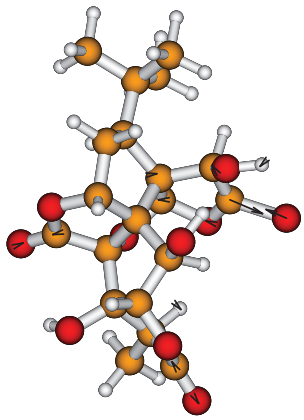


Figure 4.3: Example of a normal mode, normal mode 131 (1873.81 cm^{-1}) calculated for ginkgolide B (structure 1). The vectors illustrate the direction and amplitude of atomic displacements.

molecule can be described as an independent, multi-atom harmonic oscillator [120]. Thus the vibrational terms G_Q for a given normal mode Q can be written as:

$$G_Q(n) = \left(n + \frac{1}{2}\right)\tilde{\nu}_Q, \tag{4.12}$$

where

$$\tilde{\nu}_Q = \frac{1}{2\pi c}\sqrt{\frac{k_Q}{\mu_Q}}. \tag{4.13}$$

Here μ_Q denotes a reduced mass for the atoms participating in the normal mode and k_Q is a characteristic spring constant for the normal mode [120]. A vibrational transition can take place if an the energy of an incoming photon matches the difference between two vibrational terms, that is if:

$$\tilde{\nu}_{photon} = G_Q(n) - G_Q(n-1) = \tilde{\nu}_Q \tag{4.14}$$

4.3 VCD spectroscopy

The absolute configuration (AC) of a molecule is the assignment of the stereochemical configuration at every chiral center. Vibrational circular dichroism (VCD) spectroscopy determines the AC by measuring the difference in the absorption of left- and right-circularly polarized IR radiation associated with all 3N-6 fundamental molecular vibrational transitions [121]:

$$\Delta\varepsilon(\lambda) = \varepsilon_L(\lambda) - \varepsilon_R(\lambda), \tag{4.15}$$

where ε_L and ε_R are the extinction coefficients for left and right-circularly polarized light, respectively, at wavelength λ. Measurement of $\Delta\varepsilon(\lambda)$ over a range of wavelengths gives the VCD spectrum. The rotational strength R is experimentally defined as the integrated intensity of a VCD band:

$$R = \frac{hc}{32\pi^3 N_A}\int \frac{\Delta\varepsilon}{\lambda} d\lambda \tag{4.16}$$

Distinction between enantiomers is possible, because the VCD spectra of two enantiomers of a chiral molecule are of equal intensity and opposite sign. The full potential of VCD spectroscopy depends strongly on the availability of DFT routines for the accurate prediction of VCD spectra [122]. This is evident from the following protocol for the determination of the absolute configuration of a chiral molecule with VCD:

1. The experimental VCD(IR) spectrum is measured
2. Spectra for the two enantiomers are calculated with DFT
3. The experimental VCD spectrum is compared with the two calculated spectra to reveal the enantiomer providing the best agreement with measured signs and intensities.

For flexible molecules, a conformational analysis is performed and conformers populated at the experimental temperature are included in the generation of a *composite* VCD spectrum consisting of the population weighted sum of the spectra calculated for the individual conformers. This is demonstrated in case of two conformers of ginkgolide B in Paper V.
The starting point for the theoretical calculations is the quantum mechanical definition of the vibrational rotational strength associated with the rotational transition from an initial state $|i\rangle$ to a final state $|f\rangle$ [123]:

$$R(i \rightarrow f) = \Im \left[\langle i|\hat{\boldsymbol{\mu}}|f\rangle \cdot \langle f|\hat{\boldsymbol{m}}|i\rangle \right], \tag{4.17}$$

where $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{m}}$ are the electric dipole and magnetic dipole operators respectively. It can be shown that evaluation of the electric dipole transition moment $\langle i|\hat{\boldsymbol{\mu}}|f\rangle$ can be performed readily within the Born-Oppenheimer approximation under the assumption of harmonic vibrations. On the other hand, the evaluation of the magnetic dipole transition moment $\langle f|\hat{\boldsymbol{m}}|i\rangle$ was feasible only after theoretical contributions by Stephens in 1985, including corrections to the Born-Oppenheimer approximation. For details, see [121]. This calculational approach has been included into the Gaussian software program [64], where density functional theory calculations (particularly with Becke hybrid functionals) have shown to be superior to Hartree-Fock calculations.

4.4 GC/MS

4.4.1 Headspace analysis

Headspace analysis can be defined as analyzing a gas in contact with a liquid or a solid sample, and drawing conclusions pertaining to the nature of the sample [124]. It may be useful to think of the headspace analysis as consisting of two steps. The first step involves sampling the headspace of the system under concern. This step could be called extraction. The second step is the actual analysis of the composition of the headspace aliquot. Two methods for extraction were employed in connection with this project: Dynamic headspace and solid-phase microextraction (SPME). Dynamic headspace extraction consists of purging the headspace of a sample with an inert gas. The volatile compounds in the headspace are carried by the purge gas to a trap where they are absorbed on a suitable material. The trapped volatiles are subsequently released from the trap into the inlet of a GC. SPME [125] is a relatively new extraction technique. The analytes are extracted from the relevant matrices by partitioning them into an immobilized stationary phase coated on a silica fiber from where they are subsequently thermally desorbed in the GC injector.

4.4.2 Gas chromatography

If a sample can be vaporized without decomposition, which is the case in many situations relevant to food science, its composition can be studied with gas chromatography (GC) [126]. GC enables the separation of the constituents of a gaseous sample based on the difference in their affinities for a condensed phase preparation which for reasons soon to be apparent is known as the stationary phase. The central component of a GC instrument is a long heated hollow tube known as the column. A sample injected at one end is carried through the column by the unidirectional flow of an inert carrier gas (the mobile phase). The movement of the sample components is selectively retarded by interactions with a thin layer of a non-volatile substance, the stationary phase, coated onto the walls of the column. Due to component specific differences in the interaction with the stationary phase, each sample component leaves (elutes from) the column after a different time period (the retention time) thus providing the desired separation. Any suitable spectroscopic method based on electromagnetic radiation (NMR, IR, UV, etc.) can be employed for the identification of the eluted compounds, but the very high sensitivity of mass spectroscopy (the mass of an ion can be determined to one part per millon) is one important motivation for the ubiquitous use of this method.

4.4.3 Mass spectroscopy

Mass spectroscopy was pioneered by Thomson who determined the mass-to-charge ratio of cathode rays in deflection experiments in 1897 and hence discovered the electron [127]. In later extensions he demonstrated similar deflections of positive ions, and suggested the use of the method for the measurement of molecular weights [120].
In a mass spectrometer positively charged ions are produced from a sample injected into a ionization chamber (ion source) in the instrument. Some of the ions generated may be molecular ions, which is just the molecule with a positive charge, while others are positively charged molecular fragments with a relative abundance ("fingerprint") characteristic of the molecule. The ions are accelerated by an electric field and the beam is passed through a low pressure environment where it encounters the quadrupole mass filter (in case of quadrupole instruments, as in this study). This filter consists of four circular and perfectly parallel metal rods that the ions must pass between along the z-axis in order to reach a detector. A radio frequency voltage is applied between diagonally opposing pairs of the rods, and a direct current voltage is superimposed on the RF voltage. For a given ratio of the voltages, only ions with a certain m/z ratio will have stable trajectories and reach the detector, while other ions will collide with the rods. Thus by sweeping the voltage ratio, a range of m/z values can be scanned, and

a mass spectrum (intensity versus m/z value) can be recorded. Interpretation of the mass spectrum reveals the identity and quantity of a compound.

4.5 Equilibrium dialysis

The goal of this method is to establish a quantitative measure of the affinity of a ligand for an acceptor in an aqueous environment. The words ligand and acceptor sometimes mean a drug and its target protein, but could equally well refer to any situation where a molecule (ligand) is complexed with or entrapped by another molecular structure (acceptor). In the context of food science, the role of the receptor and the ligand could be played by a macromolecular food constituent and a flavor compound, respectively [128]. The experimental dialysis set-up consists of two chambers, the sample and assay chambers, separated by a semi-permeable membrane with a molecular weight cutoff chosen to retain the receptor component of the sample. The dialysis experiment proceeds by intro-

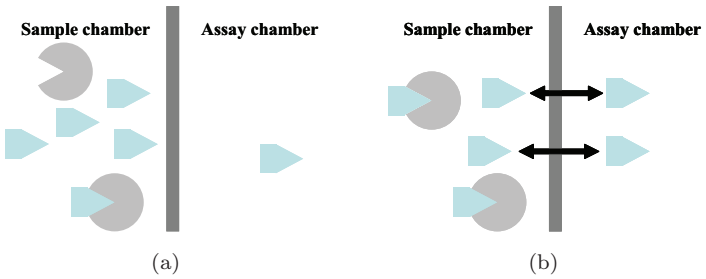


Figure 4.4: Schematic representation of equilibrium dialysis. (a) Ligands (blue) diffuse through the semi-permeable membrane. (b) At equilibrium, there is the same concentration of ligands in solution in both dialysis chambers, while some ligands are bound to receptors (pacmen) in the sample chamber.

ducing a solution of known volume and concentration of receptor, for instance β-glucan, into one of the chambers. A solution of ligand with the same volume and known concentration is placed in the remaining chamber. The ligand diffuses across the membrane until equilibrium is reached, that is to say until the concentrations of free ligand in both dialysis chambers are the same (**Figure 4.4**). In the receptor chamber, however, the receptor has bound some ligand and the total ligand concentration (free + bound) is larger than in the other chamber. The higher the affinity of the interaction, the higher the concentration of ligand in the receptor chamber at equilibrium. Thus by measuring the concentration of the free ligand in the two chambers at equilibrium, the amount of bound com-

pound can be determined. Equilibrium dialysis was employed in Paper III for the determination of the affinity of vanillin-related compounds to a BG matrix.

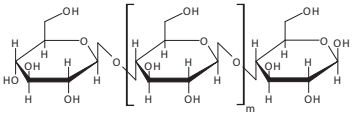
5 Applications

5.1 Molecular dynamics studies of β-glucan motifs

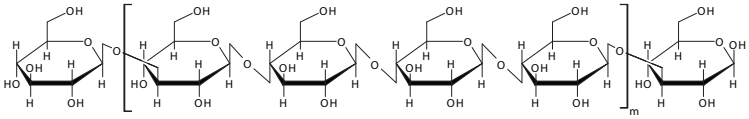
5.1.1 Mixed-linkage (1→3),(1→4)-β-D-glucan

Mixed-linkage (1→3),(1→4)-β-D-glucan (BG) is a linear homopolymer of glucose found in particular abundance in oat, barley and other cereals. Its primary structure can be viewed as a cellulose strand (**Figure 5.1(a)**) interrupted by single β(1→3) linkages, dividing the cellulose strand into cellotriosyl (**Figure 5.1(c)**) and cellotetraosyl (**Figure 5.1(b)**) units [6]. BG is an excellent texturizing agent and its consumption has been associated with positive health effects in man, such as postprandial lowering of glucose and insulin levels and lowering of serum cholesterol [6]. While the effectiveness of postprandial lowering of insulin and glucose levels depends primarily on BG dose and its ability to produce high viscosity at low concentrations, such a relationship has not been established for the cholesterol lowering effect. Nevertheless, the evidence for the positive effects of BG in lowering cholesterol has been strong enough to convince the American Food and Drug Administration to allow a health claim that confirms the positive effects of consuming oat soluble fiber on the reduction of risk of coronary heart disease in humans [129]. BG has also been associated with potentiation of the immune response, although neither the implications nor the mechanisms of this effect have been established [6]. These examples show that BG is a multifaceted biomolecule and that its presence in the human diet may lead to beneficial effects through several routes. However, if effects such as cholesterol binding cannot be accounted for by properties such as viscosity and mass of BG how can they be explained? The 2008 review by Sletmoen and Stokke [130] on the structure of β(1→3)-glucans stated:

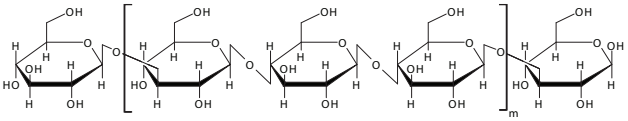
> Despite intensive research efforts, there is ... still no consensus on the basic structural requirements for biological activity, and the higher or-



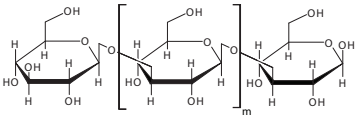
(a) Cellulose.



(b) $\beta(1\rightarrow3)$-linked cellotetraosyl units.



(c) $\beta(1\rightarrow3)$-linked cellotriosyl units.



(d) Curdlan.

Figure 5.1: Primary structure motifs in mixed linkage β-glucans. The repeating unit is shown in brackets.

der structure of the biologically active (1→3)-β-D-glucan component is not well established.

A similar statement applies to the current understanding of BG. Seen in the light of the success of molecular modeling and visualization in the elucidation of structure and mechanisms in the pharmaceutical sciences [131], it is likely that an analogous scheme may confer similar benefits in the study of BG. The simulations of the simplest BG building blocks presented in Paper I could be seen as the first steps towards gaining such a molecular level understanding of the BG polymer. The main findings in that paper will be discussed here.

5.1.2 Disaccharide BG models

Cellobiose contains a $\beta(1\rightarrow4)$ linkage and laminarabiose contains a $\beta(1\rightarrow3)$ linkage, whereby these disaccharides are the smallest building blocks of BG still containing the important glycosidic linkages. We studied the methylated versions of these sugars, methyl β-cellobioside (**Figure 5.2(a)**) and methyl β-laminarabioside (**Figure 5.2(b)**), where the methyl group ensures the anomeric β configuration at the reducing end of the disaccharide. The goal was to pro-

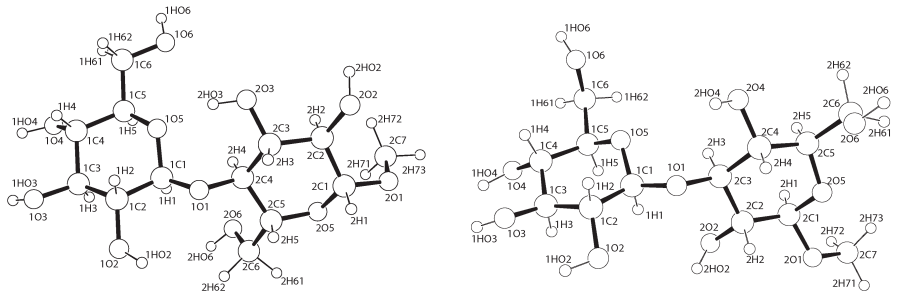


Figure 5.2: The structures of (a) methyl β-cellobioside and (b) methyl β-laminarabioside with the atom labels employed in Paper I.

vide insights into the hydration and dynamic behavior of these fundamental BG building blocks. The resulting knowledge, in addition to being interesting from a fundamental science viewpoint, may also be helpful in forwarding hypotheses pertaining to three-dimensional structure of the BG polymers. Examples of this kind of extrapolation are provided later in this section.
After building initial molecular models for the methylated disaccharides, the adiabatic maps in **Figure 5.4** and **Figure 5.5** were created. As noted in Section 2.1.4, this procedure serves to chart the energy landscape as a function of the

glycosidic dihedral angles Φ and Ψ for the isolated saccharide. In contrast, the MD simulations provide insights into the dynamic behavior including interactions between the saccharide and solvent molecules. The latter type of interaction is highly influential in determining structure of carbohydrates in solution [132]. Structures corresponding to the minima in the adiabatic maps were solvated in water boxes containing approximately 500 water molecules in accordance with the details in Paper I, and MD simulations were initiated for these systems. We felt it was desirable to use the same force field (CSFF, [28]) for both the creation of adiabatic maps and molecular dynamics simulations to exclude discrepancies when comparing results from the two types of calculation. Comparison with the literature, nevertheless, revealed that our adiabatic maps are very similar to maps produced with e.g. the MM3 force field. For instance comparison between our adiabatic map generated with CSFF for methyl β-cellobioside and the MM3(92) generated adiabatic map for cellobiose [133], shows good agreement, see **Figure 5.3**. The adiabatic maps revealed four minima (A, B, C, and D) in case of

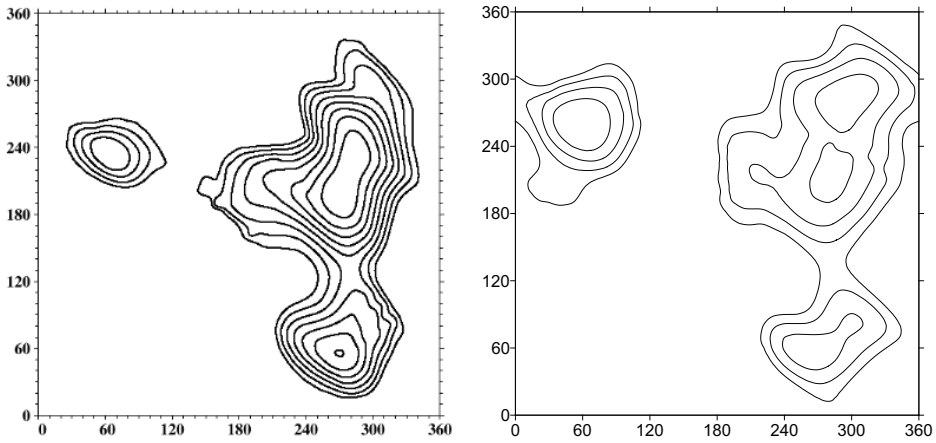


Figure 5.3: Comparison of the adibatic map for cellobiose generated with the MM3(92) force field [133] (left) and our adiabatic map for methyl β-cellobioside generated with the CSFF force field (right).

methyl β-cellobioside (**Figure 5.4**) and three minima (A, B, and C) in case of methyl β-laminarabioside (**Figure 5.5**). The simulations started at the adiabatic minimum A for methyl β-cellobioside and methyl β-laminarabioside were conducted for 50 ns. Contour plots of the two-dimensional histograms for (Φ, Ψ) across the MD trajectories showed that the simulations preferred to populate areas in vicinity of the adiabatic minimum A for both compounds, see **Figure**

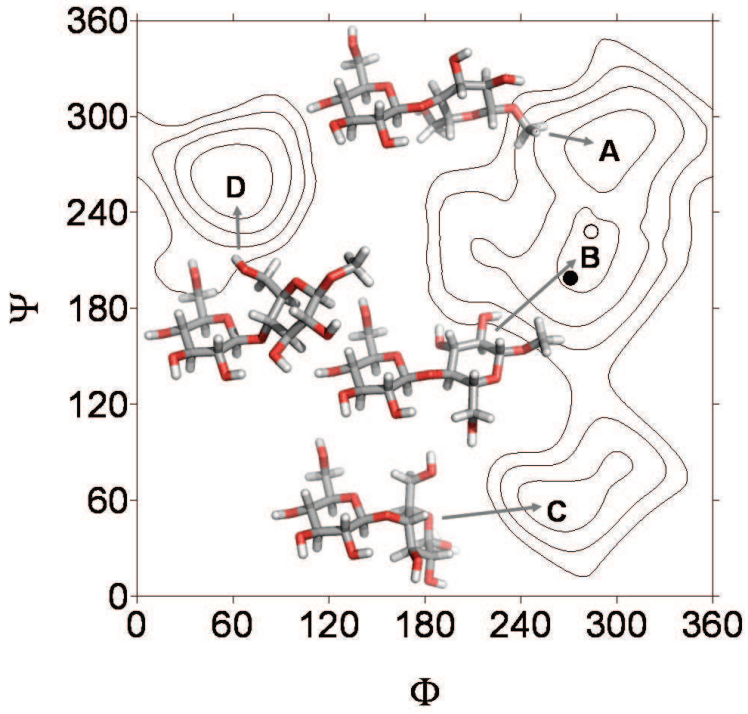


Figure 5.4: Adiabatic map for methyl β-cellobioside. The energy is contoured in intervals of 2 kcal mol^{-1} above the global minimum. Minima are indicated with the uppercase letters A, B, C, and D. The conformers corresponding to the minima are superimposed on the map. Crystal structures of methyl β-cellobioside [134] and cellobiose [135] are indicated with filled and open circles, respectively.

5.6. These simulation minima are denoted by A$'$ to indicate their connection to the adiabatic map A minima. Simulations initiated from the remaining minima in the adiabatic maps were carried out for 10 ns. Simulations started at minimum D in the adiabatic map for methyl β-cellobioside and at minimum B in the adiabatic map for methyl β-laminarabioside populated areas denoted D' and C' close to or at these adiabatic map minima during the entire simulation (10 ns). In contrast, simulations started at the adiabatic map minima B and C in case of methyl β-cellobioside and minimum C for methyl β-laminarabioside shifted from an initial population of these areas of (Φ, Ψ)-space to ultimately populating the corresponding A' minimum of each sugar. The time development of Φ and Ψ for the simulations started at the adiabatic map minimum C for

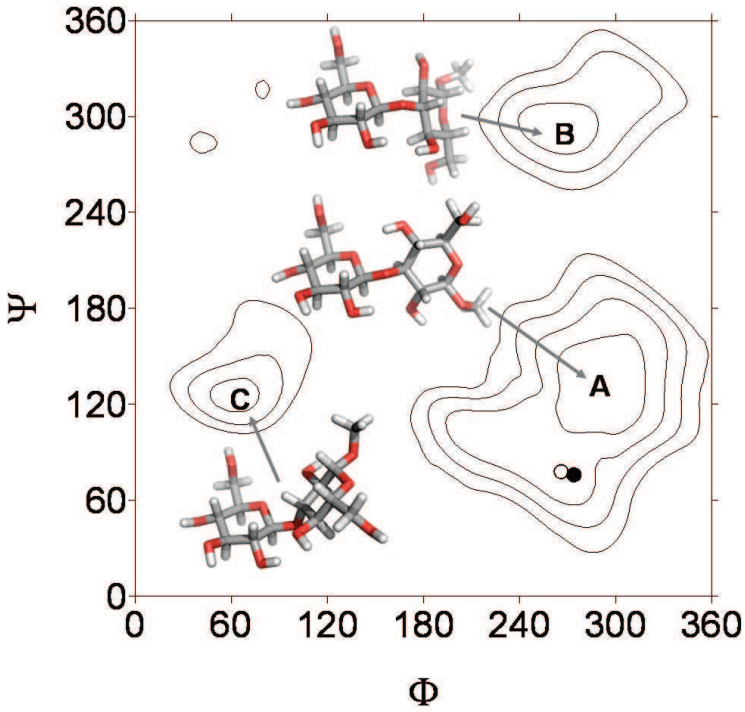


Figure 5.5: Adiabatic map for methyl β-laminarabioside. The energy is contoured in intervals of 2 kcal mol^{-1} above the global minimum. Minima are indicated with the uppercase letters A, B, and C. The conformers corresponding to the minima are superimposed on the map. Crystal structures of methyl β-laminarabioside [136] and laminarabiose [137] are indicated with filled and open circles, respectively.

methyl β-cellobioside and methyl β-laminarabioside is shown in **Figure 5.7(a)** and **Figure 5.7(b)**, respectively. It is noted from **Figure 5.7** that the transition to the preferred A' minimum occurred relatively late in the simulation of both disaccharides, emphasizing the importance of long simulation times in the investigation of even small carbohydrates.

In summary, there are two stable simulation minima for each disaccharide. In light of the observations made above, it could be suspected that the minima stable after 10 ns may also eventually move to the A' minima. This possibility is left for future research.

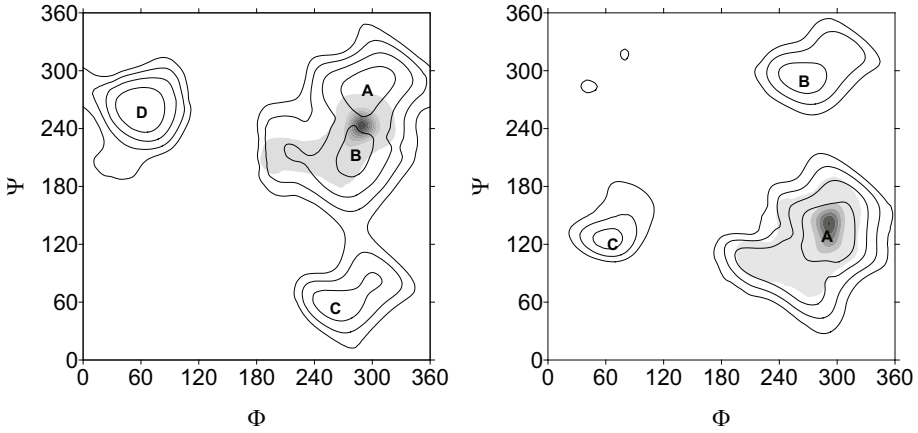
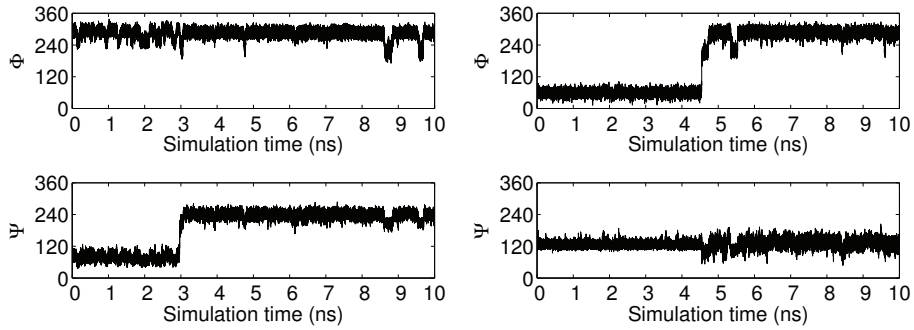


Figure 5.6: Adiabatic maps with population densities in (Φ,Ψ)-space for the 50 ns molecular dynamics simulations started at the adiabatic map minimum A for (a) methyl β-cellobioside and (b) methyl β-laminarabioside.

5.1.3 Comparison of MD and NMR results

To verify the results of the simulations, the two methylated disaccharides were synthesized and subjected to NMR spectroscopy yielding the heteronuclear scalar couplings J_{CH} for the glycosidic dihedral angles and measurements of the Nuclear Overhauser Effect (NOE). A Karplus type relationship provided the link between the measured heteronuclear scalar couplings and the glycosidic dihedral angles. **Table 5.1** summarizes the main parameters from our MD simulations and NMR measurements, and includes additional values from relevant crystal structures and MD simulations by Pereira et al. [138] for comparison. The "Structure" field in **Table 5.1** contains a designation for the system for which values are reported. Our simulations are denoted with a symbol (A, B, C, or D) corresponding to the adiabatic map minimum from where the simulation was started, and the length of the simulation in ns is indicated in parenthesis immediately after the symbol. For convenience these simulations will be referred to as simply the "A simulation" and so forth in the remainder of the text. The average dihedral angles $\langle \Phi, \Psi \rangle$ were calculated across all frames in each MD trajectory as were the $^3J_{\mathrm{CH}}$ couplings. For crystal structures and MD results from the literature, the single pair of glycosidic dihedral angles is reported in the $\langle \Phi, \Psi \rangle$ field. The heteronuclear couplings are in these cases evaluated from the single (Φ, Ψ) point.
From **Table 5.1** it is clear that there is good agreement between $\langle \Phi, \Psi \rangle$ for



(a) (b)

Figure 5.7: Time series for the glycosidic dihedral angles for the 10 ns simulation started at the adiabatic map minimum C for (a) methyl β-cellobioside and (b) methyl β-laminarabioside.

the A simulation of methyl β-cellobioside ($\langle\Phi, \Psi\rangle$=(281, 236)) and the dihedrals measured in crystal structures of methyl β-cellobioside [134] ((Φ, Ψ)=(271, 199)) and particularly cellobiose [135] ((Φ, Ψ)=(284, 228)). The results ($\langle\Phi, \Psi\rangle$=(283, 237)) from the simulations of cellobiose by Pereira et al. [138] are virtually identical with our results.

In case of methyl β-laminarabioside, the agreements between $\langle\Phi, \Psi\rangle$ for the A simulation ($\langle\Phi, \Psi\rangle$=(286, 134)) and the dihedrals measured in crystal structures are best in case of the crystal structures of hydrated curdlan [140] and anhydrous curdlan [139]. The value of Φ for the crystal structures of laminarabiose [137] and methyl β-laminarabioside [136] shows reasonable agreement with $\langle\Phi\rangle$ in our simulations, but both these crystal structures have characteristically lower values of Ψ compared to our value of $\langle\Psi\rangle$. The agreement with the simulation values ($\langle\Phi, \Psi\rangle$=(287, 163) from the study by Pereira et al. [138] is reasonable, although the difference between their $\langle\psi\rangle$ and our value is not negligible. The remaining stable simulation minimum for each disaccharide did not comply with the glycosidic dihedral angles of the crystal structures or MD values from the literature.

Considering the heteronuclear scalar couplings, it is clear that simulation A provides the best match between measured and calculated values of $^3J_{\mathrm{CH}}$ for each disaccharide. In the pairs of $^3J_{\mathrm{CH}}$ calculated from the additional stable simulation minimum, at least one of the values deviated by more than 2 Hz from the corresponding measured value. This evidence from NMR spectroscopy, together with the observations from the crystal structures, strongly suggests that the A' minimum for each disaccharide represents the conformation adopted in solution.

Table 5.1: Calculated (A, B, C, and D) and measured (NMR) heteronuclear $^3J_{CH}$-couplings in Hz. $^3J_{HC}$ and $^3J_{CH}$ refer to $^3J_{H1C4'}$ and $^3J_{C1H4'}$for methyl β-cellobioside (B14), and $^3J_{H1C3'}$ and $^3J_{C1H3'}$ for methyl β-laminarabioside (B13). The mean values of Φ and Ψ for the MD trajectories are reported and values from the crystal structures and relevant simulations from the literature are included for comparison. [a]Cellobiose MD results [138]. [b]Cellobiose crystal structure [135]. [c]Methyl β-cellobioside crystal structure [134]. [d]Laminarabiose MD results [138]. [e]Laminarabiose crystal structure [137]. [f]Methyl β-laminarabioside crystal structure [136]. [g]Anhydrous curdlan crystal structure [139]. [h]Hydrated curdlan crystal structure [140].

Compound	Structure	Torsion	Coupling	
		$<\Phi, \Psi>$	$^3\mathbf{J}_{HC}$	$^3\mathbf{J}_{CH}$
B14	A (50 ns)	281, 236	3.21	5.02
	B (10 ns)	288, 235	3.22	5.03
	C (10 ns)	285, 78	3.14	5.44
	D (10 ns)	58, 240	6.61	5.34
	Cellobiose[a]	283, 237	2.99	5.58
	Cellobiose[b]	284, 228	2.93	5.16
	Cellobioside[c]	271, 199	3.99	3.24
	NMR	-	4.05	4.96
B13	A (50 ns)	286, 134	2.99	4.65
	B (10 ns)	266, 295	4.36	6.58
	C (2 ns)	60, 127	6.65	5.19
	C (10 ns)	181, 129	4.68	4.92
	Laminarabiose[d]	287, 163	2.74	2.83
	Laminarabiose[e]	266, 78	4.46	3.56
	Laminarabioside[f]	274, 76	3.88	3.42
	Curdlan[g]	269, 127	4.34	5.44
	Curdlan[h]	275, 125	3.83	5.50
	NMR	-	3.15	4.12

This allowed us to proceed to study additional details pertaining the simulations initiated at the A minima, assuming that such observations reflected the situation of the actual compounds. This assumption was further strengthened by the second experimental contribution, the NOE restraints from NMR spectroscopy, as will be demonstrated shortly.

5.1.4 Hydrogen bonds and hydration

The crystal structure of methyl β-cellobioside and methyl β-laminarabioside both contain one prominent intra-residue hydrogen bond. In the former case, the bond

occurs between 2HO3 and 1O5, and in the latter case between 2HO4 and 1O5. The A simulations reproduce these bonds, with a duration of 35 percent and 15 percent of the simulation time, respectively. From the mapping of hydration sites of the disaccharides in the simulations in **Figure 5.8**, it is clear that significant water densities (1.8 and 3.0 times the bulk density of water, respectively) occur in each disaccharide between the oxygens involved in these hydrogen bonds. The hydration and hydrogen bonding can only occur when the involved groups are in vicinity of each other, which is exactly the case at the observed $\langle \Phi, \Psi \rangle$ in the A simulations. Thus hydration and hydrogen bonding synergistically determine the structure of these disaccharides in solution.

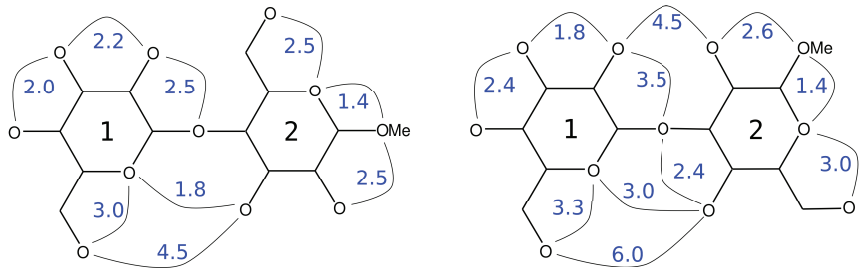


Figure 5.8: Significant water densities found between pairs of oxygens in (a) methyl β-D-cellobioside and (b) methyl β-D-laminarabioside. The analysis was carried out on the 50 ns MD trajectories started in the A minima in the adiabatic maps. Densities are reported as multiples of the bulk density of water. Figures by Peter Ibsen Hansen.

5.1.5 Nuclear Overhauser Effect

Results from measurements of the nuclear Overhauser effect are listed in **Table 5.2**, and from our perspective the most interesting NOEs occur intra-residue, that is to say between the two glycopyranoside rings. NOE restraints are typically classified in the three classes *weak*, *medium* and *strong* [141]. Only NOEs of strength *medium* and *weak* were measured in our experiments, which reflects the generally highly dynamical nature of the systems. Nevertheless, the fact that NOEs are measured in both saccharides between 1H1 and the three axial hydrogens nearest the glycosidic linkage on ring 2, are indications of the rotational constraints of the glycosidic linkage. Such observations, while qualitative in nature, are consistent with the relatively narrow distributions of Φ and Ψ noted in the simulations. One additional detail from the simulations that is compatible

Table 5.2: NOEs measured for methyl β-cellobioside (B14) and methyl β-laminarabioside (B13). The NOEs are characterized as either weak w or medium m. Atom names are in correspondence with **Figure 5.2**.

Atom	Atom	B14	B13
1H1	1H3	m	m
1H1	1H5	w	-
1H1	2H2	-	m
1H1	2H3	m	m
1H1	2H4	m	m
1H1	2H5	w	-
1H1	2H6	m	-
2H1	METH	m	m
2H1	2H3	m	m
2H1	2H4	m	-
2H1	2H5	w	-

with the NMR measurements, is the orientation of the hydroxymethyl group on residue 2 in methyl β-cellobioside. This group adopts the *gg* rotamer for 91 percent of the time, which implies a relatively fixed positioning of one of the 2H6 hydrogens at a distance of 3 Å from 1H1, which supports the experimentally observed *medium* NOE between 1H1 and 2H6. Finally, in both compounds, a NOE of *medium* strength is observed between 2H1 and the methoxy protons. The implied constraint on rotation of the methoxy group agrees well with the sharply peaked distribution of the 2O5-2C1-2O1-2C7 dihedrals observed in the MD trajectories centered at 284°. This orientation places the methoxy carbon close (2.4 Å) to 2C1 in both compounds.

5.1.6 Models of BG polymers

The discussion above showed a convincing agreement between important parameters for the molecular dynamics simulation A and NMR measurements in case of both methyl β-cellobioside and methyl β-laminarabioside. Therefore we decided to construct models of BG polymers for the purpose of visualization, using the values of $\langle \Phi, \Psi \rangle$ from simulation A of methyl β-cellobioside for the $\beta(1\rightarrow4)$-linkage and values of $\langle \Phi, \Psi \rangle$ from simulation A of methyl β-laminarabioside for the $\beta(1\rightarrow3)$-linkage. To model the successive inclusion of $\beta(1\rightarrow3)$-linkages into a cellulose strand, we first constructed the cellulose single helix in **Figure 5.9** A. Changing every fourth linkage to a $\beta(1\rightarrow3)$-linkage produced the markedly helical secondary structure in **Figure 5.9 B**. This structure has nearly six-fold symmetry, as evident when viewed along the first principal axis (left panel).

Changing every third linkage in the cellulose polymer to a $\beta(1\rightarrow3)$-linkage gives the structure in **Figure 5.9 C**, which has a smaller pitch and more turns than the previous model. Finally, the pure $\beta(1\rightarrow3)$-linked polymer (curdlan) is shown in **Figure 5.9 D**. One can envisage several types of interaction between the he-

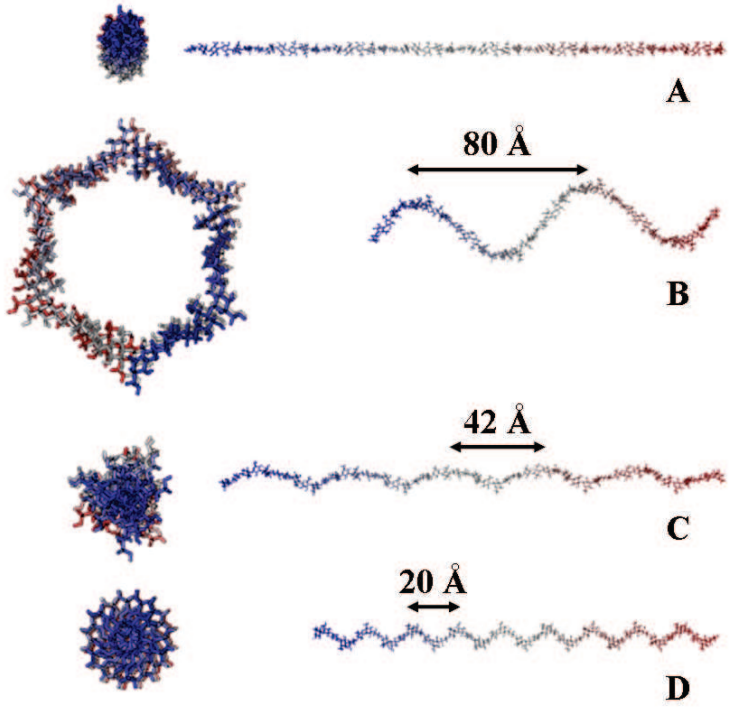


Figure 5.9: Models of β-glucans (48-mers) generated using the average values of glycosidic dihedral angles observed at A simulation minima for methyl β-cellobioside for the $\beta(1\rightarrow4)$-linkage and methyl β-laminarabioside for the $\beta(1\rightarrow3)$-linkage. (A) The cellulose polymer, (B) the $\beta(1\rightarrow 3)$-linked cellotetraosyl polymer, (C) the $\beta(1\rightarrow3)$-linked cellotriosyl polymer, and (D) the curdlan polymer. Right panel: Side view with the major principal axis of each polymer horizontally aligned. Left panel: Top view from the reducing end of each polymer along the major principal axis. The top view of (B) deviates slightly from the view along the major principal axis c, since the present depiction allows for better appreciation of the six-fold symmetry. Approximate values for the helix pitch are indicated. Models within each pane (left or right) are shown on the same scale.

lical BG models in **Figure 5.9** leading to supramolecular assemblies in aqueous environments (e.g. in the bowel) capable of entrapping smaller molecules such as

cholesterol. It is clear, however, that such speculations should merely serve as a starting point for further studies where the ideas are investigated with scientific rigor.

5.2 QSPR models for flavor release

In Paper II we performed dynamic headspace GC/MS measurements of the release into air of 15 strawberry flavoring compounds from aqueous solutions (in the following referred to as "matrices") of the two commercially available BG products PromOat [142] and Glucagel [143]. The flavor release was recorded for BG matrices containing 5, 10 and 15 weight percent of the commercial products. For reference, the flavor release was also measured from an aqueous solution of the flavor compounds, and the following measure of the relative release was defined:

$$A_{\mathrm{rel},i} = \frac{A_{\mathrm{BG},i}}{A_{\mathrm{water},i}}, \tag{5.1}$$

where $A_{\mathrm{BG},i}$ and $A_{\mathrm{water},i}$ are total peak chromatogram areas for compound i from the headspace above the BG matrix and the water solution, respectively. Plotting $A_{\mathrm{rel},i}$ yielded the flavor release profiles of **Figure 5.10**. The preparations and their associated release profiles are referred to as BG-Pxx for PromOat and BG-Gxx for Glucagel, where xx designates the weight percentage of the BG product.

5.2.1 Inspection of flavor release profiles

In **Figure 5.10** the MW for the compounds increases to the right. Disregarding a few compounds, for which consistent detection proved particularly difficult (mainly γ-decalactone and to some extent methyl cinnamate), it is clear from the figure that all preparations of BG products yield release profiles of roughly the same shape. This suggests that the change from one BG matrix to another is not associated with a qualitative change in the flavor release behavior, but rather a modulatory effect. Or in other words, a taster would probably experience a change in taste intensity rather than taste *character* when exchanging one matrix for another. This type of flavor release modulation is consistent with the effects of changes in viscosity and gel strength, as observed in other studies [144, 145]. In general, relatively high standard errors were associated with the measurements, implying that too much significance should not be given to subtle variations in the release profiles. It is clear from the flavor release profiles, that the BG matrices in most cases retain ($A_{\mathrm{rel},i} < 1$) the strawberry flavoring compounds relative to water. In general this retention increases with the concentration of the BG gel

and with the MW of the flavor compounds. The three alcohols show a different behavior than the esters, in that their retention in BG is smaller than for esters with similar MW. In fact, the only situations where compounds are "rejected" from the BG gels with respect to water ($A_{\mathrm{rel},i} > 1$) occur for alcohols. This behavior is indicated in case of *cis*-3-hexenol and 2-hexen-1-ol in the BG-G5 and BG-G10 gels, but unambiguously demonstrated for *cis*-3-hexenol, 2-hexen-1-ol and 1-hexanol in case of the five percent PromOat gel (BG-P5). Although their set-up was different, and direct comparison with our work is therefore not possible, it is interesting that Hansson et al. [144] also noted conspicuous behavior of alcohols in their static headspace study of the release of some of the same compounds from pectin matrices . In their study the release of alcohols, quantified by the air/gel partition coefficient at equilibrium, was not higher than the remaining compounds, but the partition coefficient for the alcohols increased with increasing gel strength while it decreased for esters, aldehydes, ketones, and a sulfur-compound.

Without knowledge of the results of Paper II it would seem reasonable to assume that the flavor release at decreasing BG concentrations becomes more like that of water alone. Disregarding again the highly uncertain γ-decalactone data, the barley (Glucagel) release profiles (**Figure 5.10(b)**) seem to agree to some extent with this idea. At the lowest BG concentration, the BG-G5 release profile displays relatively little variation around its mean and is closer to the line (y=1) than any other profile. In contrast, at the highest BG concentration the BG-G15 profile shows larger variations around its mean and is located far below the unity line y=1 and the other profiles. This simple picture did not hold for the oat BG (PromOat) gels. Although the release behavior by large fulfils (BG-P15 $\leq$ BG-P10 $\leq$ BG-P5), the BG-P5 profile demonstrates very characteristic variations in release and importantly releases alcohols *more* than water does. Thus further studies are required to understand the release behavior at low concentrations of oat β-glucan.
Citronellyl acetate appears to be "anchored" with nearly the same low value of A_{rel} in all gels, perhaps due to its long lipophilic tail. This property is also reflected in the high log P value (4.56) for the compound. This fact is interesting seen in another context, namely when calling the cholesterol lowering effect of BG to mind. Could cholesterol be lodged in a similar way in the BG hydrocolloid structures in the human intestinal lumen? If the lipophilicity is an important predictor, the log P of 8.74 for cholesterol [146] suggests this compound would be retained strongly.

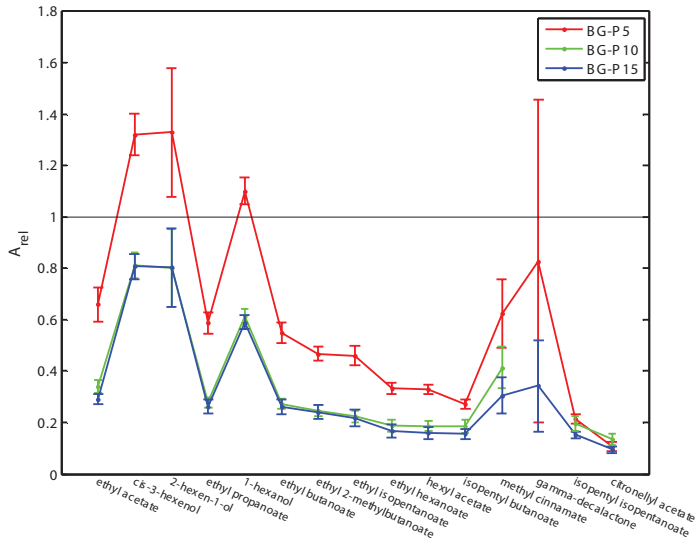
5.2.2 QSPR models

While the variation between flavor release profiles can probably be attributed mainly to viscosity changes, as indicated above, the intra-profile variation is a function of the molecular structure of the flavor compounds. For the explanation of the changes in release from one compound to another, we employed a quantitative structure property relationship approach. The purpose was two-fold. First, it was hoped that the approach would allow for elucidation of the mechanisms involved in the flavor release. Second, the study would serve as a proof-of concept, showing that this particular type of food-relevant system was amenable to the type of methodology widely employed in the pharmaceutical field.
Molecular models for all compounds were built and subjected to conformational analysis where all torsional angles were varied. The lowest conformer was selected for each molecule, and used for the calculation of molecular descriptors as described in Paper II. An important detail of the PLS regression for establishment of QSPR models, was the type of segmented cross-validation employed, where the tree alcohols were placed in their own segment and the remaining compounds (esters) where grouped in segments of three maximizing the intrasegment chemical diversity. For all six flavor release profiles, very simple and robust QSPR models emerged from the PLS approach. They were based on between two and four molecular descriptors, two or three latent variables and had prediction errors (RMSECV) in the range 0.02 to 0.04. As an example, the QSPR regression equation for the remarkable profile BG-P5 was:

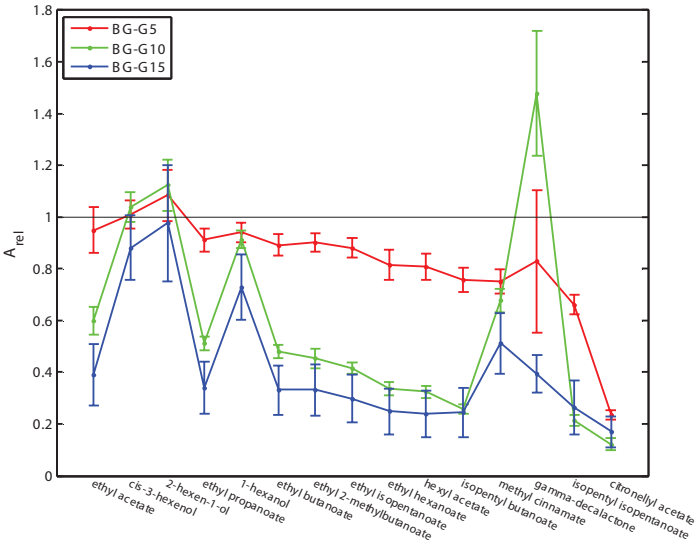
$$\mathrm{A}_{rel} = -0.0155 \cdot \mathrm{ZM1V} + 2.5412 \cdot \mathrm{PCR} + 0.2900 \cdot \mathrm{EEig01r} - 1.5410, \qquad (5.2)$$

where ZM1V is the first Zagreb index by vertex valence degrees, PCR is the ratio of multiple path counts over path counts, and EEig01r is an eigenvalue of the weighted edge adjacency matrix. The origins of the two former molecular descriptors are mentioned in Section 2.3, and as noted these descriptors are measures of molecular complexity (size, symmetry, branching, etc.). The meaning of EEig01r (and other eigenvalues of the edge adjacency matrices) is not understood at present [73]. However, as shown in Paper II, inclusion of EEig01r offers only minute improvements in the RMSECV of the QSPR model; ZM1V and PCR primarily constitute the model. The predicted versus measured plot for this QSPR model is shown in **Figure 5.11**. The quality of the plot is typical for all the QSPR models. The simple QSPR models indicate a correspondingly simple flavor release mechanism. In addition to the PLS based QSPR models, it was found that A_{rel} was highly correlated with many single molecular descriptors. This is documented by the scatter plots of A_{rel} against the most highly correlated molecular descriptors in the Supporting Information for Paper II. The high degree of correlation between molecular descriptors used in the QSPR models and the flavor release profiles is also evident from the correlation matrix in **Figure**

5.12, which graphically depicts the correlation between all the molecular descriptors used for QSPR models in Paper II and the flavor release profiles (shown in brackets). Given the many highly correlated molecular descriptors, one may for practical purposes employ Ockhams's razor and chose more intuitive molecular descriptors than ZM1V or PCR for explaining the flavor release. In the particular case of BG-P5, the predicted water/gas partition coefficient, QPlogPw [75] is a good predictor of A_{rel} as shown in the Supplementary Information for Paper II. The automated multivariate QSPR approach demonstrated here, however, holds exciting possibilities for providing insights into more complex future situations where the flavor release must be described by several variables.



(a)



(b)

Figure 5.10: Flavor release profiles for (a) oat (BG-PXX) and (b) barley (BG-GXX) BG preparations, corresponding to three concentrations (XX=5, 10, or 15 percent weight) of BG product.

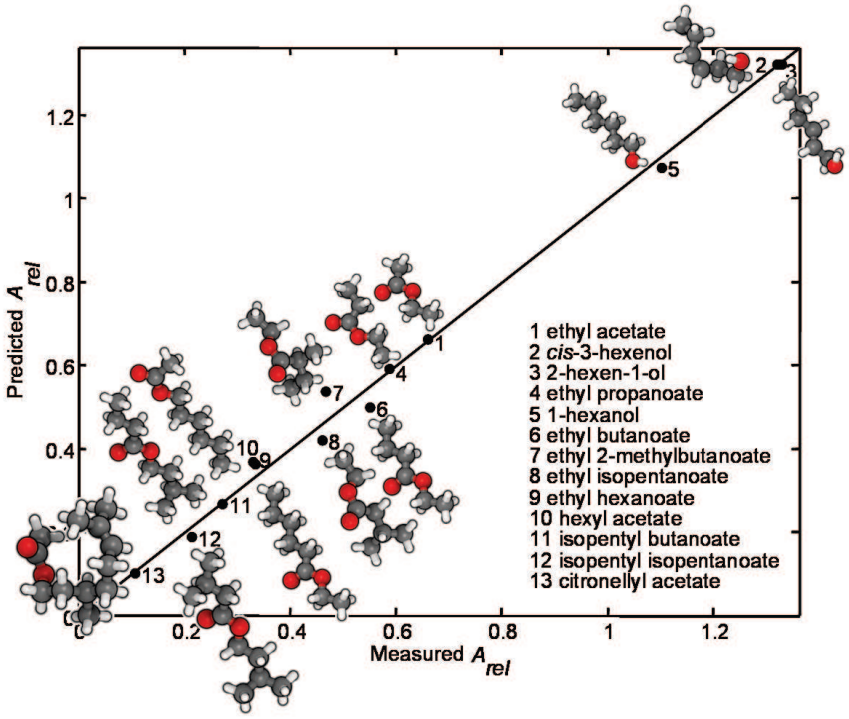


Figure 5.11: Predicted versus measured plot for the QSPR model (Equation 5.2) for flavor release from a 5% PromOat gel.

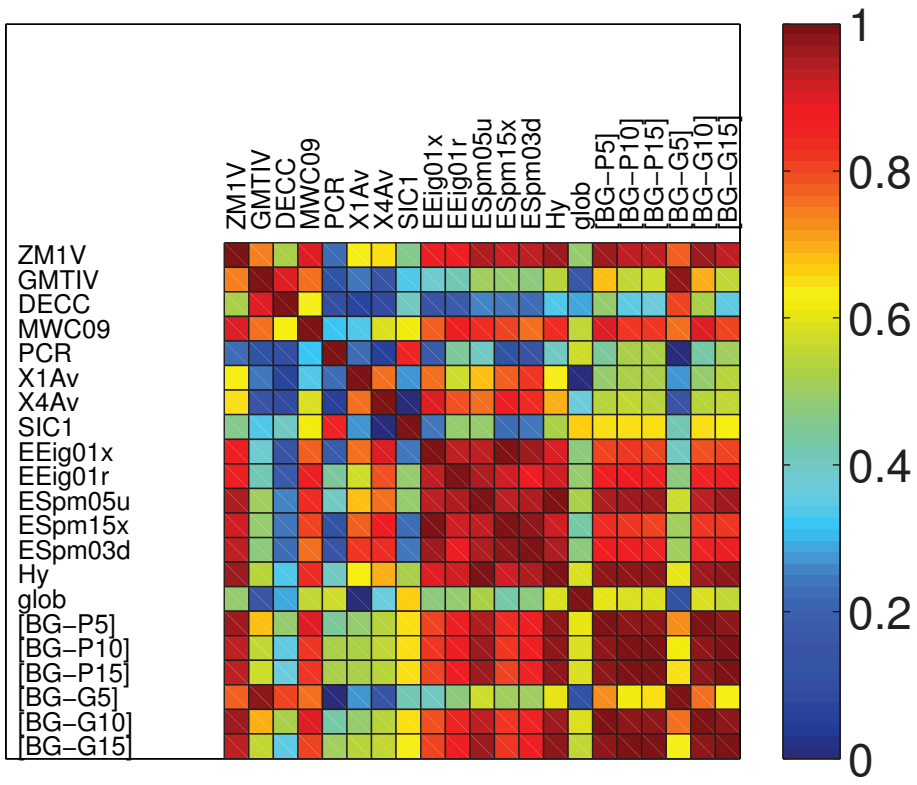


Figure 5.12: Correlation matrix for flavor release profiles (in square brackets) and molecular descriptors used in QSPR models in Paper II. Colors indicate the absolute value of the Pearson correlation coefficient [147] for each variable pair.

5.3 Dialysis

5.3.1 Dialysis experiments

While Paper II focused on the release of flavor compounds from BG matrices into air (liquid-air partitioning), the equilibrium dialysis experiments in Paper III were concerned with the partitioning of compounds between BG matrices and another liquid phase (liquid-liquid partitioning). The dialysis behavior of a series of vanillin-related aromatic compounds, selected glycosides of these, and bile salts was investigated. The discussion presented here pertains to our attempts at establishing a QSPR framework for the understanding of the dialysis results for the vanillin-related compounds. A BG sample and a reference sample for each of some twenty aromatic compounds selected were prepared, see caption of **Figure 5.13**. The reference sample consisted of just one of the aroma compounds of interest in a tricine buffered DMSO/water solution. In addition to these components, the BG sample contained an amount of the commercial barley BG product Glucagel. To test the influence of hardening of the BG gel, identically prepared BG samples were stored for 12 hours ("day 1" samples) and 48 hours ("day 2" samples). The same concentration of aroma compound was used in both reference samples and BG samples. In separate experiments the reference sample and the BG sample from day 1 or day 2 were dialyzed against the same volume of sterile buffer placed in the assay chamber. The concentration of the aroma compound in the assay chamber was monitored by UV absorbance at three wavelengths on a small amount of sample withdrawn from the assay chamber at selected intervals over 5 hours. The measurements were corrected for background absorbance from solvent and from the dialysis of small molecular constituents of the BG products and smoothed using a moving average procedure, as described in the paper. The following function was fitted to the corrected measurements:

$$A_\lambda = A(1 - e^{-kt}), \tag{5.3}$$

were t is dialysis time, k is the dialysis rate constant, and A is the asymptotic value of the function, taken to represent the equilibrium dialysis concentration of aroma compound in solution. The percentage of the compound retained by the BG matrix, ΔA, was calculated as:

$$\Delta A = \left(1 - \frac{A_{BG}}{A_{ref}}\right) \times 100\%, \tag{5.4}$$

where A_{ref} and A_{BG} are the asymptotic values for **Equation 5.3** for the reference sample and the BG sample, respectively.

5.3.2 Dialysis profiles

ΔA for the 27 vanillin inspired compounds is plotted in **Figure 5.13** against the molecular weight of the compounds. It is noted that some compounds have negative values of ΔA, suggesting they favor the BG matrix less than water (in loose terms, they are "rejected" from the BG matrix). On day 1, such behavior is limited to some of the glycosides. The curious effects of matrix ageing are reflected in the slightly increased retention of some compounds on day 2, while other compounds appear to be rejected more. An interpretation of the differences in ΔA in **Figure 5.13** in terms of molecular structure is by no means trivial. There is clearly no simple relationship between MW of the compounds and the retention in the BG matrix. Similar plots where the molecules are sorted according to their lipophilicity, or other seemingly relevant properties were also unrevealing. Furthermore, the dramatically different ΔA values for structurally similar compounds such as e.g. (19) methyl 4-hydroxy-3-methoxybenzoate (day 1 ΔA = 5) and (20) ethyl 4-hydroxy-3-methoxybenzoate (day 1 ΔA = 34) are puzzling.

5.3.3 PCA

In an attempt to explain the variation in ΔA between compounds in terms of molecular structure, more than two-hundred molecular descriptors were calculated for each compound with CDK [74] and QikProp [75] from the MM3* [148, 149] optimized structure of the conformer with lowest energy for that compound. The number of descriptors was reduced to 62 and 15 for day 1 and day 2 data, respectively, by demanding that the molecular descriptors should vary across at least 50% of the samples and that they should correlate to some extend ($r = 0.5$) with ΔA. PCA of the representations of compounds afforded by these descriptor blocks yielded the score plots in **Figure 5.14**. The scores are color coded with the values of the measurements of ΔA from day 1 and day 2, respectively. Three clusters of compounds can be identified in each PCA score plot: (III) glycosides, (II) tert-butyl compounds, and (I) the remaining compounds. It is noted that some compounds with high (yellow) and low (red) values of ΔA have scores on the extreme right and left of PC1, respectively. However, the change in ΔA along PC1 is not systematic. Inspection of the remaining PCs did also not show any evidence for explanation the variation of ΔA in terms of the molecular descriptors. Scatter plots of ΔA against the most highly correlated molecular descriptor are shown in **Figure 5.15(a)** and **Figure 5.15(c)** for day 1 and day 2 data, respectively. In both instances, the most highly correlated descriptor is Wlambda1.un, which belongs to the group of WHIM shape descriptors [150]. For an example of the correlation with a more intuitive molecular descriptor, scatter plots of ΔA against QPlogPC16 (The predicted logarithm of

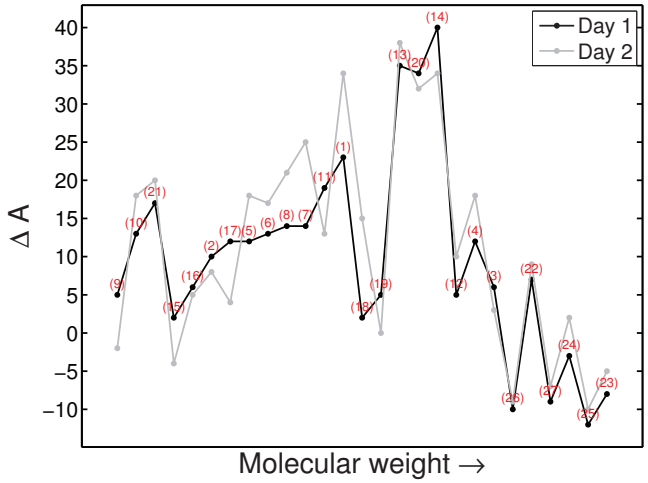
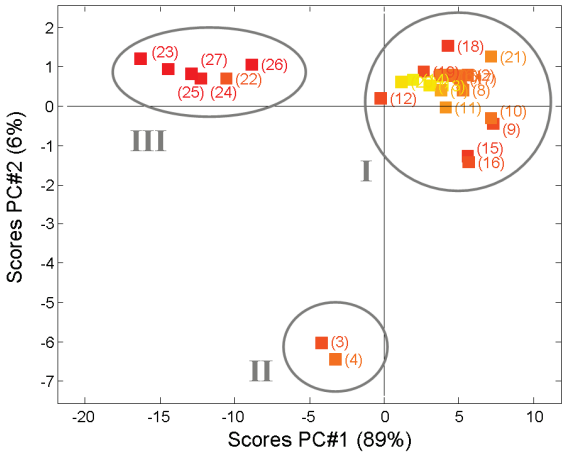


Figure 5.13: ΔA from dialysis experiments with 27 vanillin inspired aroma compounds sorted by MW. (1) 4-hydroxy-3-methoxyacetophenone, (2) 2-hydroxy-4-methoxybenzaldehyde, (3) 2,6-di-tert-butyl-4-methoxyphenol, (4) 2,6-di-tert-butyl-4-methylphenol, (5) 4-hydroxy-3-methoxybenzyl alcohol, (6) 3-hydroxy-4-methoxybenzyl alcohol, (7) 3-ethoxy-4-hydroxybenzaldehyde, (8) 3,5-dimethoxyphenol, (9) 2,3-dimethylphenol, (10) 2,5-dimethylphenol, (11) 2-hydroxy-4-methoxyacetophenone, (12) ethyl 4-ethoxy-2-hydroxybenzoate, (13) 3,5-dimethoxy-4-hydroxybenzaldehyde, (14) 3,5-dimethoxy-4-hydroxybenzoic acid, (15) 2,3,5-trimethylphenol, (16) 2,3,6-trimethylphenol, (17) 4-hydroxy-3-methoxybenzaldehyde, (18) 4-hydroxy-3-methoxybenzoic acid, (19) methyl 4-hydroxy-3-methoxybenzoate, (20) ethyl 4-hydroxy-3-methoxybenzoate, (21) 4-hydroxybenzylalcohol, (22) 4-β-D-glucopyranosyloxy-3-methoxybenzaldehyde, (23) ethyl 4-β-D-glucopyranosyloxy-3-methoxybenzoate, (24) 4-β-D-glucopyranosyloxy-3-methoxyacetophenone, (25) methyl 4-β-D-glucopyranosyloxy-3-methoxybenzoate, (26) 4-β-D-glucopyranosyloxy benzyl alcohol, (27) 3-Ethoxy-4-β-D-glucopyranosyloxybenzaldehyde.

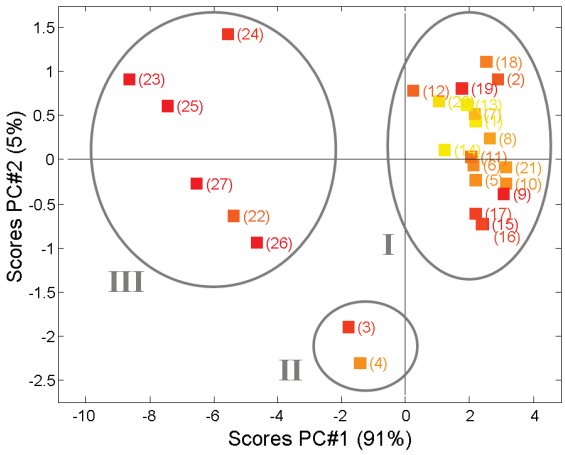
the hexadecane/gas partition coefficient, [75]) are also shown in **Figure 5.15(b)** and **Figure 5.15(d)**. However, in none of the cases does the scatter plot indicate a genuine correlation. The samples (13), (14), and (20) are conspicuous outliers in these and other scatter plots obtained for the remaining descriptors, but the correlation is generally not improved significantly after removal of these compounds.

5.3.4 The failure of QSPR

It quickly transpired that it was not possible to establish PLS models for the prediction of ΔA using the above mentioned descriptor blocks with 62 and 15



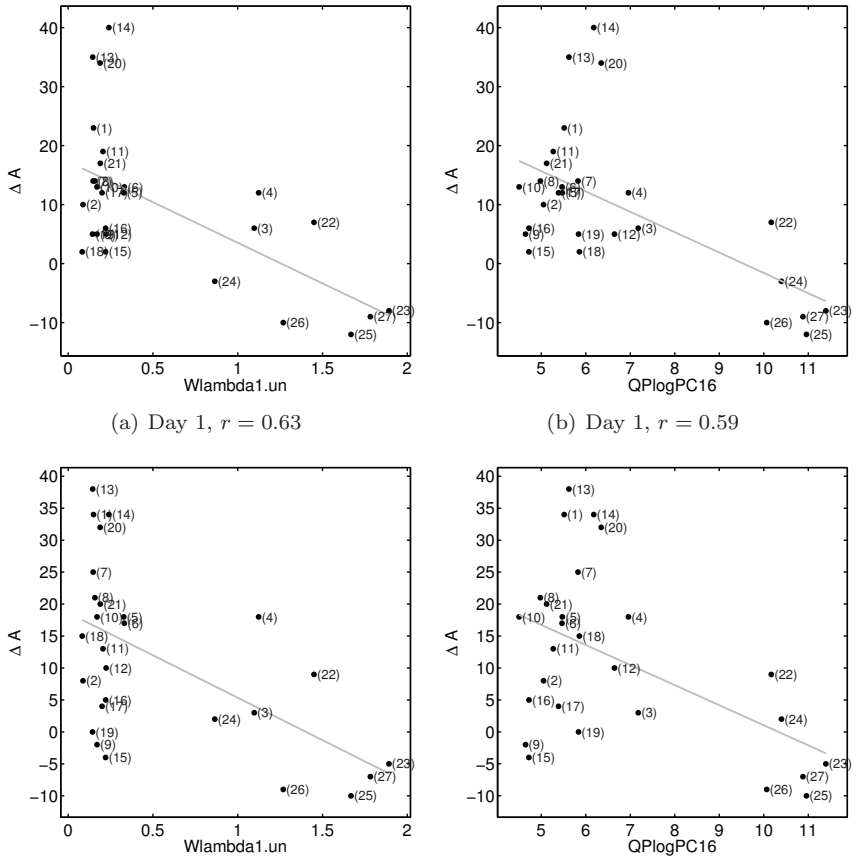
(a) Day 1, 62 molecular descriptors.



(b) Day 2, 15 molecular descriptors.

Figure 5.14: PCA score plots showing the interrelations between 27 compounds represented by selected molecular descriptors, see text. The samples are color coded with red and yellow corresponding to low and high values of ΔA, respectively. The three groups are: (III) glycosides, (II) tert-butyl compounds, and (I) the remaining compounds.

molecular descriptors for day 1 and day 2 data, respectively. Various approaches for variable selection were attempted, manual as well as automated (forward selection) and we experimented with excluding putative sample outliers from the



(a) Day 1, $r = 0.63$ (b) Day 1, $r = 0.59$

(c) Day 2, $r = 0.57$ (d) Day 2, $r = 0.51$

Figure 5.15: Scatter plots for ΔA on day 1 and day 2 against selected correlated descriptors. Sample numbers refer to the caption of **Figure 5.13**. A least-squares fit is shown.

data set, but these efforts were all fruitless. This failure of regression analysis based on molecular descriptors may have several explanations. The relationship between the molecular descriptors and ΔA could be highly non-linear, in which case (ordinary) PLS does not perform well. This possibility could be investigated in future studies by employing variants of PLS such as polynomial PLS, nonlinear PLS [151] or the use of artificial neural networks could be attempted. Another, and perhaps more likely explanation, is that the descriptors chosen are not ade-

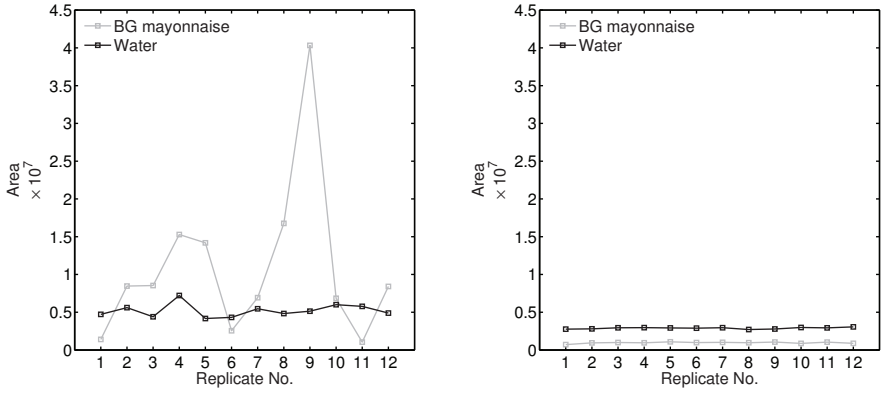
quate for the problem at hand. Alternatively, there may be a problem with the experimental setup or the measurements. The detection problem of vanillin is mentioned elsewhere in this thesis, and it cannot be ruled out that the vanillin-related compounds in the dialysis are problematic in a similar way. This could be tested by performing the dialysis experiments with molecular sets structurally unrelated to vanillin.

A literature survey reveals that the combination of equilibrium dialysis and QSAR or QSPR methodology has been employed only rarely in food science. In one of the few studies reported, Guth and Fritzler [152] achieved good results in their investigation of odorant-biopolymer binding properties using a 3D-QSAR approach. In contrast to our study, however, both the receptors (β-lactoglobulin and bovine serum albumin) and ligands (γ- and -δ-lactones) in that study had well-defined three-dimensional structures. This class of problems is well known from medicinal chemistry and is often tractable within the framework of 3D-QSAR [153]. There is no receptor to model in our study, at least not in the conventional sense. Instead the BG matrix offers a complex and dynamic meshwork with a yet poorly elucidated three dimensional structure.

5.4 Carbohydrate-aromatic interactions

5.4.1 Introduction

Encouraged by the positive outcome of the strawberry flavor release investigations in Paper II, we hoped to employ a similar approach for the study of flavor release of vanillin. Vanillin is one of the world's major flavoring compounds, enjoying a wide consumer acceptance and exhibiting several attractive properties such as antioxidant and bacteriostatic activity [154]. Recently, Lirdprapamongkol et al. [155] demonstrated that vanillin reduces the metastatic potential of human lung cancer cells. As such, there are several good reasons for investigating the potential for enriching BG preparations with the compound. However in our preliminary experiments it turned out to be very problematic to consistently detect vanillin in the headspace with GC/MS methods, despite employing different type of BG matrices and various methodologies for sampling the headspace. The results shown in **Figure 5.16(a)** are typical. The figure shows chromatogram areas obtained from sampling the headspace above 12 replicates of a simple fat replacement system. This system is a vanillin flavored β-glucan "mayonnaise" (BG mayonnaise) prepared from mixing of PromOat, rapeseed oil and water [156] and introducing a small amount of vanillin solution. The sampling/detection method was SPME GC/MS and approximately one sample was measured per hour. The reproducibility of the measurements on the BG mayonnaise is clearly too poor to allow for quantitative determination of the vanillin concentration in the headspace. In contrast, measurements of the internal standard 4-methyl-1-pentanol (**Figure 5.16(b)**) reproduce well. When compounds structurally unrelated to vanillin, such as the strawberry compounds in Paper II, generally showed significantly better reproducibility under similar experimental conditions, it seems reasonable to ask if there might be a unique aspect either to the interaction between vanillin and the measurement apparatus or between vanillin and the BG matrix, or both. In fact a detection problem pertaining to vanillin may be indicated by the small but noticeable fluctuations in measurements on the "Water" samples of vanillin in **Figure 5.16(a)**. Recent publications [157, 158] support the idea of a general detection problem with vanillin. This issue will not be pursued in the following, but instead it will be attempted to provide further insight into the BG-vanillin interaction. To this end, a greatly simplified model system consisting of the absolutely smallest BG building block methyl β-glucopyranoside and vanillin or phenol was studied in solution with NMR spectroscopy. A pragmatic computational chemistry scheme was employed in an attempt to rationalize the experimental findings. Our approach was inspired by previous studies focusing on interactions between simple carbohydrates and benzene, phenol, and aromatic amino acids [159, 160].



(a) (b)

Figure 5.16: Illustration of problems encountered in detecting vanillin in the headspace. The release of (a) vanillin and (b) the reference compound 4-methyl-1-pentanol from 12 replicates of a simple fat replacement system ("BG mayonnaise") and 12 replicates of an aqueous solution ("Water"). "Area" is the area of the peaks in the chromatogram peaks from SPME-GC/MS measurements.

5.4.2 Background

The abundance of aromatic amino acid residues in carbohydrate binding pockets of proteins suggests that aromatic-carbohydrate interactions are important for the binding to saccharides. Indeed, it has been demonstrated through site-directed mutagenesis that these interactions are crucial for the carbohydrate binding in certain proteins [161]. However, insights into the exact nature of the carbohydrate-aromatic interaction have been made only recently. In a series of experiments Morales et al. [162] investigated single strands of DNA linked to a carbohydrate-aromatic system. The study indicated that the carbohydrate and the aromatic ring interact by stacking with stabilizing energies up to 4 kcal mol^{-1}, depending on the type of carbohydrate. Calculations at the MP2/6-311+G(d) level of theory for a wider range of complexes supported these findings, with interaction energies in the range -2.8 to -12.3 kcal mol^{-1} [163]. These results are compatible with a special type of hydrogen bond, the C-H/π interaction [164], occurring between one or more of the sugar ring hydrogens and the delocalized π-system of the aromatic ring.
Prior to the present work the C-H/π interaction has been the subject of a number of computational studies mainly focused on the calculation of accurate geometries and energies. For instance Sujatha et al. [165] investigated the geometry

of the interaction between the aromatic residue and sugars in galactose-binding sites of various proteins at the MP2/6-311++G** level of theory. The most accurate calculations performed to this date have been reported by Tsuzuki et al. [166]. They performed geometry optimization of a benzene-fucose complex at the MP2/6-31G(0.25) level of theory, and evaluated the interaction energy of the complex with CCSD(T) at the estimated basis set limit. Due to limited computational resources, the application of such high levels of theory are beyond the scope of the present project. However, other studies have indicated that more pragmatic methods may provide useful results. The study by Spiwok et al. [163] on 20 carbohydrate-aromatic complexes taken from X-ray structures of glycosidases and carbohydrate-binding proteins showed that several force fields were capable of providing interaction energies correlating strongly with the MP2 values for the carbohydrate-aromatic interactions.
In addition it has been shown that GIAO [72] calculations at the modest HF/6-31G(d,p) level of theory, despite the lack of treatment of electron correlation, are useful in the prediction of relative chemical shift effects in aromatic systems [119, 167–169].

5.4.3 Methods and materials

Sample preparation

Before transfer to 5 mm (o.d.) NMR tubes, all samples were prepared in Eppendorf tubes by addition of appropriate amounts of D_2O stem solutions of phenol, vanillin or methyl β-glucopyranoside. *Mixed* samples of methyl β-glucopyranoside and vanillin or phenol had a final concentration of 10 mM methyl β-glucopyranoside and 40 mM phenol or vanillin, thus giving a pyranose:aromatic ring ratio of 1:4. In addition, a sample containing methyl β-glucopyranoside (15 mM) and phenol (225 mM) was prepared. *Pure* samples contained just the solution of vanillin (40 mM), phenol (40 mM or 225 mM) or methyl β-glucopyranoside (10 or 15 mM). All samples included internal standard (TSP-d4) at a concentration of 0.1 mM. To avoid deprotonation of vanillin, the pH was adjusted to 5 ± 0.5 with the addition of minute quantities (few uL) of 0.05 M HCL or 0.05 M NaOH in H_2O. This adjustment was performed also on samples without vanillin to eliminate pH bias in the ^{1}H spectra. The final total volume in each NMR tube was 550 μL. GC grade vanillin and phenol were purchased from Sigma-Aldrich. Methyl β-glucopyranoside was a gift from Søren Balling Engelsen.

NMR measurements

^{1}H spectra were recorded with a Bruker 500 MHz spectrometer using single pulse experiments (16 scans, D1 = 5 s, acquisition time 1.639 s) at 25°C for the samples

of sugars with and without added vanillin or phenol, respectively. All spectra were referenced to TSP-d4 at 0.0 ppm. The identification of ^{1}H resonances of methyl β-glucopyranoside was done using the assignments by Flugge et al. [170] for the terminal non-reducing glucopyranoside residue in cellotriose. The spectra were recorded using a double tuned BBI probe.

Models of carbohydrate-aromatic complexes

The crystal structure [171] of a *Clostridium Thermocellum* cellulosidase in complex with its cellopentaose substrate (PDB-code 1KWF) was used as the source of initial molecular geometries for modeling the carbohydrate-aromatic interaction. Tyrosine residues interact with the glucopyranoside rings in the bound cellopentaose from alternatingly the α- and β-face, see **Figure 5.17**. Counting from the left, the stacked interaction between tyrosine and the glucopyranoside ring occurs via the β-face of the second glucose residue and via the α-face of the third glucose residue. Both types (α and β-face interaction) of stacked tyrosine-pyranose

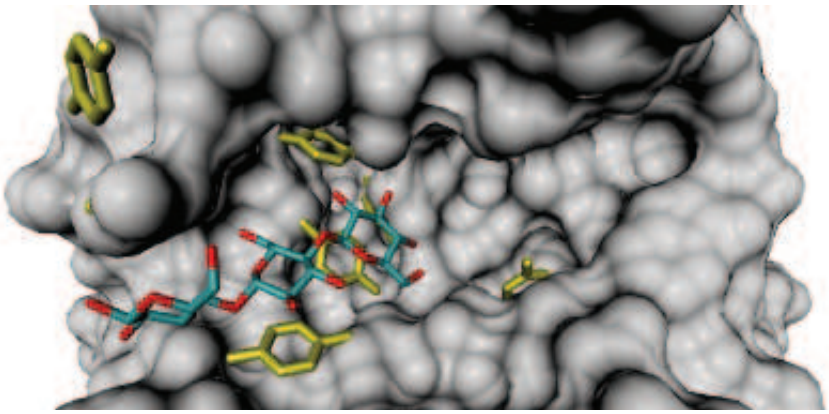
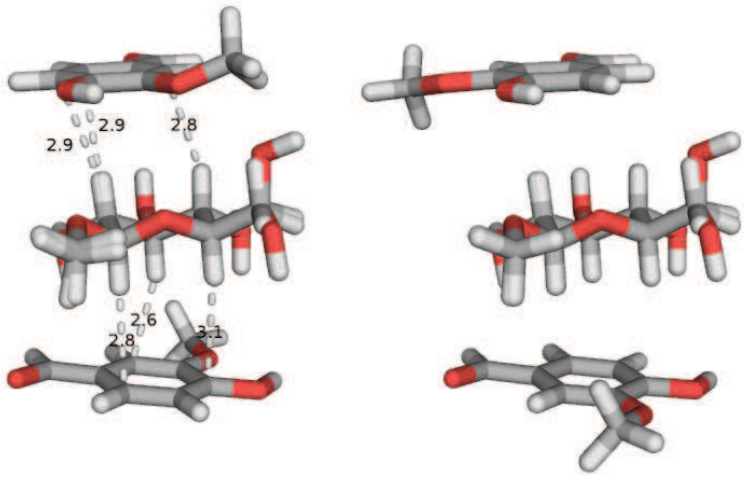


Figure 5.17: View of the carbohydrate binding cleft in the *Clostridium Thermocellum* cellulosidase [171]. A trimeric fragment of the bound cellopentaose is shown, with the reducing end to the left. Tyrosine residues within 10 Å of the substrate are shown in yellow liquorice representation while the bulk of the protein is shown in a gray surface representation.

dimer complexes were extracted from the crystal structure, and modified to model complexes between methyl β-glucopyranoside and phenol or vanillin. Changing tyrosine to phenol is straightforward, but changing tyrosine to vanillin allows for placement of the methoxy group on either side of the hydroxyl group. Structures corresponding to both possibilities (arbitrarily referred to as L or R) were built.

Subsequently, all model complexes were subjected to constrained geometry optimization with the AMBER force field [172–174] in HyperChem [175] where only atoms outside the pyranose or aromatic ring were allowed to move. Only very small adjustments of these atoms occurred during optimization. A model of the uncomplexed methyl β-glucopyranoside was optimized without constraints. The model complexes of methyl β-glucopyranoside and vanillin are shown in **Figure 5.18**. Apart from the aromatic ring substituents, the model complexes of methyl β-glucopyranoside and phenol have identical geometries.



(a) (b)

Figure 5.18: Complexes of vanillin and methyl β-glucopyranoside prepared from the crystal structure [171]. Each figure shows a superposition of two complexes, one complex representing interaction with the β-face (top-face) of the sugar, the other complex representing interaction with the α-face (bottom-face) of the sugar. The superposition was based on the best fit of sugar ring atoms. (a) Superimposed α- and β-face complexes of the L type, see text. Minimum distances (Å) between C-H hydrogens and aromatic ring carbons are indicated. (b) Superimposed α- and β-face complexes of the R type, see text.

Calculation of chemical shifts

Chemical shieldings were calculated with the GIAO method at the HF/6-31G(d,p) level of theory for all complexes and for uncomplexed methyl β-glucopyranoside. Solvent effects were included using a polarizable continuum model (PCM) [176–178] with the default water parameters in G03 [64]. The changes in chemical

shifts of the axial C-H hydrogens on the sugar due to complexation, $\Delta\delta$, were calculated by subtracting the isotropic shieldings of the complexed sugar from the isotropic shieldings of the uncomplexed sugar.

5.4.4 Results and discussion

Measured shift differences

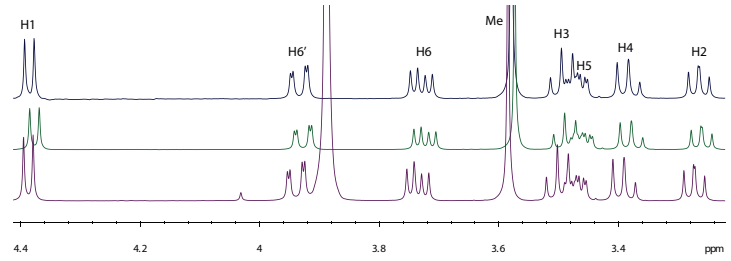
The ^{1}H NMR spectra for methyl β-glucopyranoside with and without aromatic compounds displayed some subtle but interesting differences. The presence of 4 equivalents of phenol caused decrease of all proton shifts of methyl β-glucopyranoside with respect to the pure sugar solution, see **Figure 5.19 a, middle**. On the contrary, in the presence of 4 equivalents of vanillin, all proton shifts in methyl β-glucopyranoside increased relative to the pure sugar solution (**Figure 5.19 a, bottom**). The differences in chemical shifts are summarized in **Table 5.3**. These

Table 5.3: Measured changes in chemical shifts ($\Delta\delta$) in Hz upon addition of vanillin (VAN) and phenol (PHE) to methyl β-glucopyranoside (B).

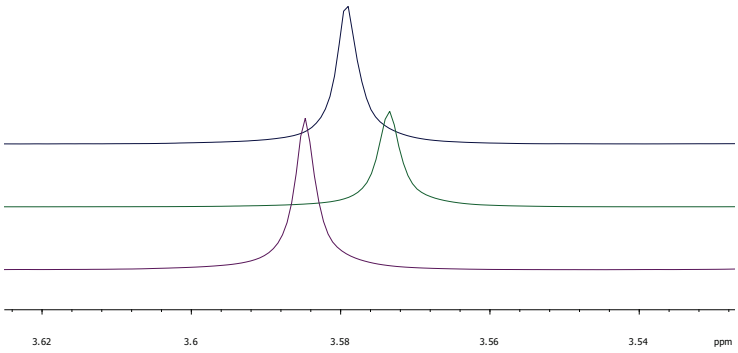
	Ratio	H1	H2	H3	H4	H5	H6A	H6B	OCH3
B:VAN	(1:4)	+1.3	+3.8	+3.3	+3.8	+1.1	+2.6	+3.0	+2.7
B:PHE	(1:4)	-3.7	-2.3	-2.6	-2.3	-4.0	-3.4	-2.6	-2.4
B:PHE	(1:15)	-8.0	+3.6	+1.5	+2.5	-7.1	-1.9	+0.2	-0.1

small but consistent changes imply that all the sugar protons are *deshielded* by vanillin and shielded by phenol. The picture is more complex when there are 15 equivalents of the aromatic compound. This is demonstrated only for phenol, due to the relatively low aqueous solubility of vanillin. In this case, some shifts move upfield and others move downfield. The changes in chemical shifts occurring at these concentrations are not easily interpreted, but could reflect the formation of complexes involving more molecules. It should be noted, however, that these results are in good agreement with values reported by Vandenbussche et al. [160] for the same system. In fact our motivation for including the (1:15) methyl β-glucopyranoside:phenol system in this study was that it allowed this comparison with literature. To maintain simplicity, the remainder of this discussion focuses only on the less ambiguous results from the (1:4) pyranose-to-aromatic ring ratio samples.

The study by Carmen Fernández-Alonso et al. [159] suggests that benzene can form a stacked complex with fucose through C-H/π interactions with the three axial hydrogens on the α-face of fucose. In methyl β-glucopyranoside, all hydrogens are axial, and it seems reasonable to assume that a putative stacking



(a)



(b)

Figure 5.19: 1D ^{1}H NMR spectra obtained from solutions of methyl β-glucopyranoside alone, and in the presence of 4 equivalents of phenol or vanillin. (a) Top: methyl β-glucopyranoside. Middle: Methyl β-glucopyranoside and phenol. Bottom: Methyl β-glucopyranoside and vanillin. (b) Detailed view of the sugar methoxy region.

interaction between methyl β-glucopyranoside and an aromatic ring in solution would occur via the three axial hydrogens (H1, H3, and H5) on the α-face. A second mode of interaction may occur via the two axial hydrogens (H2 and H4) on the β-face. The experimental values of $\Delta\delta$ for the axial α-side hydrogens in B:PHE (**Table 5.3**) are -3.7, -2.6, and -4.0 Hz for H1, H3, and H5, respectively. The smaller values (-2.3 Hz for both H2 and H4) measured for the β-face hydrogens implies a smaller shielding of these protons. These observations might be explained as complexation between phenol and the sugar via α-face stacking due to CH/π interactions. As noted above, the positive values of $\Delta\delta$ for the B:VAN complex implies deshielding of the involved protons. This is puzzling, since the

structure of vanillin at first glance does not seem to differ vastly from other aromatic molecules shown to shield protons of simple sugars in solution [160]. It is possible, however, that the aromatic ring in vanillin may not stack face to face with the pyranose ring. Alternatively, the stacking may occur but for some reason results in deshielding rather than shielding. While the latter hypothesis is the most speculative, it is the easiest to investigate. Rejection of the hypothesis will necessitate much more elaborate theoretical studies, involving the investigation of a plethora of possible geometries for the complex. The theoretical calculations of $\Delta\delta$ discussed in the following, assume that the relative geometrical arrangement of the phenol moiety and the sugar is exactly the same in a vanillin-methyl β-glucopyranoside complex and a phenol-methyl β-glucopyranoside complex.

Calculated chemical shift differences

The results of GIAO calculations of $\Delta\delta$ at the HF/6-31G(d,p) level theory for the model complexes are presented in **Table 5.4**. It is immediately noted that

Table 5.4: Changes in chemical shifts ($\Delta\delta$) in Hz calculated at the HF/6-31G(d,p) level of theory for locally optimized model complexes of methyl β-glucopyranoside with phenol and vanillin, respectively.

Complex	H1	H2	H3	H4	H5
α-face phenol	-0.6	-0.4	-0.6	0.2	-0.9
α-face vanillin-L	-1.2	-0.4	-0.6	0.3	-1.1
α-face vanillin-R	-1.1	-0.4	-0.5	0.3	-1.0
β-face phenol	-0.4	-2.2	-0.4	-0.6	0.0
β-face vanillin-L	-0.4	-2.3	-0.4	-0.8	0.0
β-face vanillin-R	-0.4	-2.3	-0.3	-0.9	0.0

all changes in chemical shifts upon complexation are negative, except in case of H5 in β-face interactions ($\Delta\delta = 0$) and H4 in α-face interactions where $\Delta\delta$ is 0.2 or 0.3. Generally, the changes in shifts for vanillin are even more negative than in case of phenol, which seems to supports the notion that vanillin shields (rather than deshields) the protons of methyl β-glucopyranoside, provided that the geometrical arrangement allows for it.

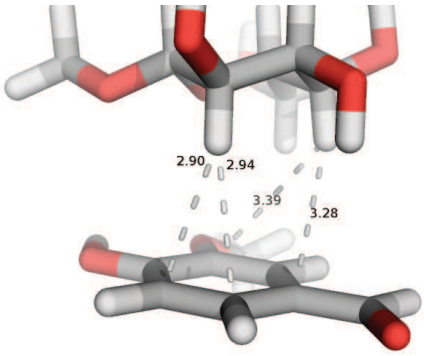
Robustness of the calculations

If the results in **Table 5.4** can be trusted, the stacking interaction between vanillin and methyl β-glucopyranoside does not occur in our experiments. But *can* the results be trusted? After all, the geometries of the complexes are clearly

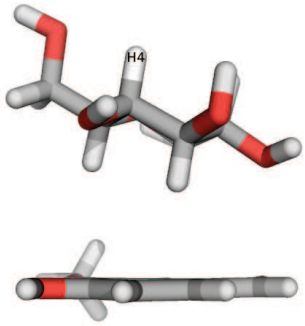
somewhat artificial, since they are constructed by modification of a crystal structure complex and subjected to minimal geometry optimization. The inter ring distance, for instance, remained fixed from the crystal structure. If the calculated changes in shifts are extremely sensitive to variations in geometry, there might exist a narrow range of stacked configurations where the calculations would indeed provide a positive $\Delta\delta$ for vanillin. This possibility should be the subject of future studies, where $\Delta\delta$ is calculated from complexes where the geometrical parameters are systematically varied. In the meantime, several observations seem to support the robustness and consistence of the present calculations. First, the position (L or R) of the methoxy group in vanillin does not exert a significant influence on calculated $\Delta\delta$. This indicates that the method is relatively unaffected by structural changes close to but not directly involved in the CH/π interaction. Although care should be taken not to stretch the evidence, this might partially justify our non-rigorous (i.e. force field based) optimization of groups attached to the rings. Second, there is agreement between the structure of model complexes and the associated calculated shift changes: H1, H3, and H5 in all α-face complexes are more shielded than H2 and H4, in agreement with the fact that H1, H3, and H5 in the model complexes face the π electron system of the aromatic ring, and the two other hydrogens point in the opposite direction. For β-face complexes, H2 is shielded significantly more than the other protons. This seems plausible, since the geometry for this type of complex positions H2 almost directly over the middle of a bond in the aromatic ring, see **Figure 5.20(a)**. The deshielding of H4 in α-face complexes requires more investigation, but may be related to the fact that H4 is the hydrogen farthest from the aromatic ring and/or that it is close to the hydroxymethyl group which has the orientation shown in **Figure 5.20(b)**.

5.4.5 Conclusion

The most important finding in this study is that the pattern of $\Delta\delta$ measured for axial carbohydrate protons in the B:PHE (1:4) sample qualitatively agrees with the values of $\Delta\delta$ calculated for the phenol-methyl β-glucopyranoside α-face complex as well as the vanillin-methyl β-glucopyranoside α-face complex. In contrast, the calculated values of $\Delta\delta$ for model complexes where the sugar interacts through the β-face with phenol/vanillin are *not* compatible with the measured changes in shifts. This supports the notion that phenol interacts with the α-face of methyl β-glucopyranoside in solution and forms a stacked complex. It also suggests that vanillin and the sugar do not form a stacked complex; a geometry yielding calculated values of $\Delta\delta$ consistent with the NMR measurements thus remains to be found.



(a)



(b)

Figure 5.20: Closer inspection of some distances between carbohydrate C-H hydrogens and the aromatic ring. (a) In the aromatic-methyl β-glucopyranoside β-face complex, H2 is positioned above a bond in the aromatic ring with H2-C distances of 2.90 Å and 2.94 Å. (b) In aromatic-methyl β-glucopyranoside α-face complexes, H4 is located far from the ring.

5.5 The influence of geometry optimization on QSAR

If a molecular descriptor depends on the three-dimensional molecular structure, it is natural to assume that the representation provided by that descriptor improves with a more accurate molecular geometry. Taking that assumption a step further, it also seems plausible that QSAR or QSPR models constructed from such descriptors improve with the quality of the three dimensional molecular structures the descriptors were calculated from. The implication is that one should

always choose a high quality method for the geometry optimization whenever using geometrically dependent descriptors. But does this hold in practice? And if it does, how accurate should the geometry optimization method be? Although the link between molecular geometry and quality of the QSPR model has been addressed in some specialized studies (e.g. [179], or [180]) the literature does not report any systematic attempts at clarifying this issue in the case of a "generic" QSAR protocol like the one outlined in Section 3.6. Given that a very large span exists in the computational time required for the various geometry optimization methods (For "typical" drug-sized molecules (30-50 atoms) [181] a molecular mechanics based optimization method may yield the minimized structure in seconds while full-blown quantum chemical methods require hours to days to perform the task, depending on the level of theory.), we felt it was appropriate to address this aspect of QSAR. The details are reported in Paper IV, and the main findings are reviewed here.

5.5.1 Computational approach

We selected three diverse sets of compounds with associated activities/properties from the literature: (**a**) 290 toxicological compounds with acute aqueous toxicity in fathead minnow [182], (**b**) 79 aromatic compounds with penetration of a dimethylsiloxane membrane [183], and (**c**) 12 PPAR-γ agonists with values of pK_i [184]. All structures were built and subjected to a Monte Carlo based conformational search, followed by optimization of the lowest energy conformer using 11 different methods, either semi-empirical, DFT, HF, or force-field based. The differences in molecular structure resulting from the different optimization methods were quantified by the root mean-square distance (RMSD) from the structure optimized at the highest level of theory (B3LYP/6-31G(d,p)), calculated using hydrogen-depleted molecules, see **Figure 5.21**. Although differences in intramolecular hydrogen bonds and close contacts were found for a few compounds across the optimization methods, such variations were in general inconspicuous. Molecular descriptors known to depend on the three-dimensional molecular structure were calculated for the 11 $\times$ three sets. Subsequently PLS models were built for the prediction of the dependent variable. Descriptors with no variation across the 11 different optimization methods were removed prior to PLS modeling, resulting in 293, 464, and 661 descriptors for the (**a**), (**b**) and (**c**) set, respectively. No variable selection was employed in the construction of PLS models, since the goal was not to produce the best QSAR model but rather to detect any differences in the final models due to the molecular optimization method employed.

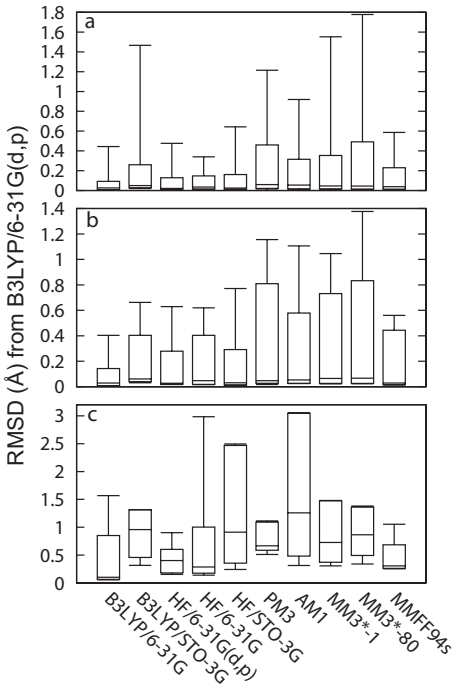


Figure 5.21: Structural variation as a function of geometry optimization method measured by the root mean square distance (RMSD) to the structure optimized at the B3LYP/6-31G(d,p) level of theory. Letters **a**, **b**, and **c** denote the three data sets. Minimum and maximum values are indicated by whiskers. The boxes are vertically limited by the 5 and 95 percentile in **a** and **b** and by the second lowest and second highest value in **c**. Median values are indicated with black bars in the interior of the boxes.

5.5.2 Quality of QSAR/QSPR models

For q^2 the ranges [0.55, 0.57], [0.58, 0.62], and [0.69, 0.75] were found for sets (**a**), (**b**), and (**c**) respectively. These narrow intervals immediately show that the optimization method does not exert an appreciable influence on the final QSAR model. The variations in q^2 across the data sets are shown in **Figure 5.22**, expressed as percentages of the maximum value of q^2. Qualitatively, the variation in q^2 between sets is in accordance with the order of flexibility between sets seen in the RMSD plots (**Figure 5.21**): In general the toxicological compounds are smaller and more rigid (lowest RMSD) than the two remaining molecular sets and also show the smallest variations in q^2 between the optimization methods.

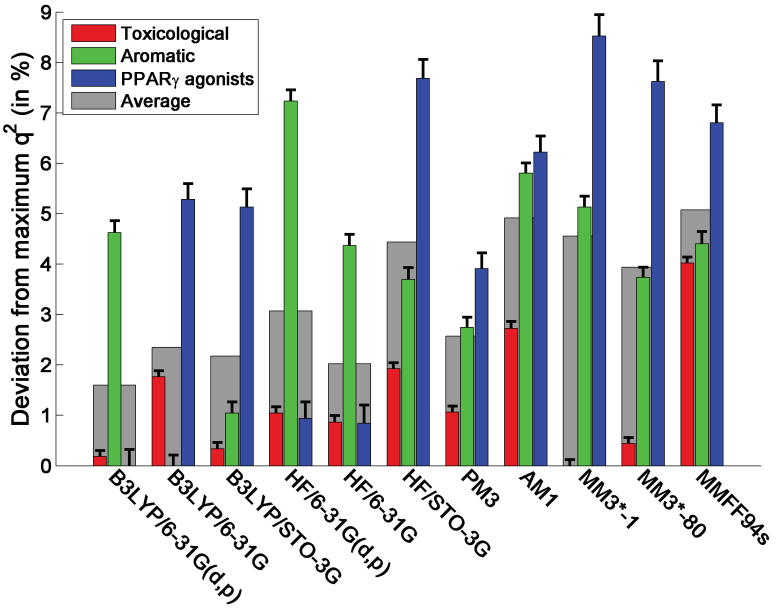


Figure 5.22: Variation in q^2 as a function of geometry optimization method for the three data sets Toxicological (**a**), Aromatic (**b**), and PPAR-γ-agonists (**c**). The variation is given as percentage short of the maximum q^2 (i.e. q^2 for the best method). Figure by Åsmund Rinnan.

On the other hand, the PPAR-γ agonists are highly flexible and show the largest variations in both RMSD and q^2. In comparison, the aromatic compounds show intermediate variations in RMSD as well as q^2. In the consideration of the results from Paper IV there are two reasons to argue that slightly more attention should be given to data set (**c**). First, the PPAR-γ agonists are larger and more flexible than molecules in the two other sets, and hence less likely to be similar to the parametrization sets used in the development of some of the (semi-)empirical optimization methods. They are thus presumably good "probes" of the performance of the optimization methods when these are employed on a molecule distantly related to the ones used in their parametrization (which would be the case for a set of novel compounds). Second, the q^2 values for the models built on this set are well above 0.5, which have been considered to signify QSAR models with predictive power (although this issue is controversial, see e.g. [185]). Other things being equal, it is more attractive to attribute importance to variations in q^2 for the model with most statistical significance. In other circumstances, one could

object to conclusions based on data set (**c**) due to its low samples to variable ratio (12:661), but the inherent resistance of PLS to colinearity and noise should reduce such concerns.

5.5.3 Conclusion

Whether considering the flexible PPAR-γ data set or one of the remaining sets, the influence of the geometry optimization method on the final QSAR model is small. Furthermore, the small variations in q^2 are not intuitively linked to the choice of geometry optimization method. For example, while the highest level of theory, B3LYP/6-31G(d,p), in case of the PPAR-γ agonists does produce a lower q^2 than the other methods, it is not clear why HF/6-31G on the same set performs better than B3LYP/6-31G.

5.6 VCD and DFT studies of ginkgolide B

5.6.1 The relevance of chirality

The thalidomide tragedy [186] was a particularly horrific demonstration of the impact of chirality in man; inversion of the only chiral center in thalidomide toggles between the sedative (R)-enantiomer and the teratogenic (S)-enantiomer [123]. Although the interchange of two enantiomers rarely has consequences as grave as those observed for thalidomide, knowledge of the stereochemical configuration of molecules in food can be valuable in such contexts as understanding flavor perception and detecting food adulteration [187]. Furthermore, since the future will presumably see a significant degree of overlap between food research and biomedicine, the determination of chirality in biomolecules with many stereocenters may become increasingly important. In Paper V we reported the application of VCD spectroscopy and DFT calculations in the determination of the absolute configuration of a natural product with 11 chiral centers: Ginkgolide B. The following is an overview of that work.

5.6.2 *Ginkgo biloba*

The *Ginkgo biloba* tree is unique in several ways. For instance, the tree has been left remarkably untouched by 200 million years of evolution and is hence regarded as a living fossil. Furthermore, *Ginkgo biloba* has an impressive history of over 2000 years of use in traditional Chinese medicine as well as cuisine [188]. The ginkgo leaf contains many compounds with potential biological activity including flavonol and flavone glycosides, diterpene lactones, ginkgolides, sesquiterpenes, iron-based superoxide dismutase, p-hydroxybenzoic acid, ascorbic acid, catechin

and the potentially toxic ginkgolic acids [189]. While the health claims made by modern manufacturers of *Ginkgo biloba* products are numerous and diverse, their validity and connection to the chemical constituents in many cases await verification. It is well-established, however, that the ginkgolides which constitute some 5 percent of the standardized *Ginkgo biloba* extract Egb761, are antagonists of the platelet activating factor (PAF) receptor [190] and thus posses anti-thrombotic properties. The general structure of ginkgolides is shown in **Figure 5.23**. Seen

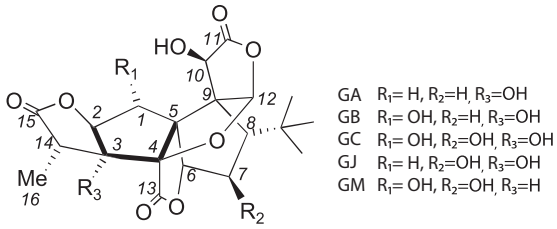


Figure 5.23: Structure of the ginkgolides.

in the light of the recurring carbohydrate theme of this thesis, it is interesting to note that it has been attempted to classify the ginkgolides as carbohydrates [191], although they are typically described more revealingly as terpene trilactones. Ginkgolide B (GB) is the most potent PAF antagonist, and the demonstration of VCD and DFT calculations in the determination of the stereochemical configuration at each of its 11 chiral centers was the subject of Paper V. Although He et al. [192] recently employed a VCD/DFT combination for their complexation studies of GB and other compounds, they focused on the 1795-1826 cm^{-1} range of the IR spectrum and did not provide a thorough investigation of the fingerprint region. The only other reported application of DFT and spectroscopy to Ginkgolide B is the study by Zhu et al. [193] which presented an experimental IR spectrum of rather low resolution and its DFT calculated counterpart.

5.6.3 Measurement of VCD spectra

In brief, the VCD and IR spectra for GB were recorded in three different media (DMSO-d_6, CD_3CN, and KBr pellets) since no single medium combines high solubility of GB with a wide transparency range. The three measured VCD and IR spectra are shown in the upper and lower panel, respectively, of **Figure 5.24**. The spectra display some effects due to media, such as the limited transparency range in case of DMSO-d_6, and the jagged appearance of the KBr spectrum, but are otherwise in good internal agreement.

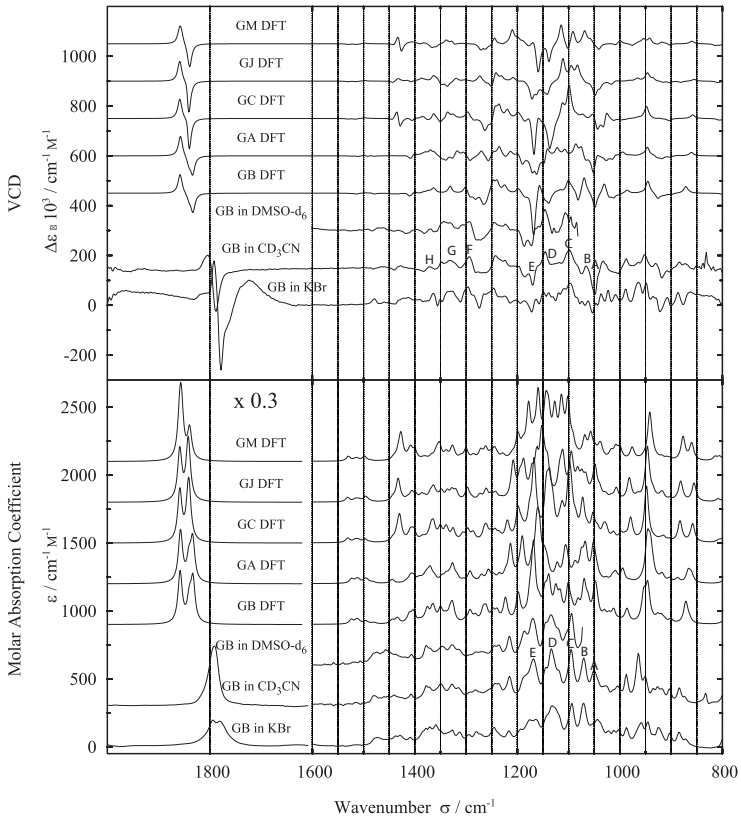


Figure 5.24: Comparison of experimental IR and VCD spectra for GB with calculated (DFT) spectra for GB, GA, GC, GJ, and GM. Prominent peaks are indicated with the letters A to H.

5.6.4 Calculation of VCD spectra

The initial structures for the ginkgolides were generated from published crystal structures. After building the structures of GB, its two stereoisomers GBC_1i and $GBC_{10}i$ with inversions at C1 and C10, respectively and the additional ginkgolides GA, GC, GJ and GM, a Monte Carlo based conformational search was carried out for each molecule followed by DFT (B3LYP/6-31G(d)) optimization of the conformers within 20 kcal mol^{-1} of the global minimum. The relative DFT energies showed large differences (>17 kJ mol^{-1}) between the lowest and next-lowest conformer for all ginkgolides except GB, GC, and GM. For the latter three ginkgolides, two conformers (the lowest and next lowest) were separated by

roughly 8 kcal mol^{-1}, which suggests both are populated at room temperature and hence may contribute to the measured VCD and IR spectra. The difference between the two conformers is the interchange of donor/acceptor roles in a hydrogen bond, as shown in **Figure 5.25** in the case of GB. In accordance with

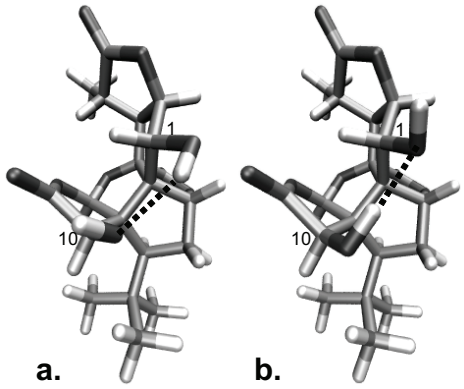


Figure 5.25: The relevant conformers of ginkgolide B at room temperature. The energy of conformer **b** is 7 kJ mol^{-1} higher than conformer **a**.

these observations, VCD and IR spectra were calculated at the B3LYP/6-31G(d) level of theory for the first conformer of GA, GJ, GBC_1i and $GBC_{10}i$, and for the two first conformers for GB, GC and GM. All calculated spectra were scaled with a constant based on previous experiences, due to the systematic error inherent to the DFT calculations. Subsequently, the spectrum for each of the two conformers of GB, GC and GM was weighted with the Boltzmann factor based on the B3LYP/6-311+G(2d,2p) Gibbs free energy for that conformer and the two spectra were added. These composite GB, GC, and GM spectra are shown in **Figure 5.25** together with the spectra for GA, GJ, GBC_1i and $GBC_{10}i$ calculated for the single lowest conformer.

5.6.5 Comparison

It is clear from comparison of the calculated and measured VCD spectra (**Figure 5.24**, upper panel) and IR spectra (**Figure 5.24**, lower panel), that the agreement between calculated and measured spectra for GB in general is excellent in the fingerprint region (850-1300 cm^{-1}). The characteristic peaks A (1045 cm^{-1}), B (1065 cm^{-1}), C (1098 cm^{-1}), D (1134 cm^{-1}), and E (1170 cm^{-1}) have been indicated on the experimental VCD and IR spectra for GB in CD_3CN in **Figure 5.24**, and these peaks are clearly reproduced in the calculated spectra for GB. In addition, the three peaks F (1300 cm^{-1}), G (1330 cm^{-1}), and H (1370 cm^{-1}) are

seen in both the calculated and measured VCD spectrum for GB. Importantly, the patterns of signs of the VCD signals are conserved between measured and calculated spectra for GB, which supports that the absolute configuration of the structure used in the DFT calculation of spectra is identical to the absolute configuration of the actual GB molecule in solution.
To test the influence of the inversion of a single chiral center on the calculated VCD spectra, we compared the calculated spectra for GB with the calculated spectra for its two stereoisomers GBC_1i and $GBC_{10}i$. These stereoisomers are of importance in the synthesis of ginkgolides where addition occurs at double bonds at these sites. The resulting changes in the VCD spectra are clearly noticeable, as evident from **Figure 5.26**.

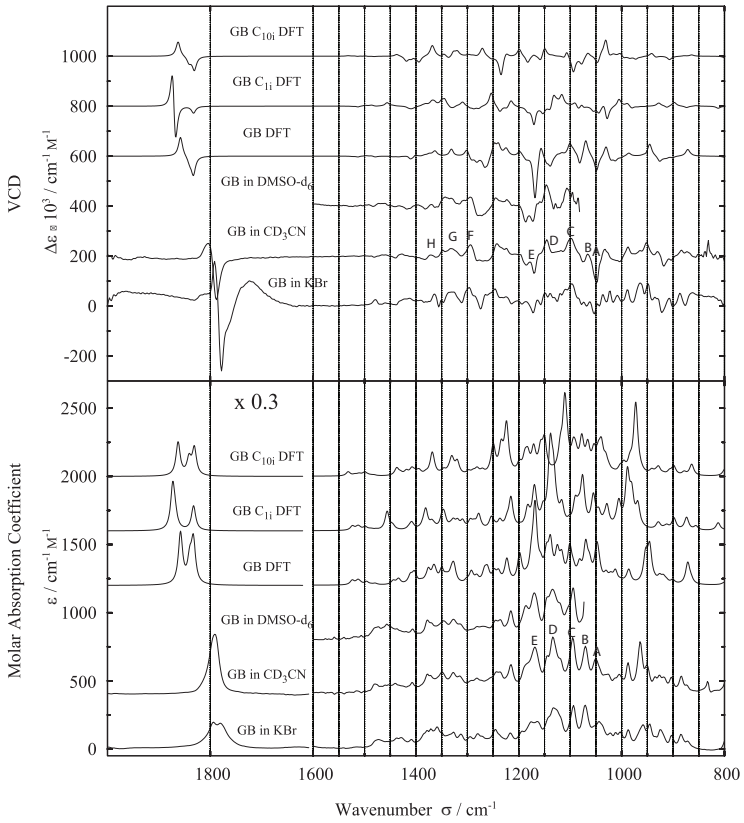


Figure 5.26: Comparison of experimental IR and VCD spectra for GB with calculated (DFT) spectra for GB, GB-C_1i and GB-C_{10}i. Prominent peaks are indicated with the letters A to H.

6 Thesis outlook

6.1 BG building blocks

The powerful analytical tool resulting from combining NMR spectroscopy with molecular dynamics simulations was demonstrated convincingly in the solution studies of the BG building blocks methyl β-cellobioside and methyl β-laminarabioside in Paper I. For the elucidation of structure and dynamics of the BG polymer, it is unlikely that much more can be learned from the disaccharides. Hence, the natural extension of the study is to perform similar combined spectroscopic and molecular modeling studies on increasingly larger BG building blocks. The larger building blocks should preferentially reflect the distribution of $\beta(1\rightarrow3)$ and $\beta(1\rightarrow4)$ linkages observed in degradation studies of BG [194, 195]. Clearly the acquisition of atomic resolution information from such oligosaccharides with NMR spectroscopy presents a challenge, due to the poor chemical shift dispersion and the degenerate structure of the oligomer [196]. Nevertheless, it is possible that information such as NOEs and heteronuclear scalar couplings may be obtainable at sufficiently high field.
Purely theoretical studies may also provide some insights. For instance explicit solvent MD simulations of the mixed-linkage BG polymer models in **Figure 5.9**, both as single polymers and in various putative polymer-polymer and polymer-small molecule configurations, may suggest stabilities of the interactions and aid in the formulation of hypotheses pertaining to the quaternary structure and functionality of BG.

6.2 Carbohydrate-aromatic interactions

Despite the qualitative nature of the results obtained in Section 5.4 with the employed methodology (borderline-questionable geometries, low level of theory for the calculation of chemical shifts), the results suggest that more accurate cal-

culations of geometries for the complexes and relative chemical shifts may yield further insights into the vanillin-carbohydrate interaction. An obvious extension is to investigate the stability of stacked vanillin-methyl β-glucopyranoside complexes at a high level of theory, for instance MP2 in combination with a large correlation consistent basis set. One could also further explore the potential of the pragmatic approach outlined in Section 5.4, by performing shift calculations for the vanillin molecule placed at various points in a grid surrounding the carbohydrate. Such a procedure should reveal geometries where $\Delta\delta$ is positive (as observed in the experimental data). Such a search, however, presents a significant computational barrier and may be guided by the preliminary construction of a map of the shielding and deshielding zones of vanillin using a simple probe such as e.g. methane.
Apart from the obvious calculational improvements, the study in Section 5.4 can also be extended experimentally in several ways. First, the stoichiometry of putative complexes might be revealed from a series of NMR measurements on solutions with different carbohydrate:aromatic compound ratios. Second, the experiments could be repeated using structural intermediates between phenol and vanillin in order to highlight the structural features causing the transition from shielding to deshielding of the sugar protons. With enough analogues, it may even be possible to construct a type of QSPR model for the explanation of shift changes in terms of structure.
Finally, the NMR experiments may be extended for the measurement of NOEs and ^{13}C shifts which might provide additional information on the structure of complexes.

6.3 QSPR/QSAR

6.3.1 Flavor release

Paper II introduced QSPR for the first time in the investigation of flavor release from BG-based matrices, and proved the methodology to be a viable option for studying this type of system. Nevertheless, a QSPR model based on 13 relatively similar compounds does not hold much promise for the prediction of flavor release for novel compounds. Also the insights in the mechanisms of flavor release are hampered by principally two factors: First the lack of chemical diversity due to using only two classes of compounds (alcohols and esters) and second the presence of a significant fraction of non-BG material (65 and 13 percent for PromOat and Glucagel, respectively) in the commercial BG products used. To provide a clear cut characterization of the role of the BG polymer on flavor release, future studies should focus on systems based on BG products with high purity.

Some compounds could not be detected, or were inconsistently detected, in the headspace above aqueous solutions or the BG matrices. This does not mean that the compounds are not present nor that their release is not modulated by the BG matrix. In particular the human nose, which is arguably more sensitive than any detector [197], may well be capable of detecting these compounds and the subtle influence of the BG matrix on their release may be reflected in the perception of aroma. Hence the dynamic headspace GC/MS method employed in Paper II will benefit from efforts directed at lowering the detection threshold for the problematic compounds.

6.3.2 Dialysis

In case of the dialysis experiments in Paper III, QSPR failed to describe the underlying phenomena. This failure of one particular QSPR approach does of course not imply that QSPR per se is unsuitable for the explanation of results from this type of experimental setup. Instead, the outcome of the study should inspire the exploration of other types of molecular descriptors or different kinds of (nonlinear) relationships between the molecular descriptors and the property of interest (ΔA). In addition, parameters pertaining to the experimental setup should be investigated, as equilibrium dialysis is novel in the study of BG matrices. As in case of Paper II, future studies will benefit from the employment of BG matrices of higher purity. In addition, a different set of small-molecule ligands could be tried since vanillin and its various structural analogues exhibited problematic behavior in other contexts of this work (see e.g. Section 5.4.1). The set of alkaloids used in early equilibrium dialysis studies on hydrocolloids [198] might be a good starting point.

6.3.3 Influence of geometry optimization

We believe that the study of the relationship between geometry optimization method and influence on the final prediction model in Paper IV is the first of its kind within the framework of a "generic" QSAR/QSPR approach, where models are constructed from PLS regression to a activity/property using large blocks of molecular descriptors obtained from dedicated applications. The study can be extended in a number of ways, most obviously by the addition of more geometry optimization methods, but also by including additional data sets from a variety of areas. In particular, it will be of importance to focus on data sets from where QSAR/QSPR models with a reasonable quality (e.g. $q^2 >> 0.5$) can be constructed.

6.4 VCD and DFT studies of ginkgolide B

The VCD/DFT study of ginkgolide B in Paper V provided the second demonstration in this thesis of a strong combination of spectroscopy and molecular modeling. Although the calculated and experimental VCD spectra for GB were strikingly similar, that observation alone does strictly speaking not imply that calculated spectra for the other ginkgolides (including variants of ginkgolide B with inversions at single chiral centers) closely resemble the corresponding experimental spectra. Hence an obvious extension of the study would be to record VCD spectra for the additional ginkgolides and perform the comparison with the calculated spectra.

7 Conclusions

The following conclusions can be drawn from this work:

Structure

BG: The combined evidence from MD simulations and NMR spectroscopy (Section 5.1.1) suggested that the glycosidic dihedral angles for the two BG motifs methyl β-cellobioside and methyl β-laminarabioside in solution predominantly populate areas around (Φ=280°, Ψ=240°) and (Φ=286°, Ψ=134°), respectively. Measured NOEs agreed qualitatively with hydration maps and hydrogen bonds suggested by the simulations, and underline the importance of the combined effect of these factors in determining the solution structure of the disaccharides.
Plausible three-dimensional models of BG polymers were created using putative primary structure sequences and the average simulation values of $\langle\Phi, \Psi\rangle$ from methyl β-cellobioside and methyl β-laminarabioside for the $\beta(1\rightarrow4)$ and $\beta(1\rightarrow3)$ linkages, respectively. The models had well-defined secondary structure and indicated the possibility of packing of polymers.

Combined NMR spectroscopy and pragmatic quantum chemical calculations (Section 5.4.1) supported the existence in solution of a stacked complex between phenol and the smallest conceptual BG building block methyl β-glucopyranoside. In contrast, calculations on a stacked vanillin-methyl β-glucopyranoside complex did not agree with the measured changes in carbohydrate ^{1}H chemical shifts after mixing of methyl β-glucopyranoside and vanillin. This suggests that vanillin and the sugar do not combine to form a stacked complex in solution. At this point it is unclear if and how this effect is related to the problematic behavior of vanillin and vanillin related compounds in some type of measurements (e.g. inconsistent detection in headspace above BG solutions mentioned in the start of Section 5.4.1), but future work will hopefully clarify the issue. Further studies are also required to elucidate the possible implications of the effect on the formulation of

novel food products containing BG and vanillin analogues.

Natural product: The combination of VCD and DFT was used to assign the absolute configuration of the natural product Ginkgolide B (Section 5.6). The agreement between the calculated spectra and the measured spectra for GB was striking. It was possible to discriminate between calculated spectra for all ginkgolides, including variants of GB with inversion at a single chiral center.

Functionality

BG: Flavor release profiles of esters and alcohols from oat and barley BG matrices (Section 5.2) showed that the compounds in most cases were retained in the BG matrices relative to an aqueous solution. The retention of esters in general increased with compound MW and with the concentration of the BG product, consistent with viscosity induced retention. Alcohols were less strongly retained than esters with similar MW, and were released more from some matrices (particularly the 5 percent PromOat matrix) than from aqueous solution. Apart from scaling of magnitudes across all compounds, there were generally no conspicuous changes in the shape of flavor release profiles between oat and barley matrices, suggesting little or no influence of the BG molecular weight.
Molecular descriptors for the flavor compounds were successfully related to the variation in flavor release with QSPR. The QSPR models were simple, and the flavor release was highly correlated with many molecular descriptors, indicating a simple flavor release mechanism.

The data from the dialysis of vanillin-related compounds and bile salts in BG matrices (Section 5.3.1) showed compound specific retention in the BG matrix. However, multivariate data analysis based on PCA and PLS failed to explain this variation in terms of molecular descriptors. The main conclusion is the qualitative observation that strongly hydrophillic compounds, such as glycosides of vanillin analogues, displayed an increased tendency of being rejected from the BG matrix.

Methodology

QSAR/QSPR: The influence of the choice of geometry optimization method on the final QSAR/QSPR model (Section 5.5) within the framework employed in Paper IV was minimal. The small changes in q^2 observed did not vary predictably with the sophistication of the geometry optimization method. Hence under similar circumstances, the application of a pragmatic geometry optimization method based on molecular mechanics or semi-empirical theory may lead to satisfactory QSAR/QSPR models.

The future

In summary this thesis has demonstrated some early steps toward elucidation of structure and functionality of a promising health promoting dietary fibre, the mixed-linkage (1→3),(1→4)-β-D-glucan. It is clear, however, that further work is required. In particular the nature of the reported results do not allow for bridging the gap between top-down and bottom-up approaches. Such integration of results from the macroscopic and microscopic domains is crucial for a holistic understanding of the BG polymer and a full realization of its potential, and should be the goal of future research.

Bibliography

[1] Miller, B. L. and Cummings, J. L. *The Human Frontal Lobes - Functions and Disorders.* The Guilford Press, (1998). Second Edition.

[2] Clark, J. E. *Proceedings of the Nutrition Society* **57**, 639–643 (1998).

[3] Goldberg, J. P., Belury, M. A., Elam, P., Finn, S. C., Hayes, D., Lyle, R., Jeor, S. S., Warren, M., and Hellwig, J. P. *Journal of the American Dietetic Association* **104**(7), 1141 – 1147 (2004).

[4] Ellwood, D. *The Australian and New Zealand Journal of Obstetrics and Gynaecology* **48**(3), 227 – 227 (2008).

[5] Frewer, L., Scholderer, J., and Lambert, N. *British Food Journal* **105**(10), 714 – 731 (2003).

[6] Wood, P. J. *Journal of Cereal Science* **46**(3), 230–238 (2007).

[7] Iqbal, R., Anand, S., Ounpuu, S., Islam, S., Zhang, X., Rangarajan, S., Chifamba, J., Al-Hinai, A., Keltai, M., Yusuf, S., and on behalf of the INTERHEART Study Investigators. *Circulation* **118**(19), 1929–1937 (2008).

[8] Cordain, L., Eaton, S. B., Sebastian, A., Mann, N., Lindeberg, S., Watkins, B. A., O'Keefe, J. H., and Brand-Miller, J. *American Journal of Clinical Nutrition* **81**(2), 341–354 (2005).

[9] Lawther, M. `http://www.functionalfoodnet.org/images/site/assets/beta_glucan.pdf`, (1996). β-Glucans and cereal soluble fibres: their role in functional foods.

[10] Lee, S., Inglett, G. E., Palmquist, D., and Warner, K. *LWT - Food Science and Technology* **42**(1), 350 – 357 (2009).

[11] Cardello, A. V. In *Food choice, acceptance and consumption,* Meiselman, H. L. and MacFie, H. J. H., editors, 1–82. Blackie Academic & Professional (1996).

[12] Rondanelli, M., Opizzi, A., and Monteferrario, F. *Minerva medica* **100**(3), 237–245 (2009).

[13] Limbach, H. J. and Kremer, K. *Trends in Food Science & Technology* **17**(5), 215 – 219 (2006). Functionality in Complex Foods.

[14] Adunsky, A. and Hershkowitz, M. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **53**(4), B306–309 (1998).

[15] Lichtenstein, A. H., Appel, L. J., Brands, M., Carnethon, M., Daniels, S., Franch, H. A., Franklin, B., Kris-Etherton, P., Harris, W. S., Howard, B., Karanja, N., Lefevre, M., Rudel, L., Sacks, F., Van Horn, L., Winston, M., and Wylie-Rosett, J. *Circulation* **114**(1), 82–96 (2006).

[16] Labanowski, J. K. `http://www.ccl.net/cca/documents/DFT/dft-overview/`, (1996). Simplified and Biased Introduction to Density Fucntional Approaches in Chemistry.

[17] Ban, F., Gauld, J. W., Wetmore, S. D., and Boyd, R. J. In *EPR of Free Radicals in Solids,* Lund, A. and Shiotani, M., editors, 246. Kluwer Academic Publishers (2001).

[18] Hansen, P. I., Larsen, F. H., Motawia, S. M., Blennow, A., Spraul, M., Dvortsak, P., and Engelsen, S. B. *Biopolymers* **89**(12), 1179–1193 (2008).

[19] French, A. D. and Johnson, G. P. *Canadian Journal of Chemistry* **84**, 603–612(10) (2006).

[20] MacKerell, A. D., Bashford, D., Bellott, Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. *The Journal of Physical Chemistry B* **102**(18), 3586–3616 (1998).

[21] Freddolino, P. L., Arkhipov, A. S., Larson, S. B., McPherson, A., and Schulten, K. *Structure* **14**(3), 437–449 (2005).

[22] Germann, T. C. and Kadau, K. *International Journal of Modern Physics C* **19**(9), 1315–1319 (2008).

[23] Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., and Woods, R. J. *Journal of Computational Chemistry* **29**(4), 622–655 (2008).

[24] Lins, R. D. and Hünenberger, P. H. *Journal of Computational Chemistry* **26**(13), 1400–1412 (2005).

[25] Damm, W., Frontera, A., Tirado-Rives, J., and Jorgensen, W. L. *Journal of Computational Chemistry* **18**(16), 1955–1970 (1997).

[26] Guvench, O., Greene, S. N., Kamath, G., Brady, J. W., Venable, R. M., Pastor, R. W., and Jr, A. D. M. *Journal of Computational Chemistry* **29**(15), 2543–2564 (2008).

[27] Hatcher, E. R., Guvench, O., and Mackerell, A. D. *Journal of Chemical Theory and Computation* **5**(5), 1315–1327 (2009).

[28] Kuttel, M., Brady, J. W., and Naidoo, K. J. *Journal of computational chemistry* **23**(13), 1236–1243 (2002).

[29] Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. *Journal of Computational Chemistry* **4**(2), 187–217 (1983).

[30] Brooks, B. R. R., Brooks, C. L. L., Mackerell, A. D. D., Nilsson, L., Petrella, R. J. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W. W., Post, C. B. B., Pu, J. Z. Z., Schaefer, M., Tidor, B., Venable, R. M. M., Woodcock, H. L. L., Wu, X., Yang, W., York, D. M. M., and Karplus, M. *Journal of Computational Chemistry* **30**(10), 1545–1614 (2009).

[31] Palma, R., Himmel, M. E., Liang, G., and Brady, J. W. In *Glycosyl Hydrolases in Biomass Conversion: ACS Symposium Series,* Himmel, M. E., editor, 112–130. American Chemical Society (2001).

[32] Ha, S. N., Giammona, A., Field, M., and Brady, J. W. *Carbohydrate Research* **180**(2), 207–211 (1988).

[33] Corzana, F., Motawia, M. S., du Penhoat, C. H., Perez, S., Tschampel, S. M., Woods, R. J., and Engelsen, S. B. *Journal of Computational Chemistry* **25**(4), 573–586 (2003).

[34] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. *Journal of Chemical Physics* **79**(2), 926–935 (1983).

[35] Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. *Journal of Computational Physics* **23**(3), 327 – 341 (1977).

[36] Thompson, M. A. *ArgusLab 4.0.1.* Planaria Software LCC, Seattle, WA, (2004). Available at: `http://www.arguslab.com`.

[37] Fletcher, R. *Practical Methods of Optimization.* Wiley, (1980).

[38] Gill, P. E., Murray, W., and Wright, M. H. *Practical Optimization.* Academic Press, (1981).

[39] Chang, G., Guida, W., and Still, W. C. *Journal of the American Chemical Society* **111**(12), 4379–4386 (1989).

[40] Saunders, M., Houk, K. N., Wu, Y.-D., Still, W. C., Lipton, M., Chang, G., and Guida, W. C. *Journal of the American Chemical Society* **112**(4), 1419–1427 (1990).

[41] Dowd, M. K., Zeng, J., French, A. D., and Reilly, P. J. *Carbohydrate Research* **230**(2), 223 – 244 (1992).

[42] Swope, W. C., Andersen, H. C., Berens, P. H., and Wilson, K. R. *The Journal of Chemical Physics* **76**(1), 637–649 (1982).

[43] Andersson, C. and Balling Engelsen, S. *Journal of Molecular Graphics and Modelling* **17**(2), 101–105(5) (1999).

[44] Karplus, M. *The Journal of Chemical Physics* **30**(1), 11–15 (1959).

[45] Karplus, M. *Journal of the American Chemical Society* **85**(18), 2870–2871 (1963).

[46] Tvaroska, I., Hricovini, M., and Petrakova, E. *Carbohydrate Research* **189**, 359–362 (1989).

[47] Hansen, P. I. PhD Thesis, Department of Food Science, Faculty of Life Sciences, University of Copenhagen, (2008). Molecular Modeling and NMR studies of Starch Structural Motifs.

[48] Born, M. and Huang, K. *Dynamical Theory of Crystal Lattices*. Oxford University Press, (1954).

[49] Kolos, W. and Wolniewicz, L. *Journal of Chemical Physics* **41**(12), 3663–3673 (1964).

[50] Payne, M. C., Teter, M. P., Allan, D. C., Arias, T. A., and Joannopoulos, J. D. *Reviews of Modern Physics* **64**(4), 1045–1097 (1992).

[51] Hartree, D. R. *Proceedings of the Cambridge Philosophical Society* **24**(1), 89–110 (1928).

[52] Ballentine, L. E. *Quantum Mechanics – a Modern Development*. World Scientific, (2003).

[53] Kohn, W. *Reviews of Modern Physics* **71**(5), 1253–1266 (1999).

[54] Kohanoff, J. *Electronic Structure Calculations for Solids and Molecules*. Cambridge, (2006).

[55] Becke, A. D. In *The Challenge of d and f Electrons: Theory and Computation,* Salahub, D. R. and Zerner, M. C., editors, 166. American Chemical Society (1989).

[56] Becke, A. D. *The Journal of Chemical Physics* **98**(7), 5648–5652 (1993).

[57] Lee, C., Yang, W., and Parr, R. G. *Physical Review B* **37**(2), 785–789 (1988).

[58] Vosko, S. H., Wilk, L., and Nusair, M. *Canadian Journal of Physics* **58**(8), 1200–1211 (1980).

[59] Jensen, F. *Introduction to computational chemistry.* Wiley, (1999).

[60] Csonka, G. I., French, A. D., Johnson, G. P., and Stortz, C. A. *Journal of Chemical Theory and Computation* **5**(4), 679–692 (2009).

[61] Cramer, C. J. In *Essentials of Computational Chemistry, Theories and Models, Second Edition*, 111–117. Wiley (2004).

[62] Slater, J. C. *Physical Review* **36**(1), 57–64 (1930).

[63] Boys, S. F. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **200**(1063), 542–554 (1950).

[64] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, Jr., J. A., Vreven, T., Kudin, K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G. A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian, H. P., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C., and Pople, J. A. *Gaussian 03, Revision C.02.* (2004). Gaussian, Inc., Wallingford, CT.

[65] Hehre, W. J., Stewart, R. F., and Pople, J. A. *The Journal of Chemical Physics* **51**(6), 2657–2664 (1969).

[66] Davidson, E. and Feller, D. *Chemical Reviews* **86**(4), 681–696 (1986).

[67] Frisch, M. J., Pople, J. A., and Binkley, J. S. *Journal of Chemical Physics* **80**(7), 3265–3269 (1984).

[68] Francl, M. M., Pietro, W. J., Hehre, W. J., Binkley, J. S., Gordon, M. S., DeFrees, D. J., and Pople, J. A. *Journal of Chemical Physics* **77**(7), 3654–3665 (1982).

[69] Ramachandran, K. I., Deepa, G., and Namboori, K. *Computational Chemistry and Molecular Modeling*. Springer, (2006).

[70] Cross, G., Lloyd, T., Salter-Duke, B., Wong, M., and Yates, B. *Basic ab initio quantum chemistry*. The Australian Computational Chemistry Via the Internet Project Team, (2005). Available at: `http://www.chem.swin.edu.au/modules/mod5/index.html`.

[71] Olsen, L. PhD Thesis, Institute of Mathematics and Physics, The Royal Veterinary and Agricultural University, Copenhagen, Denmark, (2003). Computational Studies on Metallo-β-lactamases.

[72] Wolinski, K., Hinton, J. F., and Pulay, P. *Journal of the American Chemical Society* **112**(23), 8251–8260 (1990).

[73] `http://www.talete.mi.it/products/dragon_description.htm`. DRAGON 5.5, TALETE Srl.

[74] Steinbeck, C., Han, Y., Kuhn, S., Horlacher, O., Luttmann, E., and Willighagen, E. *Journal of Chemical Information and Computer Sciences* **43**(2), 493–500 (2003).

[75] QikProp, version 2.5, Schrödinger, LLC, New York, NY, 2005.

[76] Hansch, C., Steward, A. R., Anderson, S. M., and Bentley, D. L. *Journal of Medicinal Chemistry* **11**(1), 1–11 (1968).

[77] Cayley, A. *Philosophical Magazine* (47), 444 (1874).

[78] Todeschini, R., Consonni, V., Mannhold, R., Kubinyi, H., and Timmerman, H. *Handbook of Molecular Descriptors*. Wiley-VCH, (2000).

[79] Ivanciuc, O. In *Molecular Drug Properties,* Mannhold, R., editor, 85–105. WILEY-VCH (2008).

[80] `http://www.epa.gov/nrmrl/std/cppb/qsar/`, (2009). United States Environmental Protection Agency, Molecular Descriptors Guide (PDF).

[81] Gutman, I., Russic, B., Trinajstic, N., and C. F. Wilcox, J. *The Journal of Chemical Physics* **62**(9), 3399–3405 (1975).

[82] Gutman, I. *Journal of Chemical Information and Computer Sciences* **34**(5), 1087–1089 (1994).

[83] Skorobogatov, V. A. and Dobrynin, A. A. *Communications in Mathematical and in Computer Chemistry* **23**, 105–151 (1998).

[84] Konstantinova, E. and Skorobogatov, V. *Journal of Chemical Information and Computer Sciences* **35**(3), 472–478 (1995).

[85] Rücker, G. and Rücker, C. *Journal of Chemical Information and Computer Sciences* **5**(33), 683–695 (1993).

[86] Randic, M., Brissey, G., Spencer, R., and Wilkins, C. *Computers & Chemistry* **3**(1), 5–13 (1979).

[87] Randic, M. *Communications in Mathematical and in Computer Chemistry* **7**, 5–64 (1979).

[88] Randic, M. and Jurs, P. *Quantitative Structure-Activity Relationships* **8**(1), 39–48 (1989).

[89] McCullagh, P. and Nelder, J. A. *Generalized Linear Models.* Chapman & Hall/CRC, (1989).

[90] Moore, E. H. *Bulletin of the American Mathematical Society* **26**, 394–395 (1920).

[91] Penrose, R. *Proceedings of the Cambridge Philosophical Society* **51**(3), 406–413 (1955).

[92] Selassie, C. D. In *Burger's Medicinal Chemistry and Drug Discovery. Sixth Edition, Volume 1: Drug Discovery,* Abraham, D. J., editor, 1–48. Wiley (2003).

[93] Liu, P. and Long, W. *International Journal of Molecular Sciences* **10**(5), 1978–1998 (2009).

[94] Esbensen, K. H. *Multivariate Data Analysis - in practice.* CAMO, (2002).

[95] Geladi, P. and Kowalski, B. *Analytica Chimica Acta* **185**(1), 1–17 (1986).

[96] Lorber, A., Wangen, L. E., and Kowalski, B. R. *Journal of Chemometrics* **1**(1), 19–31 (1987).

[97] Martens, H. and Næs, T. *Multivariate Calibration.* Wiley, (1989).

[98] Schüürmann, G., Ebert, R.-U., Chen, J., Wang, B., and Kühne, R. *Chemical Information and Modeling* **48**(11), 2140–2145 (2008).

[99] Gonzalez, M. P., Teran, C., Saiz-Urra, L., and Teijeira, M. *Current Topics in Medicinal Chemistry* **8**(18), 1606–1627.

[100] Crum-Brown, A. and Fraser, T. R. *Transactions of the Royal Society of Edinburgh* **25**, 151–203 (1869).

[101] Martin, Y. C. *Journal of Medicinal Chemistry* **24**(3), 229–237 (1981).

[102] Winkler, D. A. *Briefing in Bioinformatics* **3**(1), 73–86 (2002).

[103] Gramatica, P. *QSAR & Combinatorial Science* **26**(5), 694–701 (2007).

[104] Rücker, C., Rücker, G., and Meringer, M. *Journal of Chemical Information and Modeling* **47**(6), 2345–2357 (2007).

[105] Taylor, A. J. and Linforth, R. S. T. *Nutrition & Food Science* **98**(4), 202–206 (1998).

[106] Lin, S. K., Lambert, J. R., and Wahlqvist, M. *Asia Pacific Journal of Clinical Nutrition* **1**(1), 37–42 (1992).

[107] Macomber, R. S. *A complete introduction to modern NMR spectroscopy.* Wiley-Interscience, (1998).

[108] Zeeman, P. *Astrophysical Journal* **5**, 332–347 (1897).

[109] Becker, E. D. *High Resolution NMR.* Academic Press, (1989).

[110] Facelli, J. C. *Magnetic Resonance Part A* **20A**(1), 42–69 (2004).

[111] Overhauser, A. W. *Physical Review* **92**(2), 411–415 (1953).

[112] Carver, T. R. and Slichter, C. P. *Physical Review* **92**(1), 212–213 (1953).

[113] Meissner, A. and Sørensen, O. W. *Magnetic Resonance in Chemistry* **39**(1), 49–52 (2001).

[114] Bax, A. and Davis, D. G. *Journal of Magnetic Resonance* **63**(1), 207–213 (1969).

[115] Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. *The Journal of Chemical Physics* **71**(11), 4546–4553 (1979).

[116] Dailey, B. P. *The Journal of Chemical Physics* **41**(8), 2304–2310 (1964).

[117] Wannere, C. S. and Schleyer, P. V. *Organic Letters* **5**(5), 605–608 (2003).

[118] Johnson, C. E. and Bovey, F. A. *The Journal of Chemical Physics* **29**(5), 1012–1014 (1958).

[119] Martin, N. H., Floyd, R. M., Woodcock, H. L., Huffman, S., and Lee, C.-K. *Journal of Molecular Graphics and Modelling* **26**(7), 1125–1130 (2008).

[120] Laidler, K. J. and Meiser, J. H. *Physical Chemistry*. Houghton Mifflin, (1999). Third Edition.

[121] Stephens, P. J., Devlin, F. J., and Pan, J.-J. *Chirality* **20**(5), 643–663 (2008).

[122] Kellenback, E. R., Dukor, R. K., and Nafie, L. A. *Spectroscopy Europe* **19**(4), 15–18 (2007).

[123] Stephens, T. D., Bunde, C. J. W., and Fillmore, B. J. *Biochemical Pharmacology* **59**(12), 1489–1499 (2000).

[124] Kolb, B. and Ettre, L. S. *Static Headspace-Gas Chromatography*. Wiley, (2006).

[125] Sostaric, T., Boyce, M. C., and Spickett, E. E. *Journal of Agricultural and Food Chemistry* **48**(12), 5802–5807 (2000).

[126] Kitson, F. G., Larsen, B. S., and McEwen, C. N. *Gas Chromatography and Mass Spectrometry - A practical guide*. Academic Press, (1996).

[127] Thomson, J. J. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, Fifth Series* , 296 (1897).

[128] Druaux, C. and Voilley, A. *Trends in Food Science & Technology* **8**(11), 364 – 368 (1997).

[129] Final rule. Federal Register: vol. 62, 3583-3601 (1997). U.S. Food and Drug Administration, Food labeling: health claims; oats and coronary heart disease.

[130] Sletmoen, M. and Stokke, B. T. *Biopolymers* **89**(4), 310–321 (2008).

[131] Richon, A. B. *Drug Discovery Today* **13**(15-16), 659–664 (2008).

[132] Pérez, S., Kouwijzera, M., Mazeaua, K., and Engelsen, S. B. *Journal of Molecular Graphics* **14**(6), 307–321 (1996).

[133] http://www.cermav.cnrs.fr/glyco3d/index.php, (2009). GLYCO3D, Centre de Recherches sur les Macromolécules Végétales.

[134] Ham, J. T. and Williams, D. G. *Acta Crystallographica Section B* **26**(9), 1373–1383 (1970).

[135] Chu, S. S. C. and Jeffrey, G. A. *Acta Crystallographica Section B* **24**(6), 830–838 (1968).

[136] Noguchi, K., Okuyama, K., Kitamura, S., and Takeo, K. *Carbohydrate Research* **49**(1), 33–43 (1992).

[137] Takeda, H., Noritake, Y., and Kasai, N. *Carbohydrate Research* **53**(2), 137–152 (1977).

[138] Pereira, C. S., Kony, D., Baron, R., Müller, M., van Gunsteren, W. F., and Hünenberger, P. H. *Biophysical Journal* **90**(12), 4337–4344 (2006).

[139] Deslandes, Y., Marchessault, R. H., and Sarko, A. *Macromolecules* **13**(6), 1466–1471 (1980).

[140] Chuah, C. T., Sarko, A., Deslandes, Y., and Marchessault, R. H. *Macromolecules* **16**(8), 1375–1382 (1983).

[141] Wuthrich, K. *NMR of proteins and nucleic acids*. Wiley, (1986).

[142] http://www.biovelop.com/products/default.asp?page=product&id=1, (2009). Biovelop.

[143] http://www.glucagel.com/, (2009). GraceLinc Ltd.

[144] Hansson, A., Luefvén, A., and van Ruth, S. *Journal of Agricultural and Food Chemistry* **51**(7), 2000–2005 (2003).

[145] Hollowood, T. A., Linforth, R. S. T., and Taylor, A. J. *Chemical Senses* **27**(7), 583–591 (2002).

[146] Berthod, A. and Hassoun, M. *Journal of Chromatography A* **1116**(1-2), 143–148 (2006).

[147] Rodgers, J. L. and Nicewander, W. A. *The American Statistician* **42**(1), 59–66 (1988).

[148] Allinger, N. L., Yuh, Y. H., and Lii, J.-H. *Journal of the American Chemical Society* **111**, 8551–8566 (1989).

[149] MacroModel, version 9.1, Schrödinger, LLC, New York, NY, 2005.

[150] Gramatica, P. *QSAR and Combinatorial Science* **25**(4), 327–332 (2006).

[151] Kohonen, J., Reinikainen, S.-P., Aaljoki, K., and Höskuldsson, A. *Chemometrics and Intelligent Laboratory Systems* **97**(2), 159 – 163 (2009).

[152] Guth, H. and Fritzler, R. *Chemistry and Biodiversity* **1**(12), 2001–2023 (2004).

[153] Kubinyi, H. *Drug Discovery Today* **2**(12), 538–546(9) (1997).

[154] Fitzgerald, D. J., Stratford, M., Gasson, M. J., and Narbad, A. *Journal of Agricultural and Food Chemistry* **53**(5), 1769–1775 (2005).

[155] Lirdprapamongkol, K., Kramb, J.-P., Suthiphongchai, T., Surarit, R., Srisomsap, C., Dannhardt, G., and Svasti, J. *Journal of Agricultural and Food Chemistry* **57**(8), 3055–3063 (2009).

[156] Lawther, M. personal communication, (2009).

[157] de Jager, L. S., Perfetti, G. A., and Diachenko, G. W. *Journal of Chromatography A* **1145**(1-2), 83–88 (2007).

[158] de Jager, L. S., Perfetti, G. A., and Diachenko, G. W. *Food Chemistry* **107**(4), 1701–1709 (2008).

[159] del Carmen Fernández-Alonso, M., Cañada, F. J., Jiménez-Barbero, J., and Cuevas, G. *Journal of the American Chemical Society* **127**(20), 7379–7386 (2005).

[160] Vandenbussche, S., Díaz, D., Fernández-Alonso, M. C., Pan, W., Vincent, S., Cuevas, G., Cañada, F. J., Jiménez-Barbero, J., and Bartik, K. *Chemistry* **14**(25), 7570–7578 (2008).

[161] Tatko, C. *Nature Chemical Biology* **4**(10), 586–587 (2008).

[162] Morales, J. C., Reina, J. J., Díaz, I., Aviñó, A., Nieto, P. M., and Eritja, R. *Chemistry* **14**(26), 7828–7835 (2008).

[163] Spiwok, V., Lipovova, P., Skalova, T., Vondrackova, E., Dohnalek, J., Hasek, J., and Kralova, B. *Journal of Computer-Aided Molecular Design* **19**(12), 887–901(15) (2005).

[164] Nishio, M., Hirota, M., and Umezawa, Y. *The CH/π interaction. Evidence, nature and consequences*. Wiley-VCH, (1998).

[165] Sujatha, M. S., Sasidhar, Y. U., and Balaji, P. V. *Biochemistry* **44**(23), 8554–8562 (2005).

[166] Tsuzuki, S., Uchimaru, T., and Mikami, M. *The Journal of Physical Chemistry B* **113**(16), 5617–5621 (2009).

[167] Martin, N. H., Allen, N. W., Moore, K. D., and Vo, L. *Journal of Molecular Structure (Theochem.)* **454**(2-3), 161–166 (1998).

[168] Martin, N. H., Allen, N. W., and Moore, J. C. *Journal Molecular Graphics and Modelling* **18**(3), 242–246 (2000).

[169] Martin, N. H., Loveless, D. M., Main, K. L., and Wade, D. C. *Journal of Molecular Graphics and Modelling* **25**(4), 389–395 (2006).

[170] Flugge, L. A., Blank, J. T., and Petillo, P. A. *Journal of the American Chemical Society* **121**(31), 7228–7238 (1999).

[171] Guérin, D. M., Lascombe, M. B., Costabel, M., Souchon, H., Lamzin, V., Béguin, P., and Alzari, P. M. *Journal of Molecular Biology* **316**(5), 1061–1069 (2002).

[172] Weiner, S., Kollman, P., Case, D., Singh, U., Ghio, C., Alagona, G., Profeta Jr., S., and Weiner, P. *Journal of the American Chemical Society* **106**(3), 765–784 (1984).

[173] Weiner, S., Kollman, P., Nguyen, D., and Case, D. *Journal of Computational Chemistry* **7**(2), 230–252 (1986).

[174] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. *Journal of the American Chemical Society* **117**(19), 5179–5197 (1995).

[175] HyperChem(TM) Professional 7.51, Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA.

[176] Mennucci, B. and Tomasi, J. *Journal of Chemical Physics* **106**(12), 5151–5158 (1997).

[177] Cossi, M., Barone, V., Mennucci, B., and Tomasi, J. *Chemical Physics Letters* **286**(3-4), 253–260 (1998).

[178] Cancès, M. T., Mennucci, B., and Tomasi, J. *Journal of Chemical Physics* **107**(8), 3032–3041 (1997).

[179] O'Brien, S. E. and Popelier, P. L. *Canadian Journal of Chemistry* **77**(1), 28–36 (1999).

[180] Roy, K. and Popelier, P. L. *Bioorganic & Medicinal Chemistry Letters* **18**(8), 2604 – 2609 (2008).

[181] Winum, J.-Y., Scozzafava, A., Montero, J.-L., and Supuran, C. T. *Expert Opinion on Therapeutic Patents* **16**(1), 27–47 (2006).

[182] He, L. and Jurs, P. C. *Journal of Molecular Graphics and Modelling* **23**(6), 503–523 (2005).

[183] Agatonovic-Kustrin, S., Beresford, R., and Yusof, A. P. M. *Journal of Pharmaceutical and Biomedical Analysis* **26**(2), 241 – 254 (2001).

[184] Rücker, C., Scarsi, M., and Meringer, M. *Bioorganic & Medicinal Chemistry* **14**(15), 5178 – 5195 (2006).

[185] Golbraikh, A. and Tropsha, A. *Journal of Molecular Graphics and Modelling* **20**(4), 269–276 (2002).

[186] Speirs, A. L. *Lancet* **1**(7224), 303–305 (1962).

[187] Zawirska-Wojtasiak, R. *ACTA Scientiarum Polonorum, Technologia Alimentaria* **5**(1), 21–36 (2006).

[188] Zhang, H., Wang, Z., and Xu, S.-Y. *Journal of Food Engineering* **80**(4), 1226 – 1232 (2007).

[189] Jacobs, B. P. and Browner, W. S. *The American Journal of Medicine* **108**(4), 341–342 (2000).

[190] Strømgaard, K. and Nakanishi, K. *Angewandte Chemie International Edition* **43**(13), 1640–1658 (2004).

[191] Nakanishi, K. and Habaguchi, K. *Journal of the American Chemical Society* **93**(14), 3546–3547 (1971).

[192] He, J., Petrovic, A. G., Dzyuba, S. V., Berova, N., Nakanishi, K., and Polavarapu, P. L. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **69**(4), 1213 – 1222 (2008).

[193] Zhu, W.-L., Puah, C. M., Tan, X.-J., Jiang, H.-L., Chen, K.-X., and Ji, R.-Y. *Journal of Molecular Structure: THEOCHEM* **528**(1-3), 193 – 198 (2000).

[194] Staudte, R., Woodward, J., Fincher, G., and Stone, B. *Carbohydrate Polymers* **3**(4), 299 – 312 (1983).

[195] Wood, P. J. and Weisz, J. *Cereal Chemistry* **68**, 31–39 (1991).

[196] Blundell, C. D., Reed, M. A., Overduin, M., and Almond, A. *Carbohydrate Research* **341**(12), 1985 – 1991 (2006).

[197] van Ruth, S. M. and Roozen, J. P. *Talanta* **52**(2), 253 – 259 (2000).

[198] Graham, H. D. and Thomas, L. B. *Journal of Pharmaceutical Sciences* **51**(10), 988–992 (1962).

Paper I

Niels Johan Christensen, Peter Ibsen Hansen, Flemming Hofmann Larsen, Tore Folkerman, Mohammed Saddik Motawia, and Søren Balling Engelsen. *A combined nuclear magnetic resonance and molecular dynamics study of the two structural motifs for mixed linkage β-glucans: Methyl β-cellobioside and methyl β-laminarabioside.* Carbohydrate Research, **Submitted.**

A combined nuclear magnetic resonance and molecular dynamics study of the two structural motifs for mixed linkage β-glucans: methyl β-cellobioside and methyl β-laminarabioside

Niels Johan Christensen [a,*], Peter Ibsen Hansen [b], Flemming Hofmann Larsen [a], Tore Folkerman [c], Mohammed Saddik Motawia [c], and Søren Balling Engelsen [a]

[a] Quality & Technology, Department of Food Science, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Copenhagen, Denmark

[b] Department of Food Science, Cornell University, Ithaca, NY, 14853

[c] Plant Biochemistry Laboratory, Department of Plant Biology and Biotechnology, University of Copenhagen, Thorvaldsensvej 40, DK-1871 Frederiksberg C, Copenhagen, Denmark

*Corresponding author. E-mail address: njc@life.ku.dk.

Supplementary Data is available for this article.

Abstract

The conformational and hydration properties of the two disaccharides methyl β-cellobioside and methyl β-laminarabioside were investigated by NMR spectroscopy and explicit solvation molecular dynamics simulations using the carbohydrate solution force field (CSFF). Adiabatic maps produced with this force field display 4 minima for methyl β-cellobioside and 3 minima for methyl β-laminarabioside. Molecular dynamics simulations were initiated from all minima. For each disaccharide, simulations started from the global minimum were conducted for 50 ns, while the other minima were explored for 10 ns. The simulations revealed two stable minima for both compounds. In case of methyl β-cellobioside, the simulation minima in aqueous solution were shifted from their adiabatic map counterparts, while the simulation minima for methyl β-laminarabioside coincided with the corresponding adiabatic map minima. To validate the theoretical results, NMR derived NOEs and coupling constants across the glycoside linkage, $^{3}J_{HC}$ and $^{3}J_{CH}$, were calculated from the MD trajectories and compared with the experimental values. For each disaccharide, the best agreement was obtained for the simulations started at the global minimum. For both compounds, inter-ring water bridges in combination with the direct hydrogen bonds between the same groups were found to be determining factors for the overall solution structure of the disaccharide which differs from the solid state structure. Comparison with helical parameters showed that the preferred glycosidic dihedral configurations in the MD simulations were not highly compatible with the cellulose or curdlan helix structures observed in the solid state. Polymers generated using glycosidic dihedral angles from the simulations revealed secondary structure motifs that that may help to elucidate polymer associations and small-molecule binding.

Keywords: Mixed linkage beta glucan**;** Methyl β-cellobioside; Methyl β-laminarabioside; Molecular dynamics simulations; Adiabatic map; Hydration analysis.

1. Introduction

Exclusively β(1→4) or β(1→3) linked polymers of glucose play different roles in biological systems. For instance, cellulose provides structural rigidity for land plants, whereas laminaran acts as an energy reservoir for certain brown algae and brown sea weeds.[1] The β(1→4) linked polymers lead to a flat fibrous structure that will align into microfibrils[2], while the β(1→3) polymer yields triple helical structures.[3] Both of these arrangements of the polymers are insoluble in water, but interestingly mixed linkage β(1→3)β(1→4) glucans are excellent gelling agents and texturizers.[4] Mixed linkage β(1→3)β(1→4) glucans are found in particular high concentrations in the cell walls of oat and barley and numerous nutritional experiments support the beneficial effect of beta-glucan (BG) enriched diets on sugar and cholesterol metabolism, digestion and on absorption of carcinogenic substances.[5] Considerable research activity has been initiated by these observations directed towards understanding the bioactivity of BG. The aim of this work is to contribute to the elucidation of the structural basis for the bioactivity of mixed linkage β(1→3)β(1→4) glucans by focusing on fragments representing characteristic molecular motifs. A combination of molecular dynamics simulations and two-dimensional homo- and heteronuclear NMR spectroscopy was employed to provide theoretical and experimental characterization of the solution structure and hydration of methyl β-cellobioside and methyl β-laminarabioside (**Figure 1**). These methylated derivates of laminarabiose (D-glucopyranosyl-(1→3)-β-D-glucopyranose) and cellobiose (D-glucopyranosyl-(1→4)-β-D-glucopyranose) are the fundamental models of disaccharide subunits of β-glucans as each model compound represents one of the characteristic glucosidic β-linkages.

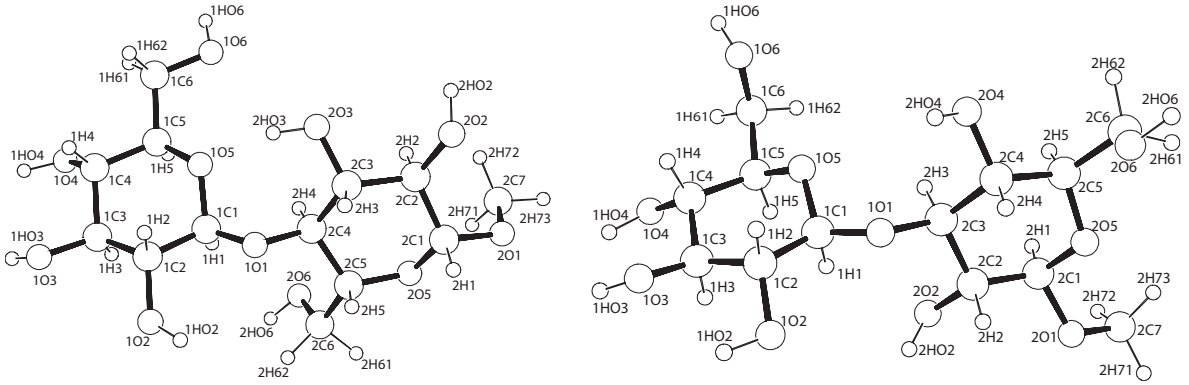


Figure 1: Molecular structure and atomic labelling for methyl β-cellobioside (left) and methyl β-laminarabioside (right).

The notion inspiring the study of building blocks of carbohydrate polymers is that the smaller fragments contain many of the conformation determining features without the prohibitively large number of atoms of the polymer.[6] In part due to its pivotal role in cellulose research cellobiose has been thoroughly investigated with computational methods. Laminaran and laminarabiose have been subjected to several computational and experimental studies. Often, parallel investigations of the β(1,3) and β(1,4) compounds are reported. The establishment of conformational maps for cellobiose, laminarabiose and related compounds has been addressed in several studies. The steric maps reported by Sathyanarayana and Rao are among the first studies in this area.[1] Examples of later studies are the relaxed residue conformational maps of cellobiose and laminarabiose based on the MM3[78] force field.[9] More recently, quantum chemical methods have been employed either in hybrid with molecular mechanics[10] or alone[11,12] for the generation of adiabatic maps of cellobiose, laminarabiose and related compounds. Molecular dynamics studies have undergone similar improvements. Since 200 ps simulations of cellobiose with and without explicit waters in 1993[13,14] increases in both simulation time and force field diversity have been reported. Some examples are 100 ps simulations of laminaran oligomers with the consistent valence force field (CVFF)[15], 1 ns simulations of all cello-oligosaccharides up to hexamer size with AMBER/GLYCAM[16], 10 ns

simulations of laminarabiose[6] with the OPLS-AA-SEI[17] force field, 30 ns simulations of laminarabiose with a modified FFGMX united atom force field in the GROMACS software package[18], and recently 50 ns simulations of cellobiose, laminarabiose and other disaccharides with a modified version of the GROMOS 45A4 force field[19] with carbohydrate parameters.[20] These studies all include explicit treatment of solvent. The combination of NMR and molecular dynamics is also reported, although accounts of such approaches are less numerous than those of purely theoretical studies. In 1985 Lipkind *et al.* demonstrated a combined computational and experimental approach, by measuring and calculating a nuclear Overhauser effect in α-cellobiose l-phosphate.[21] In a more recent effort, NMR spectroscopy and 400 ps molecular dynamics simulations were employed in the study of a series of disaccharides, including cellobiose and laminarabiose.[22] As these examples demonstrate, cellobiose and laminarabiose have been relatively thoroughly investigated computationally. In recent literature[23,24] it has been demonstrated that different carbohydrate force fields may provide different results and it is thus advisable to examine molecular carbohydrate systems by more than one force field and to validate the results experimentally. By using a different force field and O-methylated sugars, the present study complements the simulations of cellobiose and laminarabiose in the study by Pereira *et al.*.[20] This is the first combined MD/NMR comparative study of the methylated analogs.

2. Experimental

2.1 Nomenclature

In this study atoms are referred to by three letter codes. An example is 1O5, where the first digit indicates the pyranose ring number, the character is the atom symbol and the trailing digit is the standard atom numbering within the pyranose ring. Conformational flexibility around the glycosidic linkage β (1→4) in methyl β-cellobioside and β (1→3) in methyl β-laminarabioside is described by the two torsional angles Φ and Ψ. These are defined as Φ = 1O5-1C1-1O1-2C4 and Ψ = 1C1-1O1-

2C4-2C5 for methyl β-cellobioside and Φ = 1O5-1C1-1O1-2C3 and Ψ = 1C1-1O1-2C3-2C4 for methyl β-laminarabioside. The orientation of the hydroxymethyl group in residue *i* is defined by ω_i = *i*O5–*i*C5–*i*C6–*i*O6.

2.2 Synthesis of methyl β-cellobioside and methyl β-laminarabioside

The strategy for chemical synthesis of the disaccharides methyl β-D-glucopyranosyl-(1→4)- β-D-glucopyranoside (**8**) (methyl β-cellobioside **8**) and methyl β-D-glucopyranosyl-(1→3)- β-D-glucopyranoside (**9**)(methyl β-laminarabioside **9**) is outlined in Chart **1**. In this strategy, D-glucose (**1**) was converted into phenyl 2,3,4,6-tetra-O-acetyl-1-thio-β-D.glucopyranoside **(3**) as described by literature.[25] The commercially available methyl β-D-glycopyranoside (**2**) was manipulated and converted into both methyl 2,3,6-O-benzyl-β-D-glucopyranoside (**4**)[26] and methyl 2,4,6-O-benzyol-β-D-glucopyranoside (**5**)[27] according to the reported procedures. Glycosidation reaction between the phenylthio glucopyranoside derivative **3** as a glycosyl donor using the methyl glucopyranoside derivatives **4** or **5** as the glycosyl acceptors with the activator NIS / TfOH system[28] provided the corresponding disaccharide derivatives **6** and **7**. Removal of the protecting groups from both compounds **6** and **7** afforded the desired disaccharides **8** and **9**. Detailed experimental data will be published elsewhere.

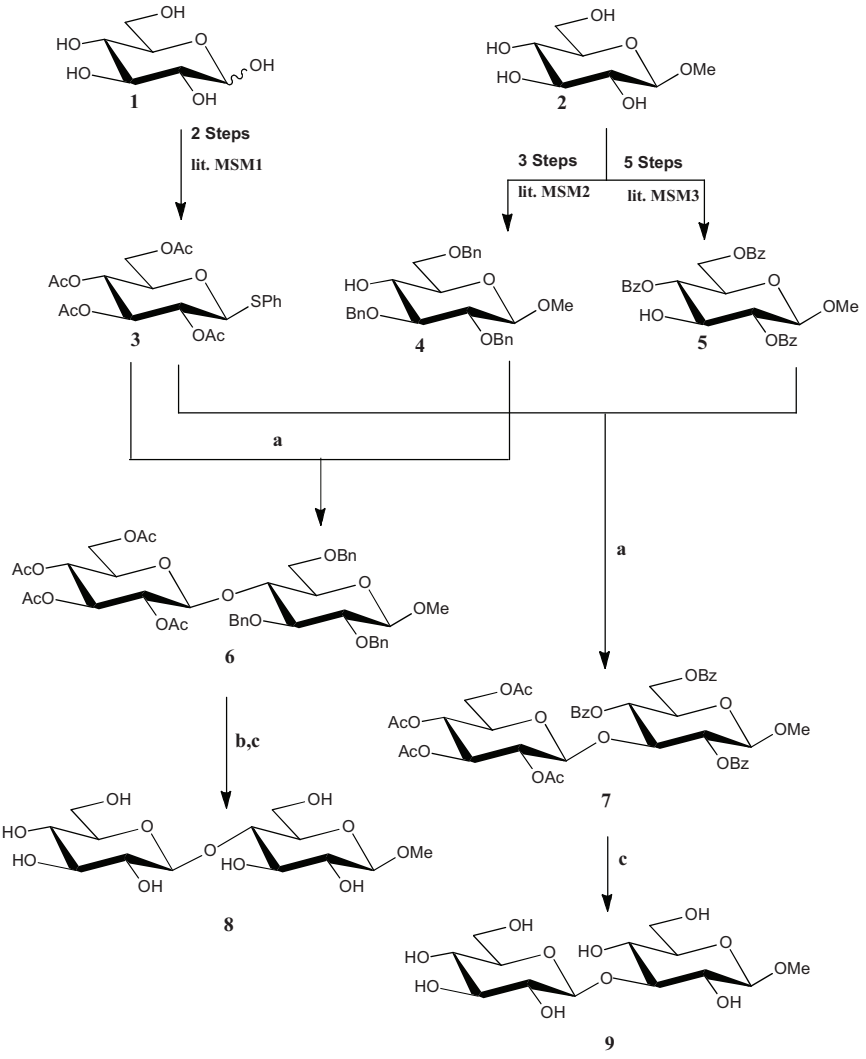


Chart 1: The strategy for chemical synthesis of the disaccharides methyl β-D-glucopyranosyl-(1→4)- β-D-glucopyranoside (**8**) and methyl β-D-glucopyranosyl-(1→3)- β-D-glucopyranoside (**9**).

2.3 NMR spectroscopy

25.7 mg methyl β-cellobioside and 43 mg methyl β-laminarabioside were dissolved in 550 µL D_2O (with 5.8 mM TSP-d4) and transferred to 5 mm (o.d.) NMR tubes. Two-dimensional homonuclear (^{1}H-^{1}H) COSY and TOCSY[29] (70 ms mixing time) and heteronuclear (^{1}H-^{13}C)-HSQC[30] and ^{13}C J-HMBC[31] spectra, as well as NOESY spectra, were recorded for both samples. COSY, NOESY, TOCSY and ^{13}C-HSQC experiments were conducted on a Bruker Avance 400 spectrometer (9.4 T)

using a double-tuned BBI probe equipped for 5 mm (o.d.) NMR tubes and operating at Larmor frequencies of 400.13 and 100.62 MHz for ^{1}H and ^{13}C, respectively. The ^{13}C J-HMBC spectra were recorded on a Bruker Avance 800 spectrometer (18.8 T) using a TCI cryoprobe equipped for 5 mm (o.d.) NMR tubes and operating at Larmor frequencies 799.96 and 201.12 MHz for ^{1}H and ^{13}C, respectively. In the ^{13}C J-HMBC spectra a scaling factor of 40 was utilized. All NMR spectra were recorded at 298 K and referenced to TSP-d4. The subsequent assignment was performed using Sparky[32] and the result is displayed in **Table 1**.

Table 1: NMR assignment obtained at 18.8 T for all ^{1}H and ^{13}C chemical shifts for methyl β-D-cellobioside and methyl β-D-laminarabioside.

	methyl β-cellobioside		**methyl β-laminarabioside**	
ring 1	**δ_H**	**δ_C**	**δ_H**	**δ_C**
1	4.51	102.6	4.68	105.7
2	3.32	73.2	3.36	76.3
3	3.51	75.6	3.53	78.4
4	3.42	69.5	3.41	72.4
5	3.49	76.0	3.49	78.9
61	3.92	60.7	3.92	63.6
62	3.74	60.7	3.72	63.6
ring 2				
1	4.41	103.1	4.41	105.8
2	3.31	72.9	3.47	75.6
3	3.65	74.4	3.75	87.6
4	3.63	78.7	3.49	78.4
5	3.61	74.8	3.5	71.1
61	4.00	60.1	3.94	63.6
62	3.82	60.1	3.76	63.6
OMe	3.58	57.3	3.58	60.1

Couplings across the glycosidic linkages were assigned for both compounds by determination of heteronuclear $^{3}J_{CH}$ spin-spin couplings from the doublet splitting of cross-peaks in ^{13}C J-HMBC spectra. $^{3}J_{C4'\text{-}H1}$, $^{3}J_{C1\text{-}H4'}$, $^{3}J_{C3'\text{-}H1}$ and $^{3}J_{C1\text{-}H3}$, were used as experimental measures of the dihedral angles and are listed in **Table 2** along with the calculated values from the simulations. The calculated values were obtained from atomic coordinates using a Karplus-type relationship (**eq. 1**) parameterized using data from both Tvaroska *et al.*[33] and Mulloy *et al.*[34].

$$^{3}J_{CH}(\theta) = 0.8 - 0.7\cos(\theta) + 5.3\cos^{2}(\theta) \qquad \textbf{(eq. 1)}$$

2.4 Molecular modeling

2.4.1 Adiabatic maps

Adiabatic maps for methyl β-cellobioside and methyl β-laminarabioside were calculated with CHARMM[35] using the CSFF force field[36] and a relative dielectric constant of 1 which is the same as the value used in the development of the force field. The adiabatic maps were generated from the isolated molecules by Φ and Ψ scans in increments of 10 degrees. After each increment, the geometry was relaxed with constrained Φ and Ψ and the total energy was recorded. Adiabatic maps were generated for all 12 combinations of the following selected configurations of exocyclic groups: All hydroxyl groups in a clockwise (*c*) or counter-clockwise (*r*) cooperative crown arrangement around the pyranose ring, gauche-gauche (*gg,* ω = 300), gauche-trans (*gt,* ω = 60), and trans-gauche (*tg,* ω = 180) rotamers for the hydroxymethyl groups. To create the final composite adiabatic map, each point in a 36 x 36 (Φ, Ψ) grid was assigned the lowest energy found at corresponding points in the 12 maps pertaining to the different exocyclic configurations. The relaxed maps were contoured at increments of 2 kcal mol^{-1} above the global minima.

2.4.2 Molecular dynamics simulations

Four starting structures for methyl β-cellobioside and three starting structures for methyl β-laminarabioside corresponding to the minima in the adiabatic maps were generated by adjustment of Φ and Ψ of the lowest energy conformer. NVE ensemble molecular dynamics simulations at 300 K were carried out with the CHARMM program using the CSFF force field. The minimum image convention was used and switching functions were enabled to ensure that all long-range interactions converged smoothly to zero at 12 Å. The TIP3P[37] potential energy function was used to model the water molecules of the solvent. Newton's equations of motion were integrated with the velocity Verlet algorithm using a timestep of 1 fs. The SHAKE[38] algorithm was used to constrain the length of bonds involving hydrogen. The neighbor-list was updated every 10 steps. Initial geometries for all simulations were created by superimposing the starting conformation of the disaccharide upon the coordinates of a box of 512 water molecules equilibrated at 300 K. Water molecules were deleted if their van der Waals radii overlapped with any atoms of the solute. The simulation box contained 488 water molecules for all simulations of methyl β-laminarabioside, and 489 water molecules for the simulations of methyl β-cellobioside, except for the simulation started at minimum *C* in the adiabatic map which contained 485 water molecules. After solvation, the system was energy minimized with 50 steepest descent iterations with the original box size to relax any steric conflicts. Subsequently the cubic box length was adjusted slightly to provide a density of 1.00 $g \cdot cm^{-3}$. Initial velocities for all atoms were assigned from a Boltzmann distribution at 300 K. The system was equilibrated for 100 ps using velocity scaling if the average temperature deviated from 300 K by more than ± 3 K. Subsequently, the production runs were started. Simulations started from global minima were run for 50 ns while simulations started from local minima were run for 10 ns.

2.4.3 Analysis of MD trajectories

Analysis of the molecular dynamics trajectories was performed with in-house software. The behavior of the glycosidic dihedral angles Φ and Ψ was investigated for simulations started from all minima in the adiabatic maps. In addition, the 50 ns trajectories from the *A* minima were subjected to analysis of hydrogen bonding and hydration. The population of the space spanned by Φ and Ψ was visualized using contour plots of the two-dimensional histograms derived from the trajectories. Sugar inter- and intra-residue hydrogen bonds (H-bonds) were analyzed by searching the trajectories for acceptable hydrogen-acceptor distances (< 3.5 Å) and donor-hydrogen-acceptor angles (> 143°). Hydration was analyzed by establishing 2D pair distributions[39] for all unique combinations of solute oxygens, and subsequently mapping the water density with contour plots. Water densities above 1.0 times the bulk density of water occurring at distances shorter than 3.5 Å from both oxygens were considered

2.4.4 Helical parameters

As mentioned in the introduction, polysaccharides containing β(1→4) and β(1→3) linkages only form regular helical arrangements. These can be characterized in terms of the helical parameters n and h, where n is the number of repeating units per helical turn and h is the translation in the length direction of the helix. Following a procedure similar to that of Hansen *et al.*[40] and Pérez *et al.*[41] the helical parameters were calculated for all combinations of the glycosidic linkage torsions using the POLYS program.[42]

3. Results and discussion

3.1 Adiabatic maps

To allow unambiguous comparison with the MD simulations in this work, adiabatic maps for the two model compounds were generated with the CSFF force field. A "true" adiabatic map for a hexopyranose disaccharide should ideally be obtained from restrained optimizations of starting configurations corresponding to all 3^{10} combinations of the staggered orientations of the exocylic groups.[43] However, this task presents an insurmountable computational barrier and the present study therefore employs the simplifications mentioned in the experimental section. Although the incomplete sampling of the conformations of exocyclic groups may introduce very small artifacts, the present approach provides significant improvements over conformational maps based on simplifications such as the use of rigid monomeric residues.[44] The most energetically favorable values of Φ occurs around 300° (−gauche) and 60° (+gauche) for both model compounds, see **Figure 2**. As remarked by Petkowicz *et al.*[45] this is in agreement with results generally observed in conformational maps for equatorially linked disaccharides.

Four minima in (Φ, Ψ)-space are discernible in the adiabatic map for methyl β-cellobioside: *A*: (Φ = 300°, Ψ =280°), *B*: (Φ = 280°, Ψ =210°), *C*: (Φ = 260°, Ψ =60°), and *D*: (Φ = 60°, Ψ =260°). In the same order, the relative energies of the minima are 1, 2, 3, and 1 kcal mol^{-1} above the global minimum. Structures corresponding to the minima are shown in **Figure 3.** The minima *A*, *B*, and *C* are merged vertically by contour lines to form the characteristic elongated island in the right side of the map. This feature, as well as the overall appearance of the adiabatic map, is in accordance with published adiabatic maps for cellobiose obtained with other force fields, e.g. MM3.[9]

Three minima are found in the adiabatic map for methyl β-laminarabioside: *A*: (Φ = 290°, Ψ =130°), *B*: (Φ = 270°, Ψ =290°), and *C*: (Φ = 60°, Ψ =120°). In the same, order the relative

energies of the minima are 1, 3, and 4 kcal mol^{-1} above the global minimum. Structures corresponding to the minima are shown in **Figure 4.**

In the adiabatic maps, the crystal structures of the model compounds and their non-methylated parent sugars are indicated with filled and open circles, respectively. For both model compounds, the crystal structures occupy positions significantly below (i.e. they have smaller values of Ψ) the adiabatic map *A* minima. This can probably be attributed mainly to packing effects in the crystal. The crystal and adiabatic map values of Φ are in good agreement.

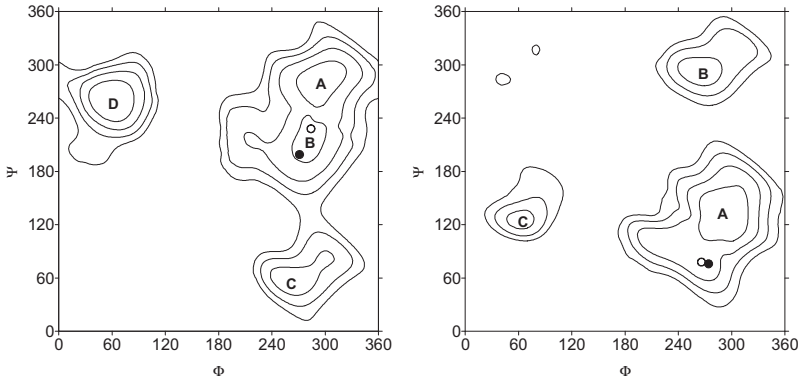


Figure 2: Adiabatic maps calculated with the CSFF force field for methyl β-cellobioside (left) and methyl β-laminarabioside (right). The energy is contoured in increments of 2 kcal mol^{-1} above the global minimum. Minima are indicated with uppercase letters. The crystal structures of methyl β-cellobioside and methyl β-laminarabioside are indicated with filled circles, while the crystal structures of cellobiose and laminarabiose are indicated with open circles.

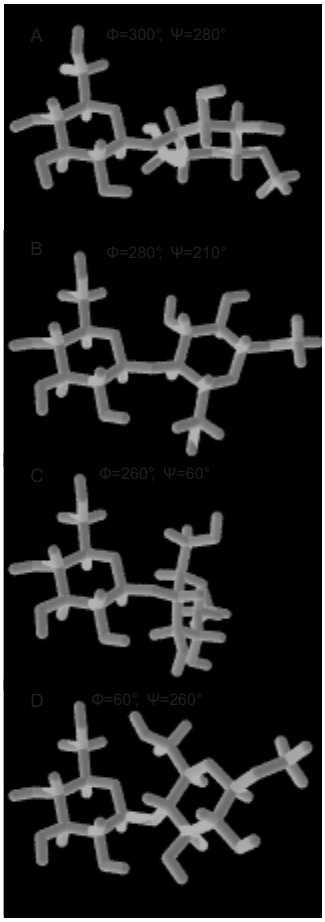


Figure 3: Structures corresponding to minima *A*, *B*, *C*, and *D* in the adiabatic map for methyl β-cellobioside (**Figure 2,** left). Values of the glycosidic dihedral angles Φ and Ψ are indicated.

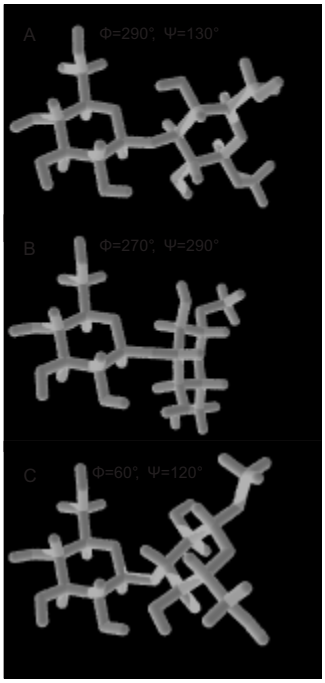


Figure 4: Structures corresponding to minima *A*, *B*, *C* in the adiabatic map for methyl β-laminarabioside (**Figure 2,** right). Values of the glycosidic dihedral angles Φ and Ψ are indicated.

3.2 Molecular dynamics simulations

3.2.1 Glycosidic dihedral angles

Methyl β-cellobioside

For the 50 ns simulation started at minimum *A* (300, 280) in the adiabatic map, the maximum density in (Φ, Ψ)-space is centered around approximately (280, 240), see **Figure 5**. The mean (Φ, Ψ) angles are practically identical with the MD simulation results from the study by Pereira *et al.*[20] (‹Φ, Ψ› = (283, 237)), see **Table 2**.

Table 2: Calculated (Karplus) and measured (NMR) heteronuclear $^3J_{CH}$-coupling coefficients in Hz. $^3J_{HC}$ and $^3J_{CH}$ refers to $^3J_{H1C4}$, and $^3J_{C1H4}$, for methyl β-cellobioside, and $^3J_{H1C3}$, and $^3J_{C1H3}$, for methyl β-laminarabioside. The mean values of Φ and Φ for the MD trajectories are reported and values from the corresponding crystal structures are included for comparison.

Compound	**Initial minimum**	**‹Φ, Ψ›**	**Karplus**		**NMR**	
			$^3J_{HC}$	**$^3J_{CH}$**	**$^3J_{HC}$**	**$^3J_{CH}$**
methyl β-cellobioside	A (50 ns)	281, 236	3.21	5.02	4.05	4.96
	B (10 ns)	288, 235	3.22	5.03		
	C (10 ns)	285, 78	3.14	5.44		
	D (10 ns)	58, 240	6.61	5.34		
	Cellobiose[a]	283, 237	2.99	5.58		
	Cellobiose[b]	284, 228	2.93	5.16		
	Cellobioside[c]	271, 199	3.99	3.24		
methyl β-laminarabioside	A (50 ns)	286, 134	2.99	4.65	3.15	4.12
	B (10 ns)	266, 295	4.36	6.58		
	C (2 ns)	60, 127	6.65	5.19		
	C (10 ns)	181, 129	4.68	4.92		
	Laminarabiose[d]	287, 163	2.74	2.83		
	Laminarabiose[e]	266, 78	4.46	3.56		
	Laminarabioside[f]	274, 76	3.88	3.42		
	Curdlan[g]	269, 127	4.34	5.44		
	Curdlan[h]	275, 125	3.83	5.50		

[a] Cellobiose MD results.[20]

[b] Cellobiose crystal structure.[46]

[c] Methyl β-cellobioside crystal structure.[47]

[d] Laminarabiose, MD results.[20]

[e] Laminarabiose crystal structure.[48]

[f] Methyl β-laminarabioside crystal structure.[49]

[g] Anhydrous curdlan, crystal structure.[3]

[h] Hydrated curdlan, crystal structure.[50]

These dihedral angles are also in accordance with values from the crystal structures for cellobiose[46] ((Φ, Ψ) = (284, 228)) and methyl cellobioside[47] ((Φ, Ψ) = (271, 199)), where interestingly cellobiose[46] provides the better match. Virtually all remaining (Φ, Ψ)-space density falls within low energy contours surrounding minima *A and B*. The adiabatic map minimum *A*, which corresponds to a nearly perpendicular orientation of pyranose rings (**Figure 3a**) remained nearly unpopulated throughout the simulation. The 10 ns simulation started at minimum *B* relaxed to simulation minimum *A'* within a few nanoseconds after which point the trajectory resembled that initiated from minimum *A*, see **Figure 6**. The simulation started at minimum *C* displayed a relatively high spread in Φ and Ψ during the first few nanoseconds, resulting in the oblong density around a simulation minimum *C'*. After approximately 3 ns an upward shift in Ψ was followed by relaxation to minimum *A'*. This transition implies surpassing a barrier of approximately 7 kcal mol^{-1}. **Figure 7** illustrates the time-development of (Φ, Ψ) with density maps obtained after 2, 4, and 10 ns trajectories started at minimum *C*. There is no reversal to the original minimum after 3 ns and the simulation populates (Φ, Ψ)–space in a manner identical to simulations started at minima *A* and *B*. The simulation started in minimum *D* gave rise to the (Φ, Ψ) –space density focused some degrees below (Ψ) the adiabatic map minimum shown in **Figure 8**, but no transitions were observed.

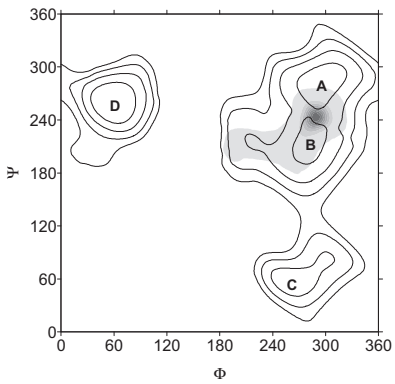


Figure 5: Methyl β-cellobioside density plot in (Φ, Ψ)-space for the 50 ns MD trajectory started at minimum *A*.

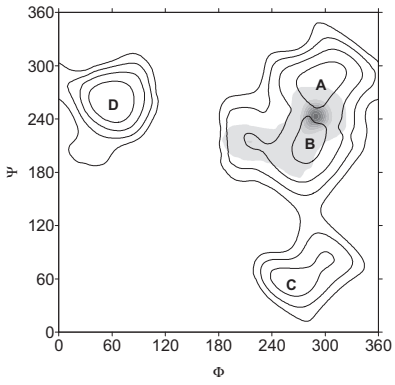


Figure 6: Methyl β-cellobioside density plot in (Φ, Ψ)-space for the 10 ns MD trajectory started at minimum *B*.

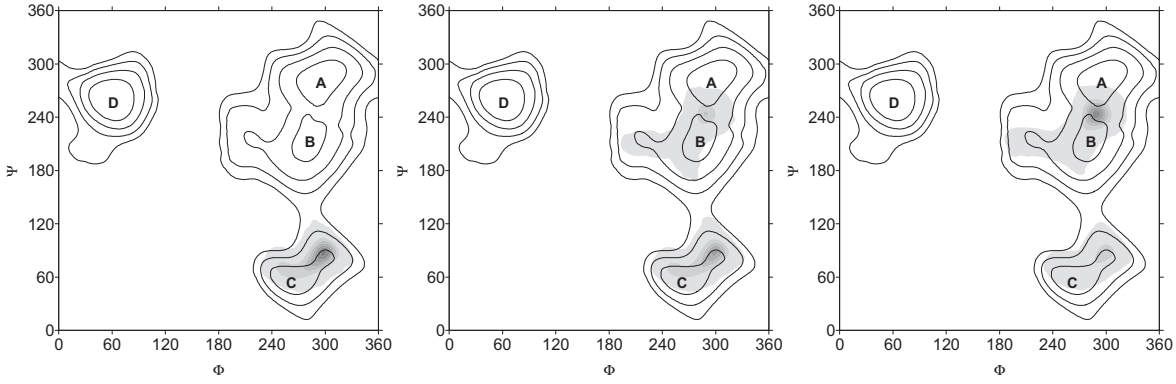


Figure 7: Methyl β-cellobioside density plots in (Φ, Ψ)-space after 2 ns (left), 4 ns (middle), and 10 ns (right) MD simulation started at minimum *C*.

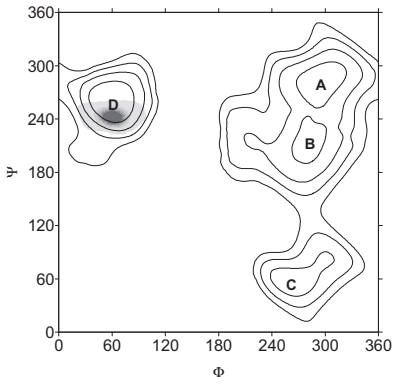


Figure 8: Methyl β-cellobioside density plot in (Φ, Ψ)-space for the 10 ns MD trajectory started at minimum *D*.

Methyl β-laminarabioside

The maximum (Φ, Ψ)-space density for the 50 ns simulation started at minimum *A* corresponds well with this adiabatic map minimum, and the majority of the density is enclosed by the 4 kcal mol^{-1} contour (**Figure 9, top**). Similarly, the simulation started at minimum *B* populates the area at this minimum in agreement with the low energy adiabatic map contours (**Figure 9, middle**). In contrast, the simulation started at minimum *C* initially populates this area, but shifted to minimum *A* after approximately 4.6 ns (**Figure 9, bottom**). This transition consists primarily of the 240° rotation in Φ and involves surpassing a barrier of significant magnitude (~12.5 kcal mol^{-1}). The part of the trajectory in (Φ, Ψ)-space involving the transition between the two simulation minima is superimposed on **Figure 9, bottom.**

Comparison of the average (Φ, Ψ) values obtained in the MD simulations in this work (**Table 2**) with crystal structures of laminarabiose[48] and methyl β-laminarabioside[49], gives the best agreement for the simulation initiated at minimum *A*, where ‹Φ, Ψ› = (286, 134). While the simulation average of this study is in good agreement with Pereira *et al.*[20] (‹Φ, Ψ› = (287, 163)), the difference between the simulation average (and adiabatic map) value of Ψ and the values from the crystal

structures[48,49] (mean 77°) is significant. Both anhydrous curdlan ((Φ, Ψ) = (269, 127))[3] and hydrated curdlan ((Φ, Ψ) = (275, 125))[50] show good agreement for both dihedrals. Interestingly it is the anhydrous version that presents the best correspondence between the average value of Ψ observed in the present simulations and an experimental value for this dihedral.

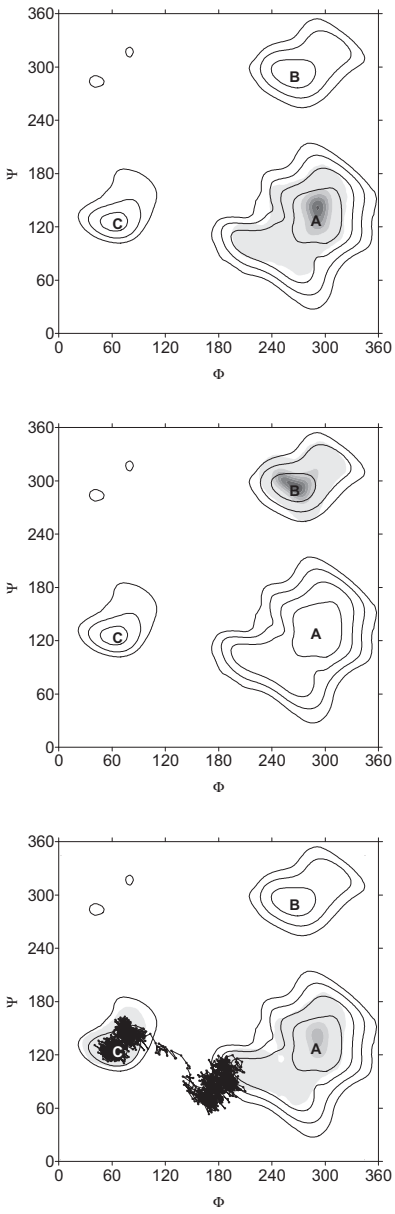


Figure 9: Methyl β-laminarabioside density plots in Φ, Ψ space for **(top)** the 50 ns MD trajectory started at minimum *A*, **(middle)** the 10 ns MD trajectory started at minimum *B,* and **(bottom)** the 10

ns MD trajectory started at minimum *C*. Part of the trajectory involving the transition from minimum *C* to *A* is superimposed on the bottom figure.

3.2.2 Other dihedral angles

The distributions of the dihedral angles ω_i of the hydroxymethyl group in each residue (**Table 3**) generally resemble the rotameric distribution reported by Kuttel *et al.* for 10 ns simulations of β-D-glucose.[36] These results are also in good agreement with average reported experimental values for β-D-glucose.[51,52] The only significant deviation from free monomer behavior occurs for ω_2 in methyl β-cellobioside where the *gg* rotamer is populated for approximately 91 percent of the time. This reduction of conformational freedom is probably a consequence of the position of the hydroxymethyl group next to the glycosidic linkage in methyl β-cellobioside.

Table 3: Distribution of orientations of the hydroxymethyl dihedral angles (ω_i) in the 50 ns trajectories for methyl β-cellobioside (B14) and methyl β-laminarabioside (B13).

	B14		**B13**		**Exp**[a]	**CSFF**[b]
	ω_1	**ω_2**	**ω_1**	**ω_2**	**ω**	**ω**
gg	68	91	67	61	52.5	69
gt	32	9	33	39	43.0	31
tg	0	0	0	0	4.5	0

[a] Average β-D-glucose values.[51,52]

[b] Distribution from 10 ns simulation of β-D-glucose.[36]

There is a bimodal distribution for the methoxy dihedrals 2O5-2C5-2O1-2C7 in both model compounds. The first mode has a mean of 64 degrees and is populated for 37 and 16 percent of the

simulation time for methyl β-cellobioside and methyl β-laminarabioside, respectively. The second mode has a mean of 284 degrees and is populated for the remaining simulation time. Transition between these rotamer populations occurred only twice for each compound on the 50 ns timescale of the simulations.

3.2.3 Intramolecular hydrogen bonds

Methyl β-cellobioside

The most important H-bond in the MD simulation started from minimum *A* occurs between 1O5 and 2HO3 and persisted for 35 percent of the simulation time (**Table 4**). This H-bond bridges the two residues in the crystal structure of cellobiose[53,54] where, according to Jacobsen, it appears to be important in establishing the relative orientation of the two rings relative twisted by approximately 26 degrees. In the study by Pereira *et al.*[20] this hydrogen bond persisted for 70 percent of the simulation time, with their less restrictive geometrical definition of the presence of H-bonds (hydrogen-acceptor distance < 2.5 Å and donor-hydrogen-acceptor angle > 135°). Based on several experimental and computational demonstrations of the 1O5···2HO3 bond, Pereira and coworkers stressed the well-established role of this hydrogen bond in determining the stability, rigidity and insolubility of cellulose. Additional H-bonds of shorter duration are observed in the MD trajectory between the primary alcohol group on one glucose residue and the hydroxyl group neighboring the glycosidic linkage on the other glucose residue. Such interactions are compatible with a relatively broad range of glycosidic dihedral angles. The simulation started at minimum *D* can be stabilized by two internal H-bonds. The most important of these bonds occurred between 1O2 and 2HO3 and persisted throughout the 10 ns simulation. The second important H-bond is more transient but recurrent and takes place between the primary hydroxyl groups of the two pyranose rings, with the two groups acting alternately as donors and acceptors.

Table 4: Methyl β-cellobioside intramolecular H-bonds from the MD trajectory started at minimum *A* in the adiabatic map. Duration is given as percentage of the simulation time.

Donor	Acceptor	Duration
2HO3	1O5	35
1HO6	2O3	5
2HO3	1O6	4
1HO2	2O6	3
2HO6	1O2	1
1HO4	1O6	<1

Methyl β-laminarabioside

The proposition that β(1→3)-linked polysaccharides may exhibit intramolecular hydrogen bonds between 1O4 and 2O5 or between 1O2 and 2O2 was put forward by Sundaralingam[55] already in 1968, and is in good agreement with the findings in the present and other studies.[6,20] The glycosidic dihedral angles of the crystal structures[48,49] imply that the pyranose rings are approximately parallel, facilitating the formation of a 2.77 Å hydrogen bond between 2HO4 and 1O5. The average glycosidic dihedral angles from the MD simulations allow for an approximately "in-line" arrangement of the atoms 2O4, 2HO4 and 1O5. With the glycosidic dihedral angles of the crystal structures, such an arrangement would result in a steric clash between 2HO4 and 1O5. Indeed the most persistent hydrogen bond in the present simulations was the bond between 2HO4 and 1O5, which was present for 15 percent of the simulation time. In the study by Pereira *et al.*[20], this bond persisted for 42 percent of the simulation time, with their H-bond definitions (see above). The preferred simulation population of the dihedral angles is associated with additional inter-residue H-bonds between 2HO2 and 1O2 and between 1HO2 and 2O2 since these atoms are in neighboring

groups to the glycosidic linkage. Less significant (~5% of the simulation time) contributors are the H-bonds between the hydroxymethyl group on 1C6 and the hydroxyl group on 2C4.

Table 5: Methyl β-laminarabioside intramolecular hydrogen bonds from the MD trajectory started at minimum *A* in the adiabatic map. Duration is given as percentage of the simulation time.

Donor	Acceptor	Duration
2HO4	1O5	15
2HO2	1O2	8
1HO6	2O4	4
2HO4	1O6	1
1HO2	2O2	1
1HO4	1O6	<1

For both model compounds, the only indication of intra-residue hydrogen bonding is observed in case of the very transient H-bond between the hydroxyl groups at position 4 and 6 on ring 1.

3.2.4 Hydration

The results of the hydration analysis of the MD trajectories are charted in **Figure 10** where the densities are given as multiples of the bulk density of water. The simulation started from minimum *A* is considered for each model compound, and only densities above 1 are shown. Considering both compounds, water densities in the range 1.4 – 3.5 occur between most oxygens separated by more than three bonds. In addition, significant densities are found between oxygens separated by more than three bonds. These situations occur between residues, giving rise to inter-residue water-bridges. There are two such bridges in case of methyl β-cellobioside and three in case of methyl β-laminarabioside. For both compounds, the highest water density is found in the inter-residue water-bridge involving the primary alcohol group on ring 1. The second group in the bridge is the hydroxyl group on 2C3 in case of methyl β-cellobioside and on 2C4 in case of methyl β-laminarabioside. Apparently these inter-ring water bridges in combination with the direct hydrogen

bonds (see above) between the same groups are determining the detailed overall solution structure of the disaccharide. For methyl β-laminarabioside, a second inter-residue zone of high water density (4.5) is found between 1O2 and 2O2 (**Figure 10, bottom**). The involved hydroxyl groups also form significant H-bonds with each other. Thus two types of stabilizing inter-residue interactions, hydrogen bonds and water bridges, co-occur between groups in both model compounds when they assume the favored dihedral configurations.

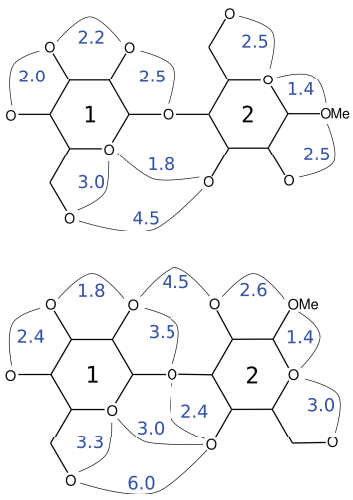


Figure 10: Significant water densities found between pairs of oxygens in methyl β-D-cellobioside (top) and methyl β-D-laminarabioside (bottom). The analysis was carried out on the 50 ns MD trajectories started in the *A* minima in the adiabatic maps. Densities are reported as multiples of the bulk density of water.

3.3 NMR spectroscopy

3.3.1 Heteronuclear coupling coefficients

For methyl β-cellobioside the best agreement between measured and calculated glycosidic heteronuclear J-coupling constants is found for the simulation started at minimum *A*, see **Table 2**. Although the calculated value of $^{3}J_{HC}$ is 0.84 Hz smaller than the experimental value, the values of

J_{CH} are virtually identical. For minimum *D,* the only other stable minimum for methyl β-cellobioside, there is reasonable agreement between the calculated and measured values of $^3J_{CH}$, but not for $^3J_{HC.}$

For methyl β-laminarabioside, minimum *A* provides the best correspondence between measured and calculated values of J_{HC} and $J_{CH.}$ The correspondence is poor between measured and calculated coupling constants for the other stable minimum, *B*.

3.2.2 Nuclear Overhauser effect

A series of *medium* (*m*) and *weak* (*w*) intensity non-neighboring NOE's are observed for the two model compounds (**Table 6**). Both inter- and intra-residue NOEs are observed, where the number of the latter type is larger in methyl β-cellobioside than in methyl β-laminarabioside. All inter- and intra-residue NOEs involving 1H5 or 2H5 are of *weak* intensity in methyl β-cellobioside. In methyl β-laminarabioside, such signals are absent. This suggests a special status of the hydrogen on C5. For both sugars, there are inter-residue NOEs between 1H1 and the three ring hydrogens closest to the glycosidic linkage on residue 2. Their intensities are *medium* (except in case 1H1-2H5 for methyl β-cellobioside which is *weak*) indicating restricted rotation around the glycosidic linkage torsions. This is in good qualitative agreement with the stabilizing inter-residue hydrogen bonding and hydration effects discussed in the previous paragraphs. The strong preference of the hydroxymethyl group in residue 2 in methyl β-cellobioside for the *gg* rotamer was mentioned in the discussion of dihedral angles. This implies a relatively fixed positioning of one of the 2H6 hydrogens at a distance of 3 Å from 1H1, which supports the experimentally observed NOE between 1H1 and 2H6. In both compounds, a NOE of *medium* intensity is observed between 2H1 and methoxy hydrogens, indicating restrictions in the rotation of the methoxy group. This agrees

well with the preferred 286 degree orientation of the 2O5-2C1-2O1-2C7 dihedral observed in the MD trajectories which positions the methoxy carbon close (2.4 Å) to 2C1 in both compounds.

Table 6: NOEs measured for methyl β-cellobioside (B14) and methyl β-laminarabioside (B13). The NOEs are characterized as either weak (*w*) or medium (*m*).

Atom	Atom	B14	B13
1H1	1H3	*m*	*m*
1H1	1H5	*w*	-
1H1	2H2	-	*m*
1H1	2H3	*m*	*m*
1H1	2H4	*m*	*m*
1H1	2H5	*w*	-
1H1	2H6	*m*	-
2H1	METH	*m*	*m*
2H1	2H3	*m*	*m*
2H1	2H4	*m*	-
2H1	2H5	*w*	-

3.3 Implications for polymer structure

3.3.1 Helical parameters

The compatibility of the favored glycosidic dihedral populations observed in the simulations of the model compounds with the dihedral configurations adopted in helical structures is visualized in **Figure 11.** Each figure is a superimposition of contour plots of the helical parameters n and h (created using the procedure of section **2.2.4**), the adiabatic map and the (Φ, Ψ) density plot for the 50 ns MD simulation started at the *A* minimum of that adiabatic map. As discussed by Pérez and Mackie[56], x-ray diffraction studies of the pure β(1→4) polymer (cellulose) reveals a highly extended structure as a consequence of the $^4C_1$ chair conformation and the β(1→4) glycosidic linkage. The parameters observed from the x-ray diffraction pattern are (n = 2, h = 5.18 Å). The intersections of these contours are enveloped by the contour 4 kcal mol^{-1} above the global minimum in the adiabatic map for methyl β-cellobioside. However, the most populated area in the 50 ns

simulation is focused near the intersection of the helical contours (n = 2.8, h = -5 Å), see **Figure 11**, **left**. Thus extrapolation from the simulation results suggests an even more extended left-handed polymer arrangement.

The crystal structures of the hydrated and the anhydrous pure β(1→3) glucose polymer (curdlan) reveals in both cases a right handed triple helix arrangement, where each strand has six-fold symmetry. In the anhydrous state the advancement per monomer is 2.935 Å. Due to the parallel, in-phase arrangement along the fiber axis, the helix repeat unit is two glucose monomers (5.87 Å).[3] Thus, conformations in agreement with these crystal structures are found at the intersection of the contours (n = 6 and h = 2.935 Å). Although this intersection is found within the contour 2 kcal mol^{-1} above the global minimum in the adiabatic map of methyl β-laminarabioside, the most populated area in the 50 ns simulation is focused near the intersection of the helical contours (n = 4.5 and h = 4 Å), see **Figure 11**, **right**. Thus the preferred population of the glycosidic dihedral angles in the simulation is not fully compatible with the triple helix arrangement observed for curdlan in the solid state.

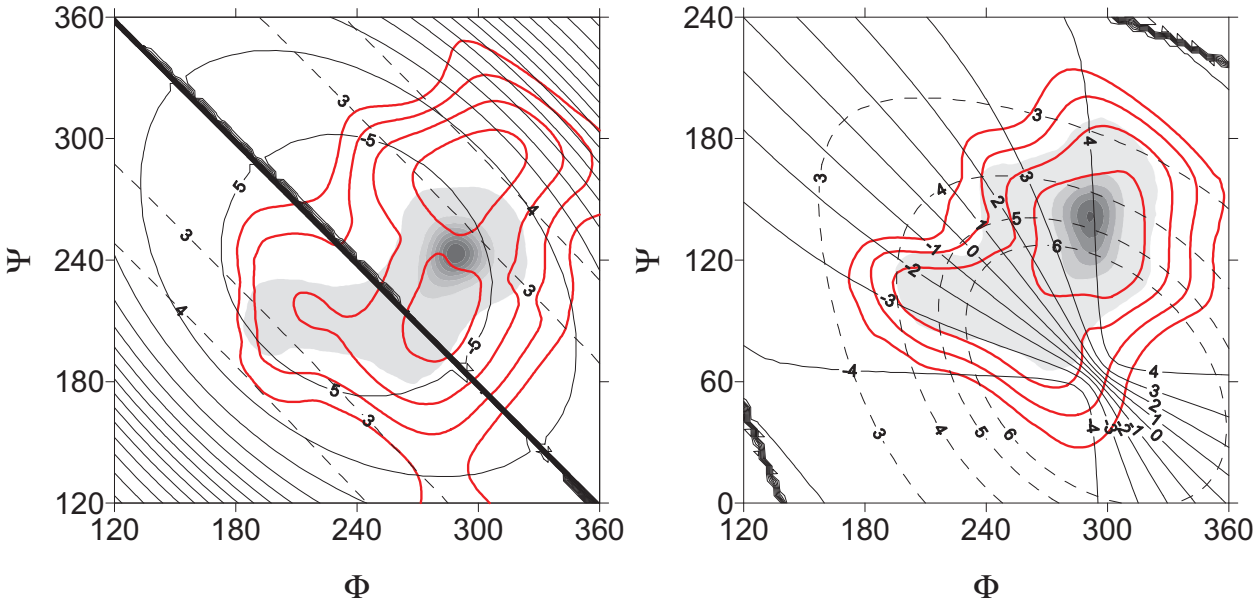


Figure 11: Superimposition of helical parameters, the adiabatic map, and the (φ, ψ) density plot of the 50 ns MD simulation for methyl β-cellobioside (left) and methyl β-laminarabioside (right). The helical parameters n (dashed contours) and h (solid contours) represent the number of repeats per helical turn and the helical repeat advancement in Å, respectively. Right-handed and left-handed helices are characterized by $h > 0$ and $h < 0$, respectively.

3.3.2 Extrapolation to polymer structure

Using lichenase degradation, it has been shown that mixed linkage BG to a large extent are composed of cellotriosyl and cellotetraosyl units linked together by single β(1→3) linkages.[57,58] Using the average glycosidic dihedral angles from the global simulation minima ((<Φ, Ψ> = 281, 236) from methyl β-cellobioside and (<Φ, Ψ> = 281, 236) from methyl β-laminarabioside) examples of regular and mixed linkage BG were generated using the POLYS program, see Figure 12.

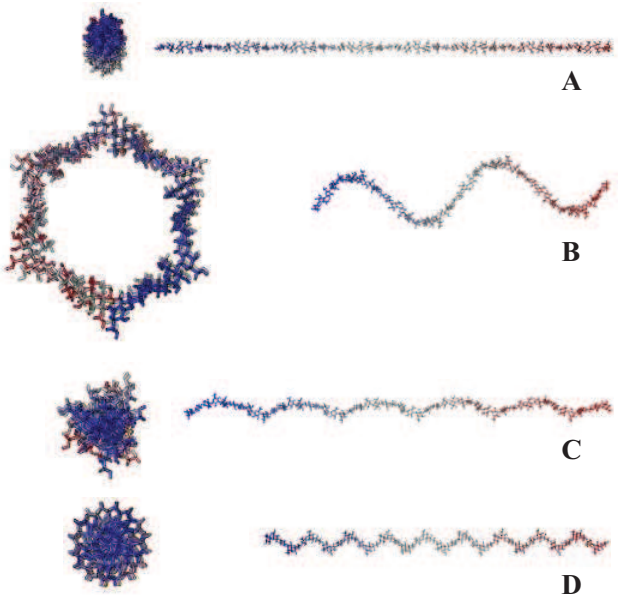


Figure 12: Models of β-glucans (48-mers) generated using the average values of glycosidic dihedral angles observed at the global MD simulation minima. (A) The cellulose polymer {G4}, (B) the mixed-linkage {G4G4G4G3} polymer, (C) the mixed linkage {G4G4G3} lechinan polymer, and (D) the laminaran polymer {G3}. Right panel: Side view with the major principal axis of each polymer horizontally aligned. Left panel: Top view from the reducing end of each polymer along the major principal axis. The top view of (B) deviates slightly from the view along the major principal axis c, since the present depiction allows for better appreciation of the six-fold symmetry. The molecular models are included as Supplementary Data and are also available for visualization and download from www.models.life.ku.dk\carbs\polymers.

The polymer examples are built from 48 D-glucopyranose monomers and include (A) the pure cellulose polymer {G4}, (B) the mixed linkage lichenan BG with repeating cellotriosyl structure {G4G4G3}, (C) the mixed linkage BG with repeating cellotetraosyl structure {G4G4G4G3} and (D) the pure laminaran polymer {G3} (in which G is β-D-Glcp, 3 and 4 indicate (1→3) and (1→4) linkages and the brackets truncate the regular repeat units). Inspection of the four generated

molecular models reveals that the pure cellulose polymer adopts an extended ribbon-like (2-fold) helical structure which will aggregate into fibres. The introduction of G3 introduces kinks in the polymer chain which will prevent the polymer from aggregating and provide a more water soluble polymer. In the case of the regular cellotetraosyl structure, the BG structure takes an open six-fold helical form. Chain-chain interactions between such structures seem ideal for introducing gel-like effects in the plant cell walls and for binding affinity to hydrophobic substances such as bile acids and carcinogenic substances in the intestines.

In conclusion, the existence of two stable minima was demonstrated for the molecular dynamics simulations of methyl β-cellobioside and methyl β-laminarabioside. The global simulation minimum (*A'*) is characterized by the average glycosidic dihedral angles (Φ = 281, Ψ = 236) for methyl β-cellobioside and (Φ = 286, Ψ = 134) for methyl β-laminarabioside. These values are in good agreement with the corresponding crystal structures. Heteronuclear $^{3}J_{COCH}$ coupling constants were measured using NMR spectroscopy and compared to coupling constants calculated from the MD trajectories using a generalized Karplus relationship. The best agreement was found for the simulation minima *A'* suggesting that the average simulation dihedral angles for these minima are representative of the model compounds in aqueous solution. Inter-residue NOEs measured in both compounds between 1H1 and the ring hydrogens closest to the glycosidic linkage on residue 2 supported the stabilizing inter-residue hydrogen bonding and hydration effects observed in the simulation trajectories at the global minima. The preferred glycosidic dihedral angles observed in the aqueous state simulations were not highly compatible with the corresponding polymer arrangements observed in the solid state. Apparently inter-residue water bridges in combination with the direct hydrogen bonds between the same groups dictated the detailed overall solution structure of the disaccharides. Nevertheless, the average glycosidic dihedrals at the simulation

minima (*A*') allowed for the construction of BG polymer models with relatively well-defined secondary structure. Such polymer models are starting points for the understanding of higher levels of BG organization which is crucial to the elucidation of interactions between BG and smaller molecules.

Acknowledgements

We thank the programs Build Your Food supported by the Danish Council for Strategic Research, Program Commission of Health, Food and Welfare, and the strategic research initiative BEST funded by the University of Copenhagen for financial support.

Literature cited:

1. Sathyanarayana, B. K.; Rao, V. S. R. *Biopolymers* **1971**, *10*, 1605-1615.

2. Marchessault, R.H. and Sundararajan, P.R. Cellulose. In *The Polysaccharides*; Aspinall, G. O., Ed.; Academic Press: New York, 1983; Vol. 2, pp. 11-95.

3. Deslandes, Y.; Marchessault, R. H.; Sarko, A. *Macromolecules* **1980**, *13*, 1466–1471.

4. Morgan, K. R.; Ofman, D. *J. Cereal Chem.* **1996**, *75*, 879-881.

5. Wood, P. Cereal β-Glucans: Structure, Properties and Health Claims. In *Advanced Dietary Fibre Technology*; McCleary, B., Prosky, L., Eds.; Blackwell Science, Oxford, England, 2001, pp 315-327.

6. Kony, D.; Damm, W.; Stoll, S.; Hünenberger, P. H. *J. Phys. Chem. B.* **2004**, *108*, 5815–5826.

7. Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551-8566.

8. Allinger, N. L.; Rahman, M.; Lii, J.-H. *J. Am. Chem. Soc.* **1990**, *112*, 8293-8307.

9. Dowd, M. K.; French, A. D.; Reilly, P. J. *Carbohydr. Res.* **1992,** *233*, 15-34.

10. French, A. D.; Kelterer, A.-M.; Johnson, G. P.; Dowd, M. K.; Cramer, C. J. *J. Mol. Graphics Modell.* **2000**, *18*, 95-107.

11. French, A. D.; Johnson, G. P. *Cellulose* **2004**, *11*, 449–462.

12. French, A. D.; Johnson, G.P. *Can. J. Chem.* **2006**, *84*, 603–612.

13. Hardy, B. J.; Sarko, A. *J. Comput. Chem.* **1993**, *14*, 831–847.

14. Hardy, B. J.; Sarko, A. *J. Comput. Chem.* **1993**, *14*, 848–857.

15. Frecer, V.; Rizzo, R.; Miertus, S. *Biomacromolecules* **2000**, *1*, 91-99.

16. Umemura, M.; Hayashi, S.; Nakagawa, T.; Yamanaka, S.; Urakawa, H.; Kajiwara, K. *J. Mol. Struct. THEOCHEM.* **2003**, *624*, 129–144.

17. Kony, D.; Damm, W.; Stoll, S.; van Gunsteren, W. F. *J. Comput. Chem.* **2002**, *23*, 1416-1429.

18. Palleschi, A.; Bocchinfuso, G.; Coviello, T.; Alhaique, F. *Carbohydr. Res.* **2005**, *340*, 2154-2162.

19. Lins, R. D.; Hünenberger, P. H. *J. Comput. Chem.* **2005**, *26*, 1400-1412.

20. Pereira, C. S.; Kony, D.; Baron, R.; Müller, M.; van Gunsteren, W. F.; Hünenberger, P. H. *Biophys. J.* **2006**, *90*, 4337–4344.

21. Lipkind, G. M.; Shashkov, A. S.; Kochetkov, N. K. *Carbohydr. Res.* **1985**, *141*, 191–197.

22. Cheetham, N. W. H.; Dasgupta, P.; Ball, G. E. *Carbohydr. Res.* **2003**, *338*, 955–962.

23. Pérez, S.; Imberty, A; Engelsen, S. B.; Gruza, J.; Mazeau, K.; Jimenez-Barbero, J.; Poveda, A.; Espinosa, J. F.; van Eyck, B. P.; Johnson, G.; French, A.; Kouwijzer, M. L. C. E.; Groothenius, P. D. J.; Bernardi, A.; Raimondi, L. ; Senderowitz, H.; Durier, V.; Vergoten, G.; Rasmussen, K. *Carbohydr. Res.* **1998**, *314*, 141-155.

24. Hemmingsen, L.; Madsen, E. D.; Esbensen, A. L.; Olsen, L.; Engelsen, S. B. *Carbohydr. Res.* **2004**, *339*, 937-948.

25. Motawia, M. S.; Olsen, C. E.; Enevoldsen, K.; Marcussen, J.; Møller, B. L. *Carbohydr. Res*. **1995**, *277*, 109-123.

26. Yoneda, Y.; Kawada, T.; Rosenau, T.; Kosma, P. *Carbohydr. Res*. **2005**, *340*, 2428-2435.

27. Williard, J. J.; Brimacombe, J. S.; Brueton, R. P. *Can. J. Chem.* **1964**, *42*, 2560-2567.

28. Veeneman, G. H.; van Leeuwen, S. H.;. van Boom, J. H. *Tetrahedron Lett.* **1990**, *31*, 1331-1334.

29. Bax, A.; Davis, D. G. *J. Magn. Reson.* **1985,** *65*, 355-360.

30. Wilker, W.; Leibfritz, D.; Kerssebaum, R.; Bermel, W. *Magn. Res. Chem.* **1993**, *31*, 287-292.

31. Meissner, A.; Sørensen, O. W. *Magn. Res. Chem.* **2001**, *39*, 49-52.

32. Goddard, T. D.; Kneller D. G. SPARKY 3, University of California, San Francisco.

33. Tvaroska, I.; Hricovini, M.; Petrakova, E. *Carbohydr. Res.* **1989**, *189*, 359-362.

34. Mulloy, B.; Frenkiel, T. A.; Davies, D. B. *Carbohydr. Res.* **1988**, *184*, 39-46.

35. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217.

36. Kuttel, M.; Brady, J. W.; Naidoo, K. *J. Comput. Chem***. 2002,** *23,* 1236-1423.

37. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.

38. Ryckaert, J.-P. *Mol. Phys.* **1985**, *55*, 549-556.

39. Andersson, C.; Engelsen, S. B. *J. Mol. Graphics Modell.* **1999**, *17*, 101-105.

40. Hansen, P. I.; Spraul, M; Dvortsak, P.; Larsen, F. H.; Blennow, A.; Motawia M. S.; Engelsen, S. B. *Biopolymers.* **2009**, *91*, 179-193.

41. Pérez, S.; Kouwijzer, M.; Mazeau, K.; Engelsen, S.B. *J. Mol. Graphics* **1996**, *14*, 307-321.

42. Engelsen, S. B.; Cros, S.; Mackie, W.; Pérez, S. *Biopolymers* **1996**, *39*, 417-433.

43. Stortz, C. A. *Carbohydr. Res.* **1999**, *322,* 77-86.

44. French, A. D. *Carbohydr. Res.* **1989**, *188*, 206-211.

45. Petkowicz, C. L. O.; Reicher, F.; Mazeaub, K. *Carbohydr. Polym.* **1998**, *37*, 25–39.

46. Chu, S. S. C.; Jeffrey, G. A. *Acta. Cryst.* **1968**, *B24*, 830-838.

47. Ham, J. T.; Williams, D. G. *Acta Crystallogr., Sect. B: Struct. Sci*. **1970**, *26*, 1373-1383.

48. Takeda, H.; Noritake, Y.; Kasai, N. *Carbohydr. Res.* **1977**, *53*, 137- 152.

49. Noguchi, K.; Okuyama, K.; Kitamura, S.; Takeo, K. *Carbohydr. Res.* **1992**, *237*, 33-43.

50. Chuah, C. T.; Sarko, A.; Deslandes, Y.; Marchessault, R. H. *Macromolecules*, **1983,** *16*, 1375-1382.

51. Bock, K.; Guzman, J. B. F.; Ogawa, S. *Carbohydr. Res.* **1988**, *174*, 354-359.

52. Nishida, Y.; Ohrui, H.; Meguro, H. *Tetrahedron Lett.* **1984**, *25*, 1575-1578.

53. Brown, C. J. *J. Chem. Soc. A*, **1966**, 927-932.

54. Jacobson, R. A.; Wunderlich, J. A.; Lipscomb, W. N. *Acta Cryst.* **1961**, *14*, 598-607.

55. Sundaralingam, M. *Biopolymers*, **1968**, *6*, 189-213.

56. Pérez, S. and Mackie, W. Structure and morphology of cellulose. In: CERMAV-CNRS, 2001. Chapter IV. http://www.cermav.cnrs.fr/glyco3d/lessons/cellulose/index.html

57. Staudte, R.G.; Woodward, J. R.; Fincher, G. B.; Stone, B. A. *Carbohydr. Polym.* **1983**, *3*, 299–312.

58. Wood, P. J.; Weisz, J.; Blackwell, B. A. *Cereal Chem.* **1991**, *68*, 31–39.

Paper II

Niels Johan Christensen, Susana Murtinheira da Trindade Leitão, Mikael Agerlin Petersen, Birthe Møller Jespersen, and Søren Balling Engelsen. *A quantitative structure-property relationship study of the release of some esters and alcohols from barley and oat β-glucan matrices*. Journal of Agricultural and Food Chemistry, 2009, 57(11), 4924-4930, **Published.**

4924 J. Agric. Food Chem. 2009, 57, 4924–4930
DOI:10.1021/jf803799z

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY
ARTICLE

A Quantitative Structure−Property Relationship Study of the Release of Some Esters and Alcohols from Barley and Oat β-Glucan Matrices

Niels Johan Christensen,* Susana Murtinheira da Trindade Leitão, Mikael Agerlin Petersen, Birthe Møller Jespersen, and Søren Balling Engelsen

Quality and Technology, Department of Food Science, University of Copenhagen, Rolighedsvej 30, 1958 Frederiksberg C, Denmark

This study investigates the release of selected strawberry flavor compounds from aqueous solutions of two barley and oat β-glucan products at concentrations of 5, 10, and 15% (w/w). The flavor release of 12 esters and 3 alcohols was measured by dynamic headspace GC-MS. For each compound the ratio of the flavor release from the β-glucan solution to the release from aqueous solution, A_{rel}, was recorded. In general, esters were retained in the β-glucan matrices in a mass-dependent manner where heavier molecules were retained more. A_{rel} for alcohols was found to be significantly larger than for the esters. Whereas A_{rel} values for esters were always below unity, this parameter was above unity for alcohols in some preparations of β-glucan. This implies that relative to esters, alcohols were rejected from some matrices. An increase in the concentration of the β-glucan products was associated with an increased retention of alcohols and esters. For solutions of oat and barley β-glucan products at the same concentration, the oat product retained the flavor compounds more strongly. This difference was more pronounced at low concentrations of the β-glucan products. To investigate the potential of a multivariate approach for the analysis of the flavor release from β-glucan products, partial least-squares regression was employed on a large selection of calculated molecular descriptors, yielding simple QSPR models capable of explaining the variation in A_{rel}. The robustness of the QSPR models was verified by cross-validation and permutation tests. The results indicate that the multivariate modeling approach might provide a useful tool for the investigation of flavor release systems similar to those studied here.

KEYWORDS: β-Glucan; QSPR; PLS; flavor release; GC-MS; strawberry

INTRODUCTION

β-Glucans (BG) are hydrocolloid-forming dietary fibers found in cereal grains, particularly oat and barley. In addition to their application as a texture enhancer in the food industry, β-glucans provide a significant range of health benefits, including promoting cardiovascular health, normalizing blood glucose levels, promoting weight loss, and enhancing immune system function (*1*). Structurally, β-glucans are linear polysaccharides of glucosyl residues connected by β-(1→3) and β-(1→4) linkages. β-Glucans in oats and barley are similar in structure, but differences in the ratio of β-(1→3) and β-(1→4) linkages, molecular weight, and possibly solubility have been reported (*2*−*4*). The oat β-glucan molecular mass can reach 3×10^6 Da, whereas the molecular mass for barley β-glucan usually is typically $(2-2.5) \times 10^6$ Da (*5*). The addition of polysaccharides to foods can modify the rate and intensity of flavor release through binding effects on volatile compounds and changes in viscosity (*6*−*8*). To predict the effect of texture agent addition and, if possible, to design the flavor release profiles of the foodstuff, a detailed understanding of the interactions between flavor compounds and texturing agent is needed. Flavor release depends not only on the nature of the food matrix but also on the structure of the volatile compounds. It is therefore desirable to choose a range of volatile compounds when flavor release from a matrix is studied so that a more global picture of the phenomenon can be obtained (*9*). Because the release of a volatile compound is a function of the molecular structure of that compound, the release phenomenon lends itself well to the methodology of quantitative structure−property relationships (QSPR). Examples of previous work in this area include QSPR models for the release of volatile compounds from solutions of sucrose (*10*), ι-carrageenan matrices (*11*), and a model dairy gel (*12*). In general, the partitioning of volatile compounds between food matrix and vapor phase has been found to be highly dependent on the hydrophobicity of the compounds (*13*). In this study we combine dynamic headspace gas chromatography−mass spectrometry (GC-MS) and QSPR to investigate the release of volatile compounds from

*Corresponding author (telephone 353 33506; fax 353 33245; e-mail: njc@life.ku.dk).

 Published on Web 04/16/2009

preparations of two commercially available barley and oat β-glucan products.

MATERIALS AND METHODS

β-Glucans. Two commercially available soluble β-glucans, a barley β-glucan and an oat β-glucan, were chosen for this study. The barley β-glucan was Glucagel, which is extracted from hull-less barley. The product was obtained from GraceLinc Ltd. (Christchurch, New Zealand). This β-glucan has a declared content of ≥75% β-glucan, <18% starch, <10% moisture, <5% protein, <2% ash, and <2% fat. The β-glucan is of moderate molecular mass [(0.12−0.18) × 10^6 Da]. The oat β-glucan was PromOat, obtained from Biovelop (Kimstad, Sweden). This β-glucan has a declared content of 30−40% β-glucan, 6% pentosans, 49% carbohydrates (described as dextrins by the supplier), 4.5% moisture, <2.5% protein, 3.5% ash, and 0.5% fat. The supplier reports a molecular mass of 1.0 × 10^6 Da, which characterizes PromOat as a high molecular weight β-glucan.

Strawberry Flavor. The BMN 42-3 model strawberry flavoring solution was obtained from Danisco. The solution has the following composition (in volume percent) designed to mimic natural strawberry flavor: 0.20% anisyl acetate, 0.20% benzyl alcohol, 0.40% *cis*-3-hexenol, 0.20% citronellyl acetate, 0.40% ethyl 2-methylbutanoate, 0.20% ethyl acetate, 1.61% ethyl butanoate, 0.40% ethyl hexanoate, 0.20% ethyl isopentanoate, 0.60% ethyl propanoate, 1.57% Furaneol, 0.60% γ-decalactone, 0.20% 1-hexanol, 0.20% hexyl acetate, 0.20% isopentyl butanoate, 0.20% isopentyl isopentanoate, 0.60% methyl cinnamate, 0.20% *trans*-2-hexenol, and 91.82% propylene glycol (solvent).

Sample Preparation. Mixtures of water and β-glucan products were prepared in three concentrations, containing 5, 10, and 15% by weight, respectively, of the commercial β-glucan product (PromOat or Glucagel). The water/β-glucan mixtures were transferred to glass beakers, and solubilization was promoted by magnetic stirring of the mixture for 30 min at 80 °C. Immediately after solubilization, 19 mL of the gel solution (still liquid and hot) was transferred to separate magnetically stirred headspace vessels, and the temperature was monitored until gelatinization. Next, the addition of flavor solution was carefully timed to ensure a homogeneous distribution of flavor compounds in the hot gel solutions while at the same time minimizing the loss of volatiles due to the elevated temperature. Immediately prior to gelatinization, 1 mL of a 0.02% aqueous solution of the BMN 42-3 model strawberry flavoring solution was added to each headspace vessel, yielding a final concentration of 0.001% of the strawberry solution in the gels. The vessels were capped immediately after the addition of flavor and stored in the refrigerator for a period of between 24 h and 1 week, according to the gelatinization time of the samples. Water reference samples were made by adding 1 mL of a 0.02% aqueous solution of the model strawberry solution to 19 mL of water. All samples were made in triplicate.

Dynamic Headspace GC-MS. Volatile compounds were collected on a Tenax-TA trap. The trap contained 250 mg of Tenax-TA with mesh size 60/80 and a density of 0.37 g mL^{-1} (Buchem bv, Apeldoorn, The Netherlands). The sample/suspension was equilibrated to 30 ± 1 °C in a circulating water bath and then purged with nitrogen (75 mL min^{-1}) for 30 min. The trapped volatiles were desorbed using an automatic thermal desorption unit (ATD 400, Perkin-Elmer, Norwalk, CT). Primary desorption was carried out by heating the trap to 250 °C with a flow (60 mL min^{-1}) of carrier gas (He) for 15.0 min. The stripped volatiles were trapped in a Tenax TA cold trap (30 mg held at 5 °C), which was subsequently heated at 300 °C for 4 min (secondary desorption, outlet split 1:10). This allowed for rapid transfer of volatiles to a gas chromatograph−mass spectrometer (GC-MS, G1800A GCD System, Hewlett-Packard, Palo Alto, CA) through a heated (225 °C) transfer line. Separation of volatiles was carried out on a 30 m DB-Wax capillary column with 0.25 mm internal diameter and 0.25 μm film thickness. The column flow rate was 1.0 mL min^{-1} using helium as a carrier gas. The column temperature program was 10 min at 45 °C, raised from 45 to 240 °C at 6 °C min^{-1}, and finally 10 min at 240 °C. The GC was equipped with a mass spectrometric detector operating in the electron ionization mode at 70 eV. Mass-to-charge ratios between 15 and 300 were scanned. Volatile compounds were identified by matching their mass spectra with those of a commercial database (Wiley275.L, HP product G1035A). The software program GCD Plus ChemStation G1074B (version A.01.00, Hewlett-Packard) was used for integrating chromatographic peaks.

Flavor Release Profiles. For each preparation (water and 5, 10, and 15% preparations of both types of β-glucan product) the mean and standard error for each chromatographic peak area were calculated from the triplicate chromatograms. The flavor release is defined as

$$A_{\mathrm{rel},i} = \frac{A_{\mathrm{BG},i}}{A_{\mathrm{water},i}} \quad (1)$$

where $A_{\mathrm{rel},i}$ is the release of compound i relative to water, $A_{\mathrm{BG},i}$ is the mean chromatographic peak area for compound i in the headspace above the β-glucan solution, and $A_{\mathrm{water},i}$ is the mean chromatographic peak area for compound i in the headspace above water. The standard error for $A_{\mathrm{rel},i}$ was estimated by

$$\frac{\Delta A_{\mathrm{rel},i}}{A_{\mathrm{rel},i}} = \sqrt{\left(\frac{\Delta A_{\mathrm{BG},i}}{A_{\mathrm{BG},i}}\right)^2 + \left(\frac{\Delta A_{\mathrm{water},i}}{A_{\mathrm{water},i}}\right)^2} \quad (2)$$

where $\Delta A_{\mathrm{rel},i}$ is the estimated standard error in $A_{\mathrm{rel},i}$, $\Delta A_{\mathrm{BG},i}$ is the estimated standard error in $A_{\mathrm{BG},i}$, and $A_{\mathrm{water},i}$ is the standard error in $A_{\mathrm{water},i}$. The compound index i is left out in the remainder of the paper for notational convenience.

Molecular Modeling. Structures of the 15 strawberry flavor compounds detected with dynamic headspace GC-MS were built with Arguslab (*14*), and conformational analysis was performed with MacroModel (*15*) using the Monte Carlo Multiple Minimum method with the MMFFs forcefield. For each compound, the lowest energy conformer was inspected with Maestro (*16*) and used for the calculation of molecular descriptors.

QSPR Modeling. A total of 647 molecular descriptors for the lowest energy conformers of the 15 volatile compounds detected in the headspace were calculated with DRAGON (*17*) and QikProp (*18*). DRAGON provides several hundreds of generic molecular descriptors within the classes of constitutional descriptors, topological descriptors, walk and path counts, connectivity indices, information indices, 2D autocorrelations, edge adjacency indices, Burden eigenvalues, topological charge indices, eigenvalue-based indices, functional group counts, atom-centered fragments, molecular properties, 2D binary fingerprints, and 2D frequency fingerprints. QikProp provides 45 descriptors, of which several are of pharmaceutical relevance (e.g., predicted skin permeability, QPlogKp), whereas others are of a more general nature [e.g., PM3 calculated ionization potential, IP(eV)]. The combined DRAGON and QikProp descriptor block was imported into MATLAB (*19*), and descriptors remaining constant across 25% or more of the compounds were removed, leaving 441 descriptors. Employing the PLS toolbox (*20*) partial least-squares regression (PLS) models for the prediction of the flavor release as defined in eq 1 were built from the autoscaled descriptor block using forward variable selection. With this method molecular descriptors are introduced one at a time until there is no improvement in the root mean square error of cross-validation (RMSECV) at the optimal number of LVs. Segmented cross-validation was used, implying that a segment of samples is left out and predicted using the remaining compounds. This is repeated until all samples have been predicted once. The choice of cross-validation segments was as follows: segment 1, ethyl acetate, ethyl 2-methylbutanoate, and isopentyl butanoate; segment 2, ethyl propanoate, ethyl isopentanoate, and hexyl acetate; segment 3, ethyl butanoate, ethyl hexanoate, and isopentyl isopentanoate; segment 4, 1-hexanol, *cis*-3-hexenol, and *trans*-2-hexenol. The robustness of each model was assessed by a permutation test in which 1000 response vectors were produced by random permutation of the original response. PLS regression against each permuted response was carried out in the same manner as on the

Table 1. Flavor Compounds Comprising the Model Strawberry Solution Used in This Study[a]

compound	CAS Registry No.	log *P*	VP	MW	solubility
anisyl acetate	000104-21-2	2.16	0.00258	180.21	582
benzyl alcohol	000100-51-6	1.10	0.094	108.14	42900
cis-3-hexenol	000928-96-1	1.61	0.937	100.16	16000
citronellyl acetate	000150-84-5	4.56	0.0526	198.31	5.69
ethyl 2-methylbutanoate	007452-7-1	1.59	0.2000	130.18	1070
ethyl acetate	000141-78-6	0.73	93.2	88.11	80000
ethyl butanoate	000105-54-4	1.85	12.8	116.16	4900
ethyl hexanoate	000123-66-0	2.83	1.56	144.22	629
ethyl isopentanoate	000108-64-5	2.26	8.3	130.19	2000
ethyl propanoate	000105-37-3	1.21	35.9	102.13	19200
furaneol	003658-77-3	0.82	0.000936	128.13	18500
γ-decalactone	000706-14-9	2.72	0.00512	170.25	292
hexanol	000111-27-3	2.03	0.928	102.18	5900
hexyl acetate	000142-92-7	2.83	1.32	144.22	511
isopentyl butanoate	000106-27-4	3.25	0.95	158.24	118
isopentyl isopentanoate	000659-70-1	3.66	0.886	172.27	44.6
methyl cinnamate	000103-26-4	2.62	0.0345	162.19	387
trans-2-hexenol	000928-95-0	1.61	0.911	100.16	1600

[a] Also shown are CAS Registry Numbers, log *P*, vapor pressure (VP) in mmHg, molecular weight (MW), and solubility (mg/L). The data were obtained from the PhysProp database [Syracuse Research Corp. (SRC)].

original response, and the RMSECV for each permuted model was plotted against the correlation coefficient between the original and permuted responses. Models with low RMSECV arising from permuted responses are due to chance, and the existence of such models calls for reconsideration of the data modeling approach.

Principal Component Analysis (PCA). Each flavor release profile can be viewed as a point (object) in a high-dimensional variable space spanned by A_{rel} for each compound. PCA allows for exploration of this variable space by finding the loadings (eigenvectors) and scores (eigenvalues) of the covariance matrix. The loadings are orthogonal vectors of maximum variance in the space spanned by A_{rel}, and the scores are the coordinates of the flavor release profiles in this basis. The scores enable a direct comparison of the similarity of flavor release profiles, whereas the loadings provide means for comparison of the release behavior of individual compounds. PCA was performed using Latentix (*21*) on the matrix of mean-centered flavor release profiles.

RESULTS AND DISCUSSION

Undetected Compounds. Anisyl acetate, benzyl alcohol, and Furaneol were not detected in the headspace above water or β-glucan solutions. It is tempting to attribute the lack of detection of anisyl acetate and Furaneol primarily to the very low vapor pressures of these compounds (see **Table 1**). However, the vapor pressure of benzyl alcohol (0.094 mmHg) is higher than the vapor pressure of citronellyl acetate (0.0526 mmHg), which, despite a low value of A_{rel}, is consistently detected in the headspace. Thus, clearly the vapor pressure is not the only variable determining the presence of volatiles in the headspace. It can be noted that the solubility of benzyl alcohol (42900 mg/L) is much higher than the solubility of citronellyl acetate (5.69 mg/L). The combination of low vapor pressure and high solubility might explain why the former compound was not detected in the headspace but the latter compound was.

Flavor Release Profiles. The flavor release profiles for 5, 10, and 15% preparations of the oat and barley β-glucan products (**Figures 1** and **2**) are plots of A_{rel} for each compound as defined in eq 1. The compounds are sorted so that their molecular weights (MW) increase to the right. In the following we employ the shorthand notations BG-G5, BG-G10, BG-G15, BG-P5, BG-P10, and BG-P15 for the release profiles, where BG stands for β-glucan and G or P immediately followed by one or two digits denotes the concentration of Glucagel and PromOat, respectively. The similarity of the main features of the flavor release profiles from barley and oat preparations indicates that there are no qualitative differences in the flavor release from these two products. In general, the retention of flavor compounds increases with the MW of the flavor compound and with higher concentration of β-glucan product. For oat and barley preparations of the same concentration, it generally holds that the oat preparations retain the flavor compounds more strongly. The only notable exceptions to this are the alcohols in BG-P5, which are released more than in BG-G5. Esters have A_{rel} below unity in all preparations of β-glucan, signifying a preference for the β-glucan matrices relative to water. The alcohols have characteristically high values of A_{rel} when compared to esters. For some preparations of β-glucan A_{rel} even surpasses unity, implying that in these cases the alcohols have a higher affinity for water than for the β-glucan preparation. This behavior is especially pronounced in the case of BG-P5. The profiles for BG-G5 and BG-G10 also show

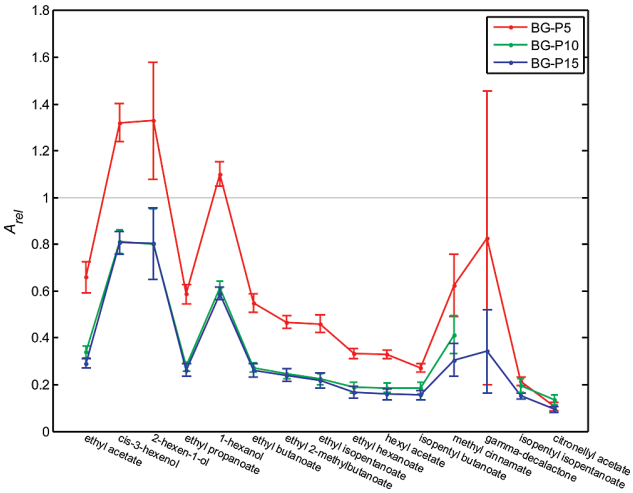


Figure 1. Flavor release profiles for the three oat β-glucan preparations BG-P5, BG-P10, and BG-P15, corresponding to 5, 10, and 15% PromOat, respectively. A_{rel} is plotted with standard error bars for each of the 15 detected flavor compounds.

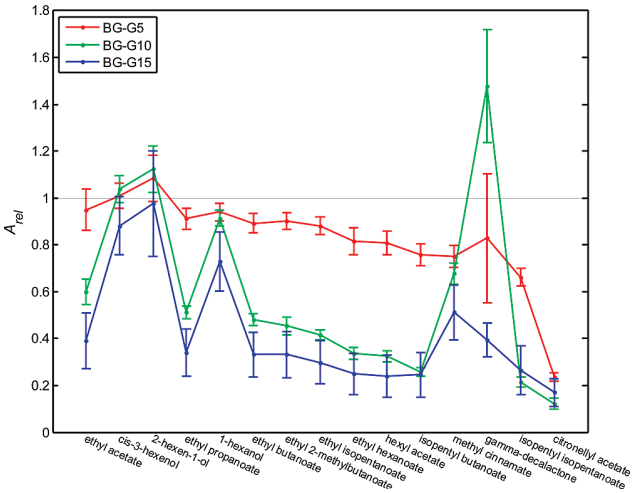


Figure 2. Flavor release profiles for the three barley β-glucan preparations BG-G5, BG-G10, and BG-G15, corresponding to 5, 10, and 15% Glucagel, respectively. A_{rel} is plotted with standard error bars for each of the 15 detected flavor compounds.

indications of the behavior for *cis*-3-hexenol and 2-hexen-1-ol, but in these cases the magnitudes of the standard errors give more ambiguous results. The most conspicuous flavor release profiles in this study are BG-G5, which is characteristically flatter than the remaining profiles, and BG-P5, which, on the other hand, displays more prominent variations than the other profiles. In slightly more quantitative terms, the average absolute deviations from the mean (MAD) A_{rel} are 0.14 and 0.31 for BG-G5 and BG-P5, respectively. For both PromOat and Glucagel the gross features of the profiles corresponding to 10 and 15% of the β-glucan product appear to be quite similar (disregarding γ-decalactone). In the former case the release profiles coincide for several compounds, and only very minor increases in retention are seen when the concentration is increased from 10 to 15%. For Glucagel this increase in concentration seems to be associated with a more significant increase in retention, but the appreciable standard errors of the BG-G15 profile could change this picture. A qualitative overview of the relative positions of all six flavor profiles can be obtained by comparing the mean value of A_{rel} across 13 compounds (excluding γ-decalactone and methyl cinnamate) for each profile. These values are 0.83 (BG-G5), 0.59 (BG-P5), 0.52 (BG-G10), 0.42 (BG-G15), 0.34 (BG-P10), and 0.32 (BG-P15). The variables MW, vapor pressure (VP), log *P*, and solubility reported in **Table 1** do not immediately explain the remarkable difference between alcohol and ester releases. The fact that alcohols can act as both hydrogen bond donors and acceptors, whereas esters can act only as hydrogen bond donors, might play a role in the significantly lower retention of the alcohols. A more detailed insight into the differences between alcohols and esters in molecular descriptor space is offered by PCA. The results of this analysis on a suitable subset of molecular descriptors have been included in Supporting Information Figure S1. This analysis shows that the group of alcohols is separated from the esters primarily on the basis of such properties as propensity for hydrogen bonding and predicted water/gas partition coefficient. The release of methyl cinnamate clearly does not have the same dependence on MW as the aliphatic esters. Scatter plots of all flavor release profiles, including methyl cinnamate, against the complete set of molecular descriptors in this study always revealed methyl cinnamate as an outlier. Most likely, the aromatic ring plays a role in determining the unique flavor release of this compound.

The largest standard error in A_{rel} is observed for γ-decalactone. This compound was inconsistently detected across the flavor release profiles, and in the case of BG-P10 it was not detected in the headspace. In some cases (BG-G5 and BG-P5) the standard error in A_{rel} is too large to allow for any quantitative assessment of the release behavior of γ-decalactone, whereas in other cases (e.g., BG-G10 and BG-G15) the standard error in A_{rel} is of a more acceptable magnitude. Regardless, in these cases the large change in A_{rel} observed from one profile to another (e.g., from BG-G10 to BG-G15) raises concerns about the reliability of the headspace detection of γ-decalactone. γ-Decalactone has the lowest vapor pressure (0.00512 mmHg) of all detected compounds. It is also relatively soluble (292 mg/L) and is unique among the detected compounds in containing a lactone ring. At this point it is unclear if and how these properties are related to the significant uncertainties associated with measuring the flavor release. Citronellyl acetate is remarkable in that A_{rel} is small and influenced very little by the β-glucan product concentration. Citronellyl acetate is the heaviest compound in the study and has the highest predicted value (4.56) for log *P*. This might explain an increased retention in the β-glucan preparations due to favorable hydrophobic interactions. This feature and the comparatively low vapor pressure and intermediate solubility might cooperatively contribute to a low (but consistently detectable) concentration of citronellyl acetate in the headspace. The pronounced variations in the release of the various flavor compounds from BG-P5 suggest that more studies should be made on the flavor release at low concentrations of PromOat. The corresponding 5% preparation of the barley β-glucan product gives a less characteristic (flatter) release profile, as would seem natural from a diluted β-glucan preparation. Finally, variations in the flavor release of some isomers (e.g., ethyl 2-methylbutanoate and ethyl isopentanoate) are indicated in some profiles, but these subtle changes are typically of the same order of magnitude as the standard errors.

PCA. PCA was employed to provide an overview of the relationships between the different flavor release profiles. Scores and loading plots for the first two principal components of this analysis are shown in **Figure 3**. The first principal component (PC1) explains 85% of the variance, whereas the second principal component (PC2) explains 15% of the variance. The first two PCs thus capture all variation in the flavor release profiles, which indicates a simple release system. Lines are drawn to indicate the relationship between scores for flavor release profiles for the same type (oat or barley) of β-glucan. The similar release/retention behaviors for the profiles BG-P10, BG-P15, and BG-G15 are evident from the grouping in upper right-hand pane of the score plot (**Figure 3a**). Conversely, the dissimilar profiles BG-G5 and BG-P5 are separated by different scores on PC1 and particularly on PC2. From the loading plot in **Figure 3b** it is clear that BG-P5 is unique in its release of alcohols. The separation between the

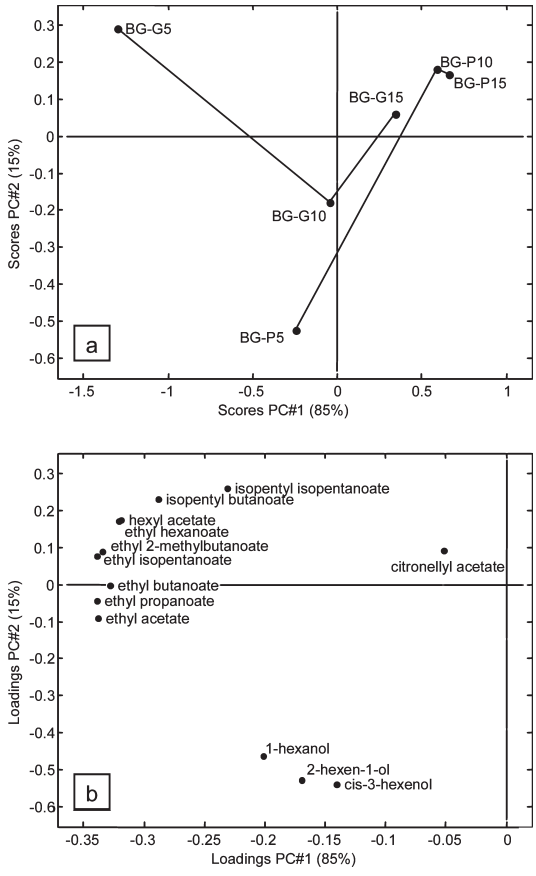


Figure 3. Principal component analysis of the flavor release from six preparations of β-glucan/water matrices: (**a**) score plot showing the relationship between flavor release profiles [lines are drawn between scores for the same type (oat or barley) of β-glucan product]; (**b**) loading plot showing the relationship between flavor release for the individual compounds.

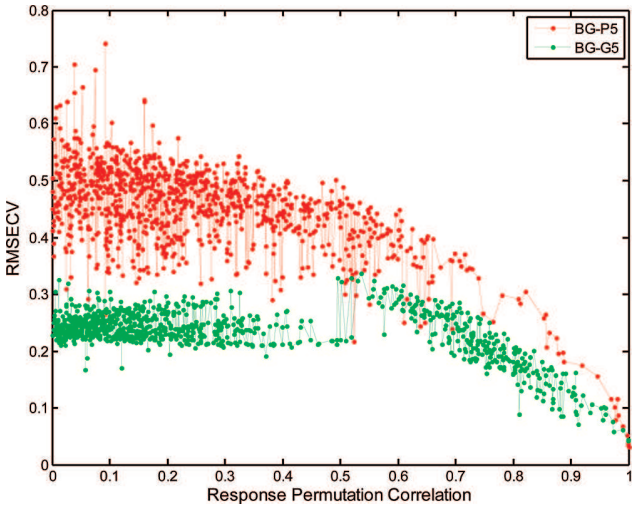


Figure 4. Response permutation test for the QSPR models for the flavor release profiles BG-P5 and BG-G5.

group of alcohols and esters in the loading plot reveals the disparate release behavior of the two classes of compounds. Along PC1, from left to right, the concentration for the β-glucan products increases. Thus, concentration appears to be the most important parameter for variation in the flavor release. The influence of type of β-glucan product is reflected in the different scores on PC1 for preparations of barley and oat β-glucan product with the same concentration.

QSPR Models. Preliminary modeling showed that γ-decalactone and methyl cinnamate could not be included in the global models, and these two compounds were excluded from the rest of the study. Presumably, the problem with including γ-decalactone in the models was principally due to the significant uncertainties in the flavor release of this compound. In the case of methyl cinnamate, the presence of an aromatic ring may have been problematic as this molecular motif is not present in the remaining compounds included in the QSPR models. The choice of cross-validation in the current study deserves attention, because the compounds fall into two classes, esters and alcohols, with markedly different release behaviors as discussed above. Because there are no intermediate response values between the two groups, the risk of overfitting is appreciable. To circumvent this problem, segmented cross-validation was employed by which the alcohols were arranged in their own segment. The remaining segments were chosen so as to achieve significant structural diversity within each segment. Without any a priori knowledge of the mechanisms involved, it is expected in analogy with a similar approach in a previous study on the function of calcitriol analogues (*22*) that the diverse set of molecular descriptors from both QikProp and Dragon contains information relevant to flavor release from the β-glucan matrices. QSPR models for all flavor release profiles were characterized by low RMSECVs ($0.02 < \text{RMSECV} < 0.04$). Between two and four molecular descriptors and a maximum of three latent variables (LVs) were used in the models, which all were well-behaved in subsequent response−permutation tests. This is a strong indication of the robustness of the models. Results of the response−permutation tests are illustrated in **Figure 4** for QSPR models for BG-P5 and BG-G5. Visual inspection of predicted versus measured plots shows that data points are generally in the vicinity of the target line $x = y$. As an example of this, the predicted versus measured plots for the QSPR models for BG-P5 and BG-G5 are shown in panels **a** and **b**, respectively, of **Figure 5**. A summary of the QSPR parameters for all models is given in **Table 2**, and definitions of the selected

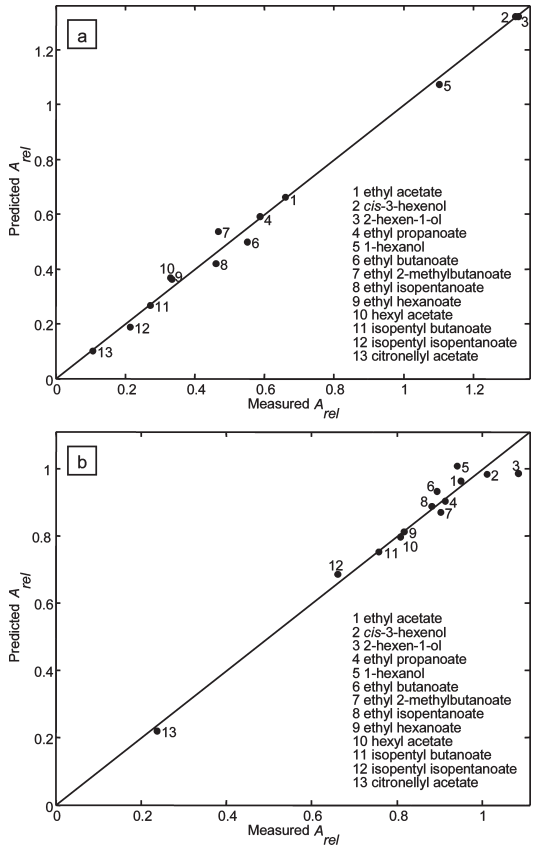


Figure 5. Predicted versus measured plots from QSPR analysis of the flavor release from two β-glucan matrices: (**a**) model for BG-P5; (**b**) model for BG-G5. The target line $x = y$ is shown.

Table 2. Summary of Parameters (RMSECV for the Final Model, Latent Variables, and Descriptors) for QSPR Models of Strawberry Flavor Release from Oat and Barley β-Glucan Matrices[a]

	oat β-glucan		barley β-glucan	
	RMSECV	descriptors	RMSECV	descriptors
5% BG-5	0.03 at 3 LV	ZM1V (0.12) PCR (0.05) EEig01r (0.03)	0.04 at 2 LV	GMTIV (0.05) DECC (0.04)
10% BG-10	0.02 at 3 LV	Hy (0.05) ESpm03d (0.04) X1Av (0.02)	0.02 at 3 LV	ZM1V (0.08) SIC1 (0.05) MWC09 (0.02)
15% BG-15	0.03 at 3 LV	Hy (0.06) EEig01x (0.04) ESpm05u (0.03) X4Av (0.03)	0.03 at 3 LV	Hy (0.06) ESpm15x (0.04) glob (0.04) ESpm05d (0.03)

[a] Descriptors are listed in the order they were selected by forward selection. The number in parentheses following a descriptor is the RMSECV after inclusion of that descriptor.

descriptors are given in **Table 3**. The value of the RMSECV after inclusion of a descriptor is noted in parentheses after that descriptor in **Table 2**. Although the inclusion of a second descriptor in the case of BG-P5 causes the RMSECV to improve from 0.12 to 0.05, in most cases a single descriptor is sufficient to provide a very low RMSECV value. This strengthens the notion of a flavor release system governed by simple mechanisms. Relationships between the flavor release profiles are reflected in the QSPR models. For instance, the hydrophilicity (Hy) descriptor was the first selected descriptor for prediction models for the

Table 3. Summary of Molecular Descriptors Used in the QSPR Models[a]

molecular descriptor	type	description
DECC	topological descriptors	eccentric
EEig01r	edge adjacency indices	eigenvalue 01 from edge adjacency matrix weighted by resonance integrals
EEig01x	edge adjacency indices	eigenvalue 01 from edge adjacency matrix weighted by edge degrees
ESpm03d	edge adjacency indices	spectral moment 03 from edge adjacency matrix weighted by dipole moments
ESpm05d	edge adjacency indices	spectral moment 05 from edge adjacency matrix weighted by dipole moments
ESpm05u	edge adjacency indices	spectral moment 05 from edge adjacency matrix
ESpm15x	edge adjacency indices	spectral moment 15 from edge adjacency matrix weighted by edge degrees
glob	QikProp	globularity descriptor
GMTIV	topological descriptors	Gutman MTI by valence vertex degrees
Hy	molecular properties	hydrophilic factor
MWC09	walk and path counts	molecular walk count of order 09
PCR	walk and path counts (block 3)	ratio of multiple path count over path count
SIC1	information indices	structural information content (neighborhood symmetry of 1 order)
X1Av	connectivity indices	average valence connectivity index chi-1
X4Av	connectivity indices	average valence connectivity index chi-4
ZM1V	topological descriptors	first Zagreb index by valence vertex degrees

[a] The descriptors are from DRAGON 5.5 (*17*) except for glob, which is from QikProp (*18*).

relatively similar profiles BG-P10, BG-P15, and BG-G15. The correlation between the Hy descriptor and these flavor release profiles is between 0.97 and 0.98. Despite these large correlations, however, the release of the alcohols does not correlate well with the Hy descriptor. The regression equations for the two most disparate flavor release profiles in this study, BG-P5 and BG-G5, are given in eqs 3 and 4, respectively.

$$A_{\mathrm{rel,BG-P5}} = -0.0155 \times \mathrm{ZM1V} + 2.5412 \times \mathrm{PCR} + 0.2900 \times \mathrm{EEig01r} - 1.5410 \quad (3)$$

$$A_{\mathrm{rel,BG-G5}} = -0.0004 \times \mathrm{GMTIV} + 0.247 \times \mathrm{DECC} + 0.9057 \quad (4)$$

ZM1V (*23*) is the first selected descriptor for the QSPR model for BG-P5. It is calculated from the adjacency matrix for a molecule and is related to the degree of molecular branching. In **Figure 6a** the behavior of the descriptors in the QSPR model for BG-P5 can be followed across the compounds. The behavior of ZM1V looks like an inverted flavor profile. The first coefficient in eq 3 provides an inversion of ZM1V, thereby establishing the general features of the predicted flavor release profile. Subsequently, the addition of the PCR (*24*) descriptor primarily attenuates the exaggerated magnitude of the peak from ZM1V corresponding to 1-hexanol. Furthermore, the value of ZM1V shows a significant change from isopentyl isopentanoate to citronellyl acetate. Because PCR shows a similar change in the opposite direction, the result from addition is a smoother behavior, which is in better agreement with the observed profile. Finally, more subtle variations in the predicted profile are accounted for by the addition of EEig01r (*17*). A similar analysis can be made for the descriptors for the QSPR model for BG-G5. The coarse outline of the flavor release profile BG-G5 is provided by variation in the GMTIV (*25*) descriptor. Upon addition of the DECC (*26*) descriptor multiplied by a suitable constant, the more distinct features due to the alcohols emerge. Finally, it can be noted that although the descriptors selected in the PLS modeling approach can be complex to interpret, high correlations between the flavor release profiles and more intuitive descriptors such as Hy (hydrophilicity) also exist. Plots of A_{rel} for each β-glucan preparation against the eight most highly correlated descriptors are provided in the Supporting Information.

In conclusion, the flavor release profiles in this study show that the retention of esters and alcohols increases with the molecular weight of the flavor compounds and with the concentration of

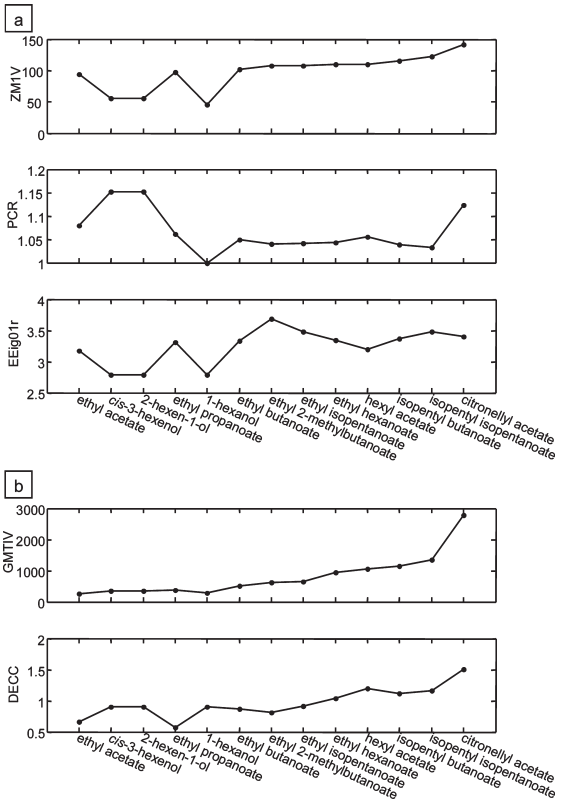


Figure 6. Variation in the molecular descriptors used in two QSPR models for flavor release: (**a**) model for BG-P5; (**b**) model for BG-G5.

β-glucan product. Increasing the β-glucan product concentration from 5 to 10% generally causes a larger increase in flavor retention than the change from 10 to 15%. Esters always have A_{rel} below unity, whereas alcohols in some cases have A_{rel} above unity. The profile BG-P5 in particular exhibits a remarkably high release of alcohols, which suggests future studies on low concentration oat β-glucan matrices. Comparison of oat and barley β-glucan products at the same concentration shows that oat preparations generally retain the flavor compounds more strongly. The different compositions of the oat and barley β-glucan products employed make it difficult to conclude whether or not the stronger retention in oat is related to the β-glucan type (low or high molecular weight). However, the direct observations of flavor release profiles and the trends noted with PCA suggest

that flavor release is not guided by subtle interactions with the β-glucan polymer. This is in accordance with similar observations from a carrageenan study (*11*). Instead, the important role of the MW of flavor compounds and concentration of β-glucan product points to a simple flavor release mechanism. It was demonstrated that the employed QSPR methodology can produce simple and robust models for the prediction of flavor release from the matrices investigated. The actual models produced are probably of limited practical use due to the strongly restricted chemical space they were constructed from. However, the successful application of the QSPR approach, including variable selection and appropriate choice of cross-validation scheme, shows that the investigated systems are amenable to such procedures. The fact that single molecular descriptors, such as the hydrophilicity descriptor or molecular weight, cannot simultaneously account for the release of esters and alcohols in this study emphasizes the importance of a multivariate approach, which may establish the connection between the release phenomenon and several less obvious theoretical molecular descriptors.

ABBREVIATIONS USED

QSPR, quantitative structure property relationship; PLS, partial least-squares regression; GC-MS, gas chromatography−mass spectrometry; LV, latent variable; RMSECV, root mean square error of cross-validation; PCA, principal component analysis; MAD, mean absolute deviation.

ACKNOWLEDGMENT

We warmly thank Mehdi Farahani for assistance during the laboratory work.

Supporting Information Available: Tables S1 and S2 contain the numerical values of A_{rel} and associated standard errors. A biplot for the PCA on a molecular descriptor space showing separation of esters and alcohols based on physicochemical properties is shown in Figure S1. Table S3 contains descriptions of the molecular descriptors used in Figure S1. Figures S2−S7 show scatter plots of A_{rel} for each preparation of β-glucan product against the eight molecular descriptors most highly correlated with the property. This material is available free of charge via the Internet at http://pubs.acs.org.

LITERATURE CITED

(1) Wood, P. Cereal β-glucans: structure, properties and health claims. In *Advanced Dietary Fibre Technology*; McCleary, B., Prosky, L., Eds.; Blackwell Science: Oxford, U.K., 2001; Vol. *1*, pp 315−327.
(2) Burkus, Z.; Temelli, F. Rheological properties of barley β-glucan. *Carbohydr. Polym.* **2005**, *59*, 459–465.
(3) Lazaridou, A.; Biliaderis, C. G.; Micha-Screttas; Steele, B. R. A. Comparative study on structure−function relations of mixed-linkage (1→3), (1→4) linear β-D-glucans. *Food Hydrocolloids* **2004**, *18*, 837–855.
(4) Vaikousi, H.; Biliaderis, C. G.; Izydorczyk, M. S. Solution flow behaviour and gelling properties of water-soluble barley (1→3, 1→4)-β-glucans varying on molecular size. *J. Cereal Sci.* **2004**, *39*, 119–137.
(5) Aman, P.; Rimsten, L.; Andersson, R. Molecular weight distribution of β-glucan in oat-based foods. *Cereal Chem.* **2003**, *81*, 356–360.
(6) Boland, A. B.; Buhr, K.; Giannouli, P.; van Ruth, S. M. Influence of gelatin, starch, pectin and artificial saliva on the release of 11 flavour compounds from model gel systems. *Food Chem.* **2004**, *86*, 401–11.
(7) Malkki, Y.; Heinio, R. L.; Autio, K. Influence of oat gum, guar gum and carboxymethyl cellulose on the perception of sweetness and flavor. *Food Hydrocolloids* **1993**, *6*, 525–532.
(8) Pangborn, R. M.; Gibbs, Z. M.; Tassan, C. Effect of hydrocolloids on apparent viscosity and sensory properties of selected beverages. *J. Texture Stud.* **1978**, *9*, 415–436.
(9) de Roos, K. B. Effect of texture and microstructure on flavour retention and release (review). *Int. Dairy J.* **2003**, *13*, 593–605.
(10) Friel, E. N.; Linforth, R. S. T.; Taylor, A. J. An empirical model to predict the headspace concentration of volatile compounds above solutions containing sucrose. *Food Chem.* **2000**, *71*, 309–317.
(11) Chana, A.; Tromelin, A.; Andriot, I.; Guichard, E. Flavor release from ι-carrageenan matrix: a quantitative structure−property relationships approach. *J. Agric. Food Chem.* **2006**, *54*, 3679–3685.
(12) Lubbers, S.; Decourcelle, N.; Martinez, D.; Guichard, E.; Tromelin, A. Effect of thickeners on aroma compound behavior in a model dairy gel. *J. Agric. Food Chem.* **2007**, *55*, 4835–4841.
(13) Juteau, A.; Tournier, C.; Guichard, E. Influence of type and amount of gelling agent on flavour perception: physicochemical effect or interaction between senses?. *Flavour Fragrance J.* **2004**, *19*, 483–490.
(14) Mark A. Thompson, ArgusLab 4.0.1, Planaria Software LLC, Seattle, WA; http://www.arguslab.com.
(15) Macromodel, vers. 9.6, Schrödinger, LLC, New York, 2008.
(16) Maestro, Schrödinger, LLC, New York, 2008.
(17) DRAGON, vers. 5.5, Talete SRL, Milan, Italy.
(18) QikProp, vers. 2.5, Schrödinger, LLC, New York.
(19) Matlab, vers. 5.3, The MathWorks Inc., Natick, MA.
(20) PLS toolbox, Eigenvector Research, Inc.
(21) LatentiX, vers. 2.0, Latent5, Copenhagen, Denmark.
(22) Jensen, B. F.; Sørensen, M. D.; Kissmeyer, A.-M.; Björkling, F.; Sonne, K.; Engelsen, S. B.; Nørgaard, L. Prediction of in vitro metabolic stability of calcitriol analogs by QSAR. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 849–859.
(23) Gutman, I.; Ruscic, B.; Trinajstic, N.; Wilcox, C. F. Graph theory and molecular orbitals, part XII. Acyclic polyenes. *J. Chem. Phys.* **1975**, *62*, 3399–3405.
(24) Rücker, G.; Rücker, C. Counts of all walks as atomic and molecular descriptors. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 683–695.
(25) Gutman, I. Selected properties of the Schultz molecular topological index. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1087–1089.
(26) Konstantinova, E. V. The discrimination ability of some topological and information distance indices for graphs of unbranched hexagonal systems. *J. Chem. Inf. Comp. Sci.* **1997**, *36*, 54–57.

Received for Review December 8, 2008. Revised manuscript received March 29, 2009. Accepted April 04, 2009. The Ministry of Food Agriculture and Fisheries is gratefully acknowledged for financial support to the project "FFS05-9: Build Your Food".

Supporting Information for:

A Quantitative Structure Property Relationship Study of the Release of Some Esters and Alcohols from Barley and Oat β-Glucan Matrices

Niels Johan Christensen*,[†] Susana Murtinheira da Trindade Leitão,[†] Mikael Agerlin Petersen,[†] Birthe Møller Jespersen, and Søren Balling Engelsen.[†]

[†] Quality & Technology, Department of Food Science, University of Copenhagen, Rolighedsvej 30, 1958 Frederiksberg C, Denmark

* Corresponding author: Niels Johan Christensen, Quality and Technology, Department of Food Science, University of Copenhagen, Rolighedsvej 30, 1958 Frederiksberg C, Denmark. Phone: 353 33506, fax: 353 33245, e-mail: njc@life.ku.dk.

Table S1: Flavor release (A_{rel}) from preparations of the barley β-glucan product Glucagel. Standard errors are calculated from triplicate samples, as described in the materials and methods section in the manuscript.

Compound	BG-G5 (Glucagel 5%)		BG-G10 (Glucagel 10%)		BG-G15 (Glucagel 15%)	
	A_{rel}	Std. Error	A_{rel}	Std. Error	A_{rel}	Std. Error
ethyl acetate	0.950	0.087	0.599	0.053	0.391	0.118
cis-3-hexenol	1.010	0.055	1.038	0.058	0.881	0.125
2-hexen-1-ol	1.084	0.100	1.125	0.099	0.976	0.226
ethyl propanoate	0.912	0.045	0.512	0.028	0.340	0.100
1-hexanol	0.940	0.038	0.914	0.033	0.729	0.125
ethyl butanoate	0.893	0.040	0.481	0.025	0.332	0.096
ethyl 2-methylbutanoate	0.901	0.037	0.454	0.038	0.331	0.100
ethyl isopentanoate	0.882	0.038	0.416	0.023	0.297	0.092
ethyl hexanoate	0.817	0.057	0.336	0.025	0.249	0.088
hexyl acetate	0.807	0.051	0.324	0.022	0.239	0.090
isopentyl butanoate	0.757	0.047	0.258	0.019	0.245	0.097
methyl cinnamate	0.752	0.047	0.677	0.044	0.512	0.117
gamma-decalactone	0.829	0.275	1.477	0.243	0.395	0.073
isopentyl isopentanoate	0.661	0.038	0.213	0.021	0.264	0.104
citronellyl acetate	0.236	0.019	0.121	0.023	0.170	0.059

Table S2: Flavor release (A_{rel}) from preparations of the oat β-glucan product PromOat. Standard errors are calculated from triplicate samples, as described in the materials and methods section in the manuscript.

Compound	BG-P5 (PromOat 5%)		BG-P10 (PromOat 10%)		BG-P15 (PromOat 15%)	
	A_{rel}	Std. Error	A_{rel}	Std. Error	A_{rel}	Std. Error
ethyl acetate	0.659	0.065	0.339	0.025	0.391	0.118
cis-3-hexenol	1.321	0.082	0.811	0.050	0.881	0.125
2-hexen-1-ol	1.329	0.250	0.801	0.152	0.976	0.226
ethyl propanoate	0.588	0.042	0.277	0.019	0.340	0.100
1-hexanol	1.102	0.054	0.615	0.028	0.729	0.125
ethyl butanoate	0.549	0.041	0.272	0.020	0.332	0.096
ethyl 2-methylbutanoate	0.467	0.027	0.246	0.020	0.331	0.100
ethyl isopentanoate	0.460	0.038	0.223	0.023	0.297	0.092
ethyl hexanoate	0.334	0.021	0.188	0.022	0.249	0.088
hexyl acetate	0.329	0.018	0.187	0.021	0.239	0.090
isopentyl butanoate	0.272	0.017	0.186	0.026	0.245	0.097
methyl cinnamate	0.624	0.134	0.413	0.082	0.512	0.117
gamma-decalactone	0.827	0.627	N/A	N/A	0.395	0.073
isopentyl isopentanoate	0.213	0.019	0.196	0.030	0.264	0.104
citronellyl acetate	0.106	0.016	0.136	0.022	0.170	0.059

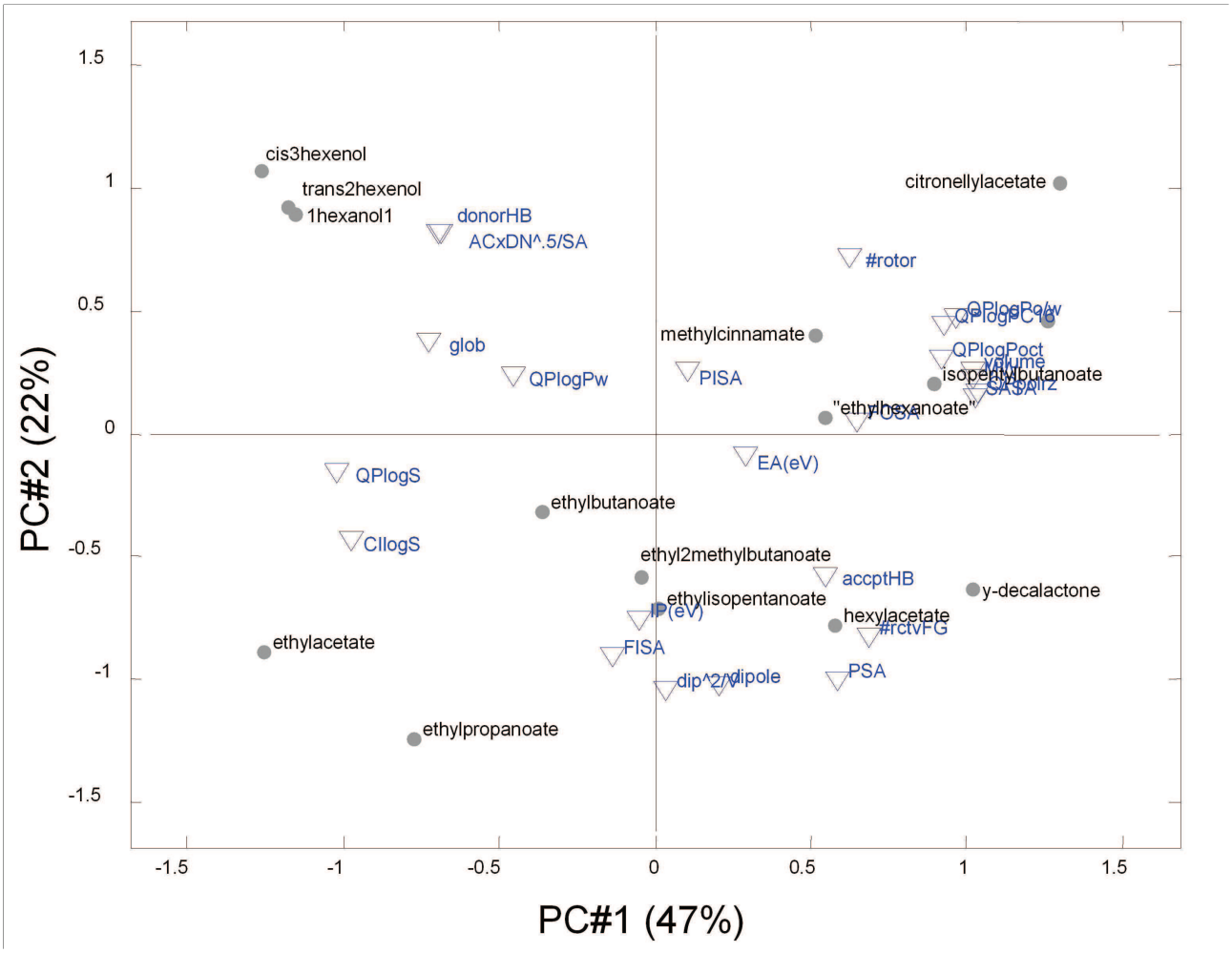


Figure S1: Bi-plot (combined loading and score-plot) showing the first two principal components for the PCA of 15 volatile compounds in a space spanned by 24 molecular descriptors. Molecular descriptors are marked with triangles and compounds are marked with dots. The three alcohols are grouped in the upper left-hand corner of the plot. Molecular descriptors particularly relevant for the grouping of alcohols are donorHB, ACxDN^.5/SA, glob, and QPlogPw.

Table S3: Descriptor of molecular descriptors (from QikProp) used in the PCA (Figure S1).

Molecular Descriptor	Description
#rotor	Number of non-trivial (not CX3), non-hindered (not alkene, amide, small ring) rotatable bonds.
#rctvFG	Number of reactive functional groups.
MW	Molecular weight of the molecule.
Dipole	Computed dipole moment of the molecule.
SASA	Total solvent accessible surface area (SASA) in square angstroms using a probe with a 1.4 Å radius.
FOSA	Hydrophobic component of the SASA (saturated carbon and attached hydrogen).
FISA	Hydrophilic component of the SASA (SASA on N, O, and H on heteroatoms).
PISA	π (carbon and attached hydrogen) component of the SASA.
Volume	Total solvent-accessible volume in cubic angstroms using a probe with a 1.4 Å radius.
DonorHB	Estimated number of hydrogen bonds that would be donated by the solute to water molecules in an aqueous solution. Values are averages taken over a number of configurations, so they can be non-integer.
accptHB	Estimated number of hydrogen bonds that would be accepted by the solute from water molecules in an aqueous solution. Values are averages taken over a number of configurations, so they can be non-integer.
Dip^2/V	Square of the dipole moment divided by the molecular volume. This is the key term in the Kirkwood-Onsager equation for the free energy of solvation of a dipole with volume V.
ACxDN^.5/SA	Index of cohesive interaction in solids. See *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1155.
Glob	Globularity descriptor. Globularity is 1.0 for a spherical molecule.
QPpolrz	Predicted polarizability in cubic angstroms.
QPlogPC16	Predicted hexadecane/gas partition coefficient.
QPlogPoct	Predicted octanol/gas partition coefficient.
QPlogPw	Predicted water/gas partition coefficient.
QPlogPo/w	Predicted octanol/water partition coefficient.
QPlogS	Predicted aqueous solubility, log S. S in moles/liter is the concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid.
CIlogS	Conformation-independent predicted aqueous solubility, log S. S in moles/liter is the concentration of the solute in a saturated solution that is in equilibrium with the crystalline solid.
IP (eV)	PM3 calculated ionization potential.
EA (eV)	PM3 calculated electron affinity.
PSA	Van der Waals surface area of polar nitrogen and oxygen atoms.

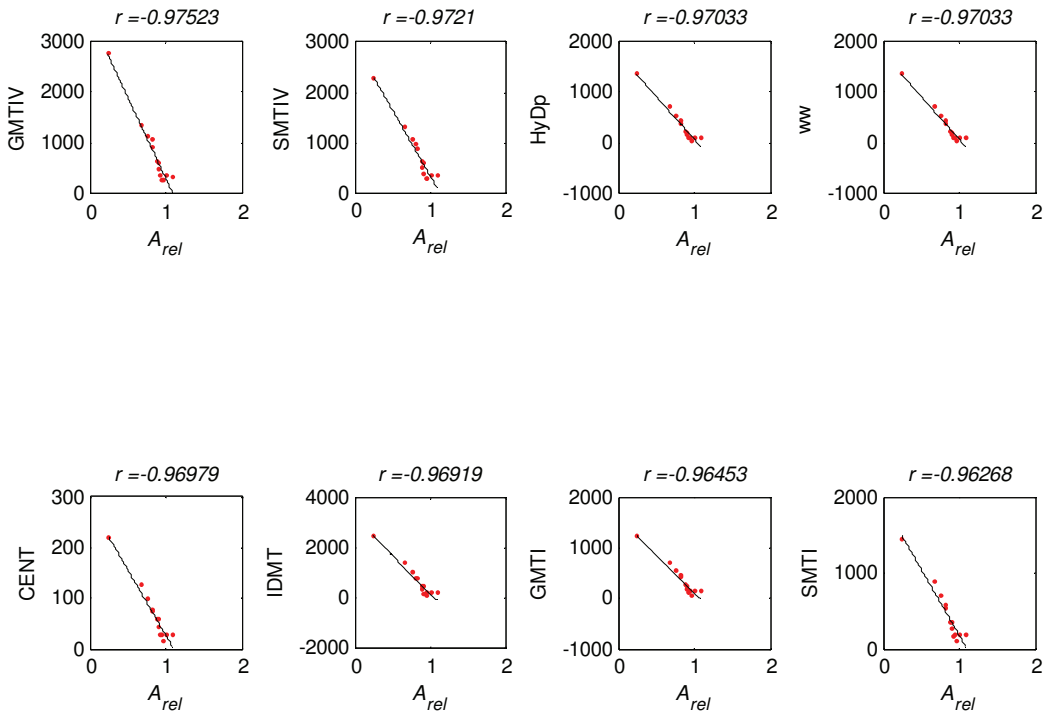


Figure S2: Scatterplots of A_{rel} for BG-G5 against the eight most highly correlated descriptors. In each case, the correlation coefficient (r) is given and a least squares fit to a line is shown.

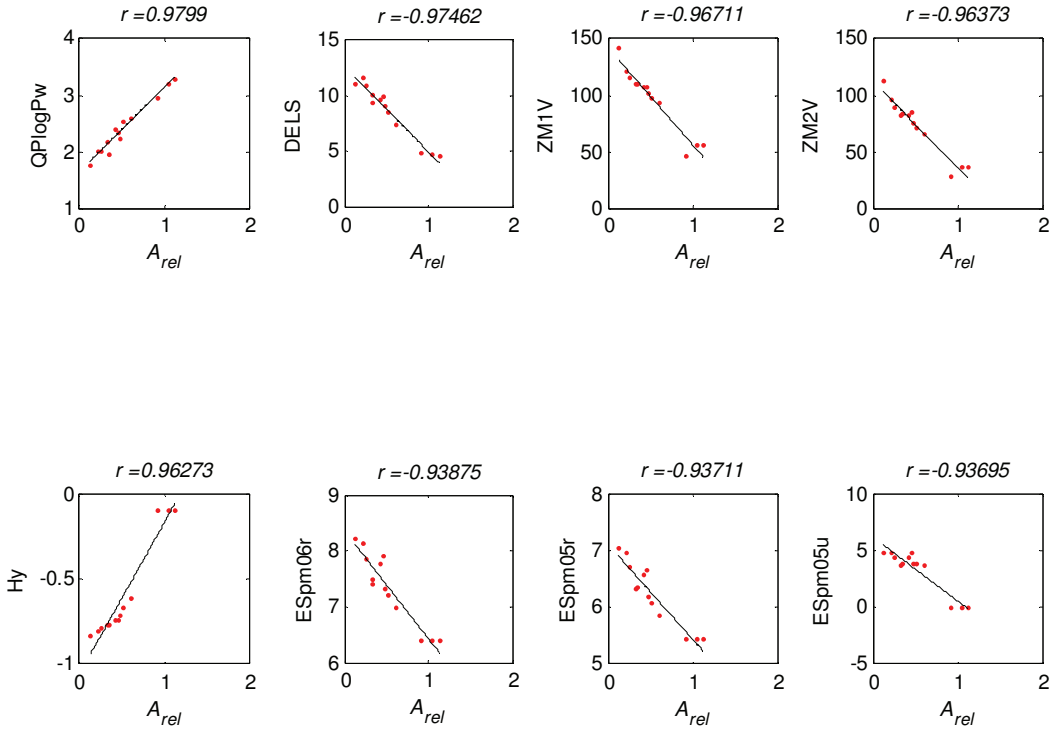


Figure S3: Scatterplots of A_{rel} for BG-G10 against the eight most highly correlated descriptors. In each case, the correlation coefficient (r) is given and a least squares fit to a line is shown.

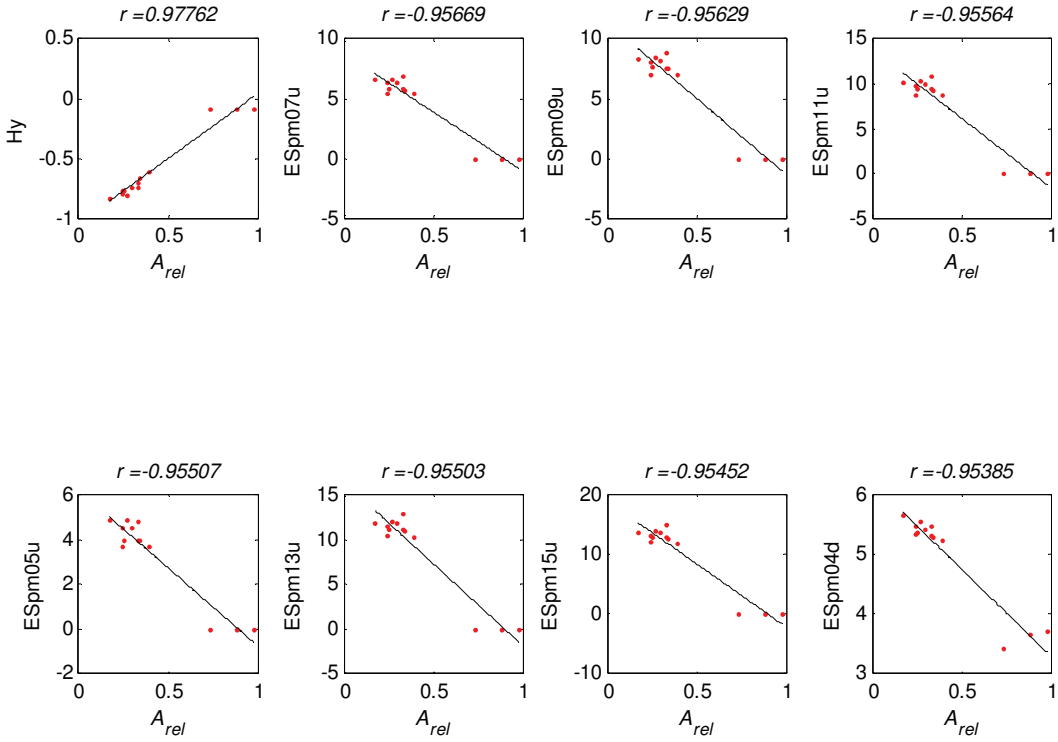


Figure S4: Scatterplots of A_{rel} for BG-G15 against the eight most highly correlated descriptors. In each case, the correlation coefficient (r) is given and a least squares fit to a line is shown.

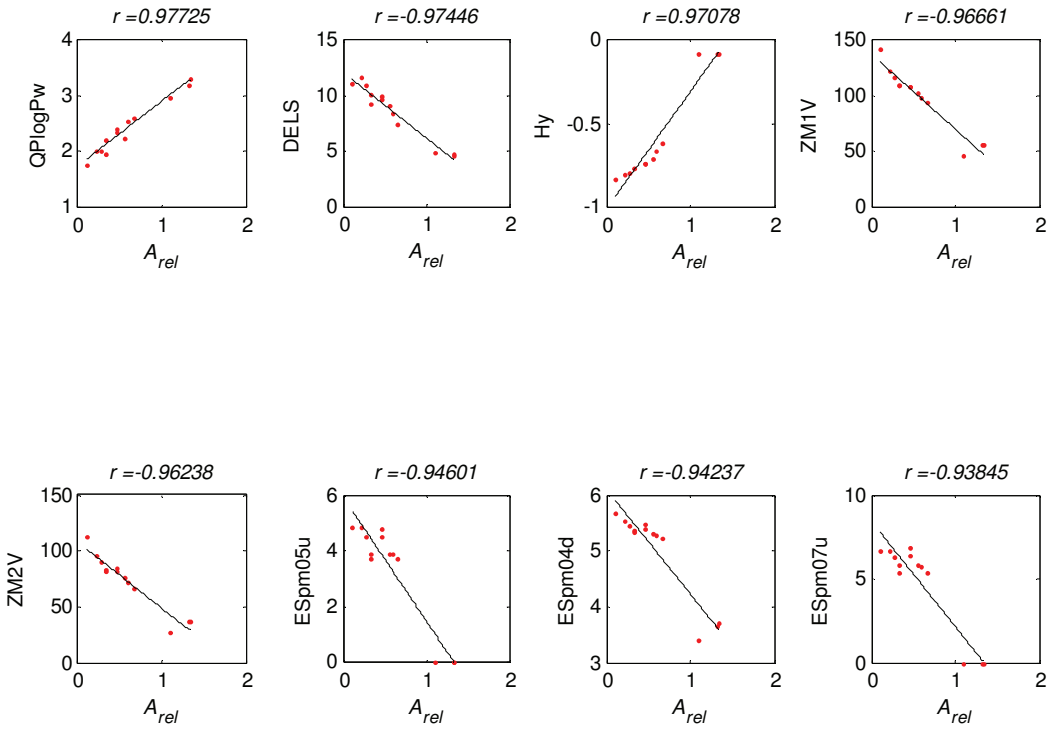


Figure S5: Scatterplots of A_{rel} for BG-P5 against the eight most highly correlated descriptors. In each case, the correlation coefficient (r) is given and a least squares fit to a line is shown.

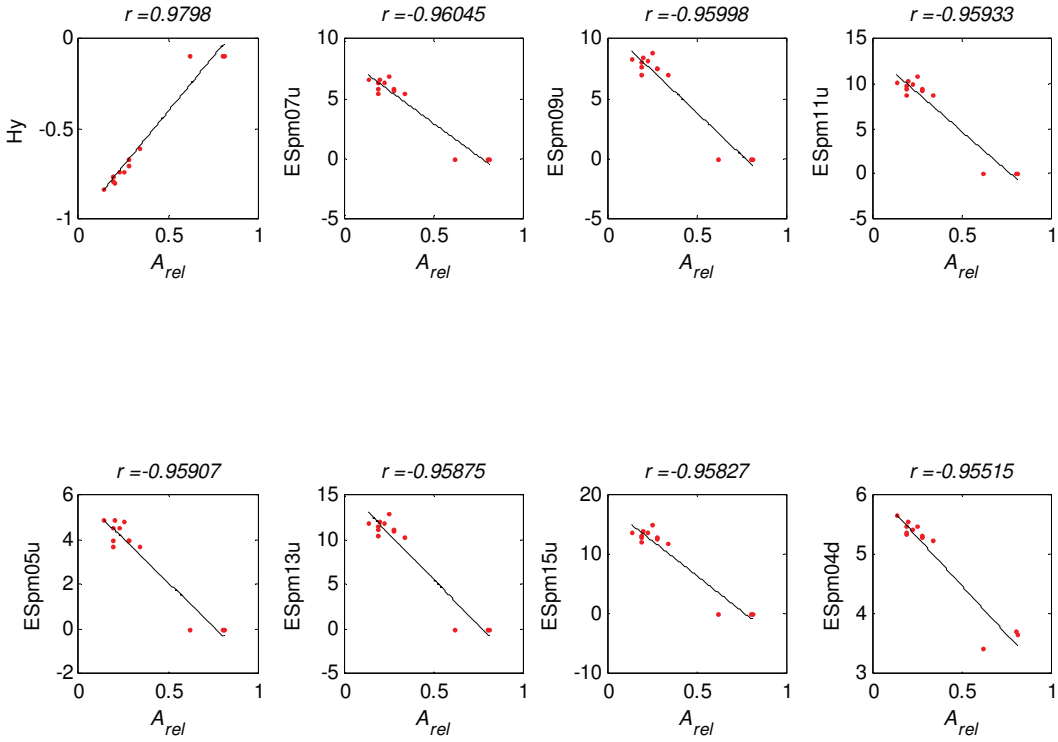


Figure S6: Scatterplots of A_{rel} for BG-P10 against the eight most highly correlated descriptors. In each case, the correlation coefficient (r) is given and a least squares fit to a line is shown.

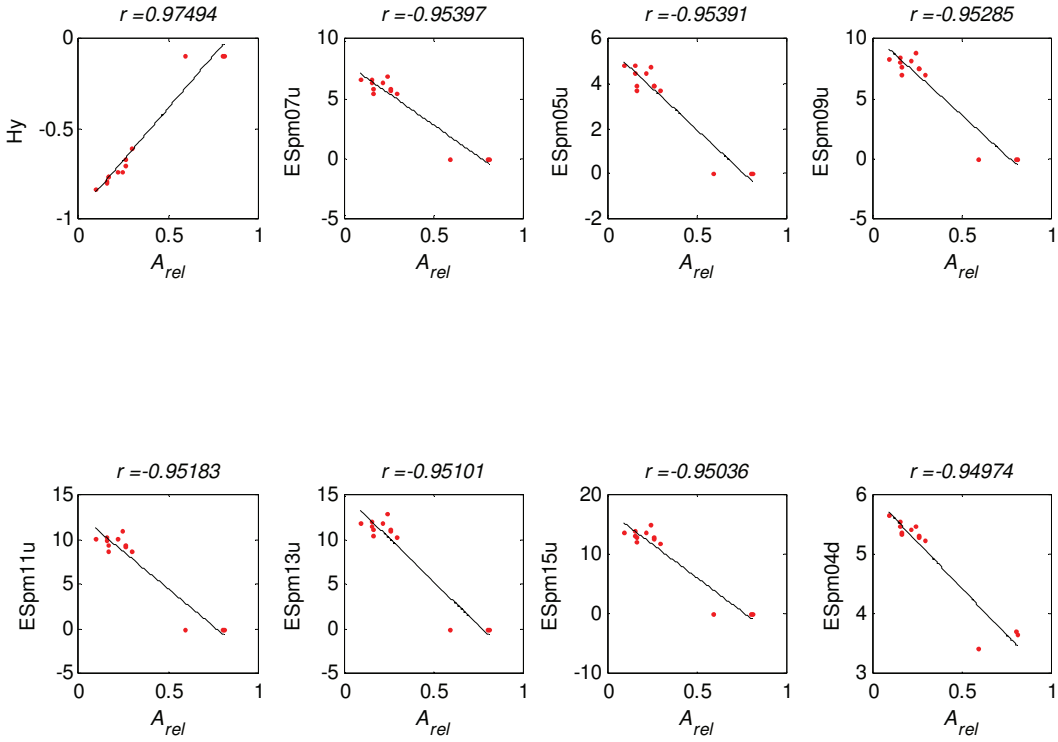


Figure S7: Scatterplots of A_{rel} for BG-P15 against the eight most highly correlated descriptors. In each case, the correlation coefficient (r) is given and a least squares fit to a line is shown.

Paper III

Henrik Toft Simonsen, Mette Skau Nielsen, Niels Johan Christensen, Ulla Christensen, Thomas V. La Cour, Mohammed Saddik Motawia, Birthe Pontoppidan Jespersen, Søren Balling Engelsen, and Birger Lindberg Møller. *Molecular interactions between barley and oat β-glucans and phenolic derivatives*. Journal of Agricultural and Food Chemistry, 2009, 11, 57(5), 2056-2064, **Published.**

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Molecular Interactions between Barley and Oat β-Glucans and Phenolic Derivatives

Henrik Toft Simonsen,[†] Mette S. Nielsen,[†,‡] Niels J. Christensen,[‡] Ulla Christensen,[†] Thomas V. La Cour,[†] Mohammed Saddik Motawia,[†] Birthe P. M. Jespersen,[‡] Søren B. Engelsen,[‡] and Birger Lindberg Møller*,[†]

Plant Biochemistry Laboratory and Laboratory for Molecular Plant Biology, VKR Research Centre Pro-Active Plants, Department of Plant Biology and Biotechnology, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg, Denmark, and Quality and Technology, Department of Food Science, University of Copenhagen, Rolighedsvej 30, 1958 Frederiksberg C, Denmark

Equilibrium dialysis, molecular modeling, and multivariate data analysis were used to investigate the nature of the molecular interactions between 21 vanillin-inspired phenolic derivatives, 4 bile salts, and 2 commercially available β-glucan preparations, Glucagel and PromOat, from barley and oats. The two β-glucan products showed very similar binding properties. It was demonstrated that the two β-glucan products are able to absorb most phenolic derivatives at a level corresponding to the absorption of bile salts. Glucosides of the phenolic compounds showed poor or no absorption. The four phenolic derivatives that showed strongest retention in the dialysis assay shared the presence of a hydroxyl group in *para*-position to a CHO group. However, other compounds with the same structural feature but possessing a different set of additional functional groups showed less retention. Principal component analysis (PCA) and partial least-squares regression (PLS) calculations using a multitude of diverse descriptors related to electronic, geometrical, constitutional, hybrid, and topological features of the phenolic compounds showed a marked distinction between aglycon, glucosides, and bile salt retention. These analyses did not offer additional information with respect to the mode of interaction of the individual phenolics with the β-glucans. When the barley β-glucan was subjected to enzyme degradation, the ability to bind some but not all of the phenolic derivatives was lost. It is concluded that the binding must be dependent on multiple characteristics that are not captured by a single molecular descriptor.

KEYWORDS: β-Glucan; barley; oat; bile salts; phenolic derivatives; β-glucosides

INTRODUCTION

The first publication on the relationship between dietary fibers and small molecules, namely, bile salts, was published by Cooksoon et al. in 1967 (*1*). Since then, the health-promoting effect of dietary fibers and the influence of dietary fibers in food mixtures has been investigated in a large number of studies. The health-promoting effects of dietary fibers are now well documented (*2*−*5*). Knowledge obtained in these studies has inspired this study and has served as a platform in the investigation of the interaction of a different set of small molecules with β-glucans. It has been observed in animal as well as human models that an increase in soluble, viscous nonstarch polysaccharides (e.g., β-glucan) in the diet is accompanied by an increase in fecal sterols, suggesting that these fibers interact with bile salts and cholesterol in the gastrointestinal (GI) tract (*6, 7*). Cellulose was shown not to possess any of these effects, but it was not possible to define the properties responsible for binding or retention (*8*). The adsorption capacities of different fiber types were shown to vary, and the drug colestyramine, a bile acid sequestrant, has been adopted as a standard for these measurements (*8*). Eastwood et al. (*9*) suggested a simple method to establish strong and reversible adsorption. A linear relationship between the percentage of bile acid adsorbed and the amount of fiber used regardless of the bile acid concentration was observed. However, differences between the adsorption of different bile acids made final conclusions difficult. Dietary fibers from different sources have been tested and shown to adsorb bile salts, but the adsorption was not correlated with the ability of the fiber to alter the cholesterol level in vivo. This suggests that several factors influence the properties that account for the adsorption of the bile acids and the lowering of blood cholesterol levels (*10, 11*).

* Corresponding author (telephone 353 33352; fax 353 33333; e-mail blm@life.ku.dk).
[†] Department of Plant Biology and Biotechnology.
[‡] Department of Food Science.

10.1021/jf802057v CCC: $40.75
Published on Web 02/11/2009

Table 1. Percentage Dialysate Retention Based on the Asymptotic UV Absorbance (ΔA) and Dialysis Rate Constants (k_d) from Mixed Solutions of 2.5% (w/v) Glucagel and the 21 Selected Phenolic Derivatives (Means of Minimum Two Replicates)

	dialysate retention[a] (ΔA, %)			dialysis rate constant[b] (k_d)		
compd (name/no.)	day 1	day 2	Δ (2 − 1)	day 1	day 2	Δ (2 − 1)
4′-hydroxy-3′-methoxyacetophenone (**1**)	23	34	11	0.008	0.009	0.001
2-hydroxy-4-methoxybenzaldehyde (**2**)	10	8	−2	0.008	0.007	−0.001
2,6-di-*tert*-butyl-4-methoxyphenol (**3**)	6	3	−3	0.013	0.011	−0.002
2,6-di-*tert*-butyl-4-methylphenol (**4**)	12	18	6	0.012	0.011	−0.001
4-hydroxy-3-methoxybenzyl alcohol (**5**)	12	18	6	0.016	0.012	−0.004
3-hydroxy-4-methoxybenzyl alcohol (**6**)	13	17	4	0.015	0.011	−0.004
3-ethoxy-4-hydroxybenzaldehyde (**7**)	14	25	11	0.008	0.010	0.002
3,5-dimethoxyphenol (**8**)	14	21	7	0.011	0.012	0.001
2,3-dimethylphenol (**9**)	5	−2	−7	0.011	0.009	−0.002
2,5-dimethylphenol (**10**)	13	18	5	0.008	0.010	0.002
2′-hydroxy-4′-methoxyacetophenone (**11**)	19	13	−6	0.009	0.008	−0.001
ethyl 4-ethoxy-2-hydroxybenzoate (**12**)	5	10	5	0.013	0.011	−0.002
3,5-dimethoxy-4-hydroxybenzaldehyde (**13**)	35	38	3	0.009	0.010	0.001
3,5-dimethoxy-4-hydroxybenzoic acid (**14**)	40	34	−6	0.010	0.006	−0.004
2,3,5-trimethylphenol (**15**)	2	−4	−6	0.011	0.008	−0.003
2,3,6-trimethylphenol (**16**)	6	5	−1	0.012	0.010	−0.002
4-hydroxy-3-methoxybenzaldehyde (**17**)	12	4	−8	0.011	0.009	−0.002
4-hydroxy-3-methoxybenzoic acid (**18**)	2	15	13	0.010	0.011	0.001
methyl 4-hydroxy-3-methoxybenzoate (**19**)	5	0	−5	0.009	0.008	−0.001
ethyl 4-hydroxy-3-methoxybenzoate (**20**)	34	32	−2	0.010	0.008	−0.002
4-hydroxybenzyl alcohol (**21**)	17	20	3	0.012	0.009	−0.003

[a] Compounds **1**, **2**, **7**, **11**, **13**, **14**, **17**−**27** were measured at 280 nm and compounds **3**−**6**, **8**−**10**, **12**, **15**, and **16** were measured at 220 nm. [b] Values represent k_d values from mixed solutions of vanillin derivatives and Glucagel.

Fiber-induced changes in fecal bile salt concentrations or composition may not be the sole mechanism involved in the lowering of serum cholesterol. No direct correlation between the viscosity of the matrix and the adsorption has been observed, whereas adjustments of pH and salt strength have been observed to alter the adsorption properties of the fibers (*12*). Multiple adsorption mechanisms are possible, mediated by the same, partly overlapping, or different molecular parameters contributing to viscosity. One major mechanism could be the formation of micelles by the bile acid and the adsorption of these within the fiber (*13, 14*).

β-Glucans are known as hydrocolloid-forming glucose polymers and are used as texture-enhancing additives in the food industry. Several studies have shown that hydrocolloids influence the rate and intensity of flavor release in foods (*15*−*17*). It is recognized that viscosity affects overall flavor perception (*15, 18*). Thickened solutions of similar viscosity do not necessarily offer the same flavor perception. This demonstrates that viscosity as well as adsorption affects flavor release and perception (*19*). The molecular mechanisms that govern the functionality of β-glucans in human health and in food matrices thus remain elusive. Knowledge of the physicochemical interactions that occur between aroma compounds and food constituents is required to be able to describe the behavior of aroma and flavor compounds in food products.

To study interactions between β-glucans and small molecules such as aroma compounds, phenolic derivatives, and bile salts, several different methods have been used, including static headspace, NMR, dynamic exponential dilution, and size exclusion chromatography (*9, 19*−*22*). Thus, thermodynamic and other dynamic approaches have been used to study the behavior of aroma compounds in model complex media that possess different microstructure.

In equilibrium dialysis, a liquid is partitioned through a semipermeable membrane that separates a cell into two compartments, a sample and an assay chamber, of which one contains the dietary fiber. If the interactions that occur between two components (e.g., small molecules such as aroma compounds and macromolecules such as dietary fiber) are strong enough, only the nonretained small molecules will participate in the equilibrium. If the total concentration of a small compound in the two compartments at equilibrium is known, it is possible to calculate the quantity that is adsorbed or retained by the macromolecule matrix (*23*). This enables quantitative assessment of molecular interactions between small compounds and macromolecular food constituents (e.g., β-glucan). Molecular affinities and mechanisms by which β-glucans function may be elucidated by combined studies of small compound retention in β-glucan matrices, molecular modeling, and multivariate data analysis. This knowledge can be correlated to the interactions of other small molecules and β-glucans and provide ideas on how β-glucans affect the aroma of foods and function as health promoters in the intestine.

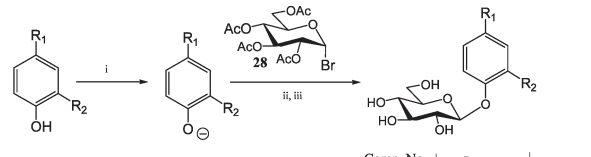
The objective of the present study was to determine the possible interaction between 21 different vanillin-inspired phenolic derivatives, 6 glucosides of these derivatives, and 4 bile salts and specific barley β-glucan (Glucagel) and oat β-glucan (PromOat) preparations using equilibrium dialysis. Differences in the β-glucan-induced retention of the small molecules were related to the specific physicochemical properties of these molecules using molecular modeling and multivariate data analysis.

MATERIALS AND METHODS

Phenolic Derivatives and Bile Acids. Vanillin and 20 different related phenolic structures (**1**−**21**), tryptophan, and the 4 bile acids deoxycholic acid, glycocholic acid, taurocholic acid, and cholic acid were obtained from Sigma-Aldrich (Copenhagen, Denmark) and selected for screening.

Glucosylation of Phenolic Compounds. The investigated phenolic glucosides **22**−**27** (**Table 2**) used in this study were chemically synthesized as shown in **Figure 1**.

Glucosylation of the aglycons **1**, **7**, **17**, and **19**−**21** with 2,3,4,6-tetra-*O*-acetyl-α-D-glucopyranosyl bromide (**28**) was performed in aqueous organic basic media using homogeneous reaction conditions and aqueous NaOH with acetone as the organic cosolvent (*24*) to provide the related aryl O-protected-glucosides, which by Zémplen deacetylation afforded the phenolic compounds **22**−**27**. The purity and structural conformation was verified by NMR spectroscopy.



Comp. No.	R_1	R_2
22	CHO	OCH_3
23	$COOC_2H_3$	OCH_3
24	COCH3	OCH_3
25	$COOCH_3$	OCH_3
26	CH2OH	H
27	CHO	OCH3

Figure 1. Chemical synthesis of the phenolic glucosides **22**−**27**: (i) NaOH, H_2O, <10 °C, 15 min; (ii) acetone, room temperature, 24 h; (iii) MeOH, MeONa/MeOH, room temperature, 1−2 h.

Barley β-Glucan. Glucagel, a commercial soluble β-glucan extracted from hull-less barley, was obtained from GraceLinc Ltd. (Christchurch, New Zealand). Glucagel has a declared content of ≥75% β-glucan, <18% starch, <10% moisture, <5% protein, <2% ash, and <2% fat. The β-glucan is of moderate molecular weight, $(0.12-0.18) \times 10^6$.

Oat β-Glucan. PromOat, a commercial soluble β-glucan extracted from oat, was obtained from Biovelop (Kimstad, Sweden). PromOat has a declared content of 30−40% β-glucan, 6% pentosans, 49% carbohydrates (described as dextrins by the supplier), 4.5% moisture, <2.5% protein, 3.5% ash, and 0.5% fat. According to the supplier, molecular weights of 1.0×10^6 are routinely achieved, which characterizes PromOat as a high molecular weight β-glucan.

Equilibrium Dialysis Assays with Glucagel. The phenolic compounds and β-glucan were dialyzed in sterile 0.5% DMSO and 10 mM Tricine buffer (pH ∼7) using 1 mL in-line equilibrium dialysis cells (Bel-Art Products, Pequannock, NJ) and dialysis membranes with size exclusion of 6−8 kDa for globular molecules (Spectrum Laboratories Inc., Breda, The Netherlands). According to the manufacturer, this cutoff is calculated for proteins, and for dextrans the cutoff is calculated to be 1−1.5 kDa. Glucagel 5% (w/v) was dispersed in distilled water in a 50 mL conical flask covered with aluminum foil and then heated for 30 min at 80−82 °C using a hot plate stirrer controlled by a thermostat. A sufficient quantity of phenolic derivatives or bile salts was dissolved in 1 mL of DMSO. Two 5 mL samples of 2.5% (w/v) β-glucan were prepared in the described buffer. To one sample were added phenolic derivatives or bile salts to achieve 2 and 10 mM final concentrations, respectively. Additional 5 mL samples of 2 mM phenolic derivative or 10 mM bile salt were prepared in buffer. The concentrations of the phenolic derivative relate to the maximum possible concentration to be kept in solution. The concentrations of the bile salts were chosen as those previously used in similar experiments (*14, 20*). All samples were stored at 4 °C prior to performance of the dialysis experiment. To minimize the differences in the gel-setting samples, the dialysis experiments were initiated 12 and 36 h (1 and 2 days) after sample preparation. All samples were then heated to 25 °C and thoroughly mixed using a vortex mixer, again to keep the gel setting to a minimum. One milliliter of each sample was applied into two individual sample chambers of the dialysis cells under aseptic conditions and subsequently dialyzed against 1 mL of sterile buffer in the assay chambers. The time of application of the first sample was denoted time zero, and a time gap of 30 s between applications of samples was maintained, thus keeping track of the exact dialysis time for each sample. After application, the dialysis cells were quickly transferred to a thermostat-regulated rotating water bath at 37 °C. Preliminary kinetic studies showed that dialysis equilibrium was reached in <5 h. Aliquots (10 μL) were withdrawn from the assay chamber after 15, 30, 45, 60, 75, 90, 120, 150, 180, 240, and 300 min (5 h), transferred to a 96-well plate (Nunc A/S, Roskilde, Denmark), and diluted with distilled water in the ratio 1:9. Absorbance of the diluted dialysate sample and of a reference of distilled water was measured at 210, 220, and 280 nm on a Spectra Max 190 plate reader (Molecular Devices Corp., Sunnyvale, CA) depending on the wavelength providing maximum absorbance for the phenolic derivative used in the individual experiments. All compounds were tested four times; that is, two replicates were dialyzed 12 h after mixing and two replicates of the same sample were dialyzed 36 h after mixing. The 12 h (day 1) and 36 h (day 2) dialysis experiments were treated as two separate experiments.

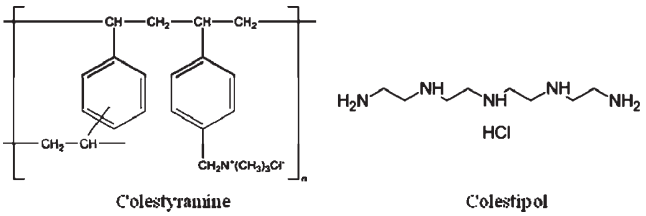


Figure 2. Chemical structures of colestyramine and colestipol.

Enzymatic Breakdown Studies. The studies were performed using Glucagel and the methodology reported above. After aliquot sampling at 240 min, 2 μL of 50 mU/mL Lichenase enzyme (Megazyme, Ireland) or 2 μL of a 1:100 dilution of the original product solution of Viscozyme L (batch KTN02140, Novozymes, Denmark) was added to the β-glucan-containing compartment of the dialysis instrument. Aliquots (10 μL) were withdrawn from the assay chamber after 255, 270, 285, 300, 315, 330, 360, 390, 420, and 450 min from the start of the experiment. The aliquots were analyzed as described above.

Comparison of Glucagel and the Bile Salt Sequestrant Drugs Colestyramine and Colestipol Hydrochloride. The studies were performed using the methodology reported above and a 2.5% (w/v) assay concentration of Glucagel. Using the commercial drug formulations, 1% (w/v) solutions of colestyramine (Questran, Bristol-Myers Squibb) and colestipol hydrochloride (Lestid, Pfizer) were prepared. Chemical structures are shown in **Figure 2**. The drugs were obtained directly from the manufacturer through the pharmacy at the University of Copenhagen, and the 1% (w/v) concentration used reflects the recommend dose of the drugs.

Comparative Equilibrium Dialysis Assays with Glucagel and PromOat. This series of dialysis experiments was carried out as above except that the concentrations of both β-glucan preparations were reduced to 1% (w/v) (the glucan concentration in the assay) to circumvent handling problems related to the high specific viscosity of PromOat.

Exponential Curve Fitting of Dialysis Data. Initial data handling was conducted in Excel (Microsoft Office 2003) where measured absorbencies were corrected for the background absorbance, mean values were calculated for the two replicates, and moving averages were calculated over five continuous measurements. Regression analysis procedures were employed to explore the relationship between dialysate absorbance and dialysis time. Analysis of moving averages reduces the impact of nonrelevant dialysis information, leads to simpler and more robust data sets for regression models, and improves interpretation of the dialysis data. Generally, linearity of the absorbance to concentration relationship over the concentration range of 0−2 mM for the phenolic compounds was observed when measured in the 210−280 nm range. Subsequent mathematical modeling was conducted in SigmaPlot (version 4.01). A curve derived from the exponential equation

$$\text{absorbance} = A(1 - e^{-kt}) \quad (1)$$

was fitted to data from each individual dialysis experiment, where A is the asymptotic or equilibrium UV absorbance value, t is the dialysis time, and k is the dialysis rate constant. All model fits were evaluated using correlation coefficients (r^2). Asymptotic values were compared between dialyses of each of the pure phenolic derivatives and for the phenolics mixed with β-glucan to quantitatively determine the level of dialysate retention (ΔA) by β-glucan. Asymptotic values derived from the pure β-glucan samples were subtracted from all of the β-glucan/aroma compound mixtures to account for the dialysate from β-glucan alone.

Multivariate Data Analysis. Dialysate retention data from day 1 and day 2 dialysis experiments were subjected to principal component analysis (PCA) and partial least-squares (PLS) regression.

In PCA, a data matrix is decomposed by consecutive orthogonal extraction of the largest variation (principal components, PCs) in data

Table 2. Percentage Dialysate Retention Based on the Asymptotic UV Absorbance (Δ*A*) from Mixed Solutions of 2.5% (w/v) Glucagel and Six Selected Glucosides of Phenolic Compounds (Means of Minimum Two Replicates, All Meseaured at 280 nm)

compd (name/no.)	aglycon no.	dialysate retention (Δ*A*, %)		
		day 1	day 2	Δ (2 − 1)
4-β-D-glucopyranosyloxy-3-methoxybenzaldehyde (**22**)	17	7	9	2
ethyl 4-β-D-glucopyranosyloxy-3-methoxybenzoate (**23**)	20	−8	−5	3
4′-β-D-glucopyranosyloxy-3′-methoxyacetophenone (**24**)	1	−3	2	5
methyl 4-β-D-glucopyranosyloxy-3-methoxybenzoate (**25**)	19	−12	−10	2
4-β-D-glucopyranosyloxybenzyl alcohol (**26**)	21	−10	−9	1
3-ethoxy-4-β-D-glucopyranosyloxybenzaldehyde (**27**)	7	−9	−7	2

until the variation left is unsystematic. The loading vectors can be considered as pure hidden profiles that are common to all measurements. Two-dimensional scatter plots of the score vectors show the covariance between samples, providing a characterization of data. The purpose of PLS regression is to build a linear model enabling prediction of a desired chemical/physical characteristic (*Y*) from measured data (*X*). During the regression, *X* is decomposed as in PCA, but the PCs are found as the underlying structures that covary best with the *Y* variable (*25*).

All molecular structures in this study were optimized with MM3* in MacroModel (*26*). A total of 234 molecular descriptors were calculated for the phenolic derivatives and glucosides with CDK (*27*) and QikProp (*28*). CDK descriptors are divided into five major classes: electronic (atomic polarizabilities, bond polarizabilities, charged partial surface areas, hydrogen bond acceptors, and hydrogen bond donors), geometrical (geometrical index, length over breadth, moments of inertia, and Petitjean shape indices), constitutional (*A*log*P*, bond, element, and atom type counts, largest chain, Lipinski's rule of five, rotable bonds count, *X*log*P*, molecular weight), hybrid (BCUT and WHIM), and topological (carbon types, Chi indices, eccentric connectivity index, fragment complexity, Kier and Hall molecular shape indices, topological polar surface area, Wiener numbers, Zagreb index, and Moreau-Broto autocorrelation descriptors). QikProp provides approximately 40 descriptors, of which several (e.g., predicted brain/blood partition coefficient, QPLogBB) are of pharmaceutical relevance, whereas others [e.g., PM3 calculated ionization potential, IP(eV)] are of a more general nature. Without any a priori knowledge of the mechanisms involved, we anticipate that this diverse set of molecular descriptors captures information relevant to the dialysis characteristics. To reduce the amount of noise in the descriptor matrix and improve the subsequent interpretation of PCA plots and PLS regressions, a simple two-step variable selection scheme was employed. First, descriptors were deleted unless they assumed distinct values for at least 12 samples. Second, the Pearson product moment correlation coefficient, *r*, between response (Δ*A* or k_d) and the descriptors was evaluated. In cases when *r* fell below a certain threshold (0.5 for Δ*A*, 0.2 for k_d), the corresponding descriptor was deleted. The lower threshold for k_d was required, because response−descriptor correlations were very low in this case. The number of descriptors produced by the variable selection was 62 for day 1 Δ*A*, 16 for day 2 Δ*A*, 82 for day 1 k_d, and 91 for day 2 k_d. The variables were autoscaled prior to data analysis, and full (leave-one-out) cross-validation was used. The PLS models were evaluated on the basis of the root-mean-square error of cross-validation (RMSECV), which is the estimation of the error of the predicted values. Multivariate data analysis was performed using Unscrambler (*29*), Matlab (*30*), and Latentix (*31*).

Molecular Mass Determination. The molecular mass of the β-glucans was estimated by size exclusion chromatography (SEC) using a Superdex 200 column (16 mm × 60 cm, GE Healthcare Bio-Sciences AB, Uppsala, Sweden) fitted with a refractive index detector. Superdex 200 is a cross-linked agarose and dextran material with a nominal bead size of 13 μm, a pore size of 100−120 Å, and an optimal separation range of 10,000−600,000 Da. The mobile phase consisted of 50 mM ammonium formate and 200 mM NaCl, and the column was eluted at room temperature at 1.6 mL/h. Dextrose standards of 5, 12, 25, 50, 80, and 150 kDa (Fluka, Buchs, Switzerland) were used for calibration.

RESULTS

Vanillin and 20 other phenolic derivatives (**1**−**21**), 6 phenolic glucosides (**22**−**27**), and 4 bile salts were tested with respect to their ability to bind to Glucagel using the dialysis setup. The phenolic derivatives can be divided into different chemical classes according to the functional groups present in the different compounds. All of them possess a central benzene ring as the core structure. The benzene ring is substituted with a minimum of two groups, of which one would be an oxy group. The aglycon structures may be grouped into 2 ketones (**1**, **11**), 4 aldehydes (**2**, **7**, **13**, **17**), of which 2 were hydroxylated on the benzene ring (**2**, **17**), 10 phenols (**3**−**6**, **8**−**10**, **15**, **16**, **21**), of which 2 had one additional hydroxyl group (**5**, **6**), 3 esters of either ethyl or methyl character (**12**, **19**, **20**), and 2 aromatic acids (**14**, **18**) (**Tables 1** and **2**). The molecules cover a broad range of log *P* values ranging from −0.5 for the glucosides to 5.1 for compound **4**. With respect to their physical dimensions, the phenolic derivatives and bile salts also span a significant range from the large bile salts (18−19 Å in diameter) to the smallest, being compound **21** (7 Å in diameter).

The ability of all of the compounds **1**−**27** to bind to Glucagel was analyzed in dialysis experiments. Exponential curve fitting of all dialysis data was performed to fit eq 1. Dialysis curves of **8** and **10** in the presence of 2.5% (w/v) Glucagel as well as the exponential curve fit are presented in **Figure 3** as typical examples. Dialysate retentions (Δ*A*) and dialysis rate constants (k_d) from the screening of all 21 phenolic derivatives are presented in **Table 1**. The values represent the mean of two replicates. Values obtained by dialysis of β-glucan in the absence of any added compound were subtracted from the values obtained with added compounds. All dialysis experiments exhibited patterns similar to those presented in **Figure 3**. The patterns are composed of an initial steep slope during the first ∼15−90 min and an asymptotic convergence toward a maximum after 90−300 min. Generally, there was good agreement between the dialysis curve and the exponential fit. Correlation coefficients from the curve fits were >0.95 in all dialysis experiments. At equilibrium/asymptotic level, the absorbance of **8** reaches a value of 0.96, whereas the asymptotic value of **8** + 2.5% (w/v) Glucagel is 0.76. This gives a relative difference of 22%, which corresponds to the retention of dialysate (**8**) by the β-glucan, as seen in **Table 1**. The different classes of chemical compounds within the 21 phenolic derivatives such as ketones, aldehydes, phenols, esters, and acids did not differentiate significantly from each other and within the groups with respect to being retained by the barley β-glucan. Only four compounds (**1**, **13**, **14**, **20**) gave rise to a Δ*A* > 30. All four compounds shared the presence of a hydroxyl group at position 4 and a CHO group at position 1. However, other compounds such as **18**, **5**, and **19** also possess these features and give rise to much lower Δ*A* values. The difference between day 1 and day 2 retentions (Δ 2 − 1) showed a weak tendency of increased

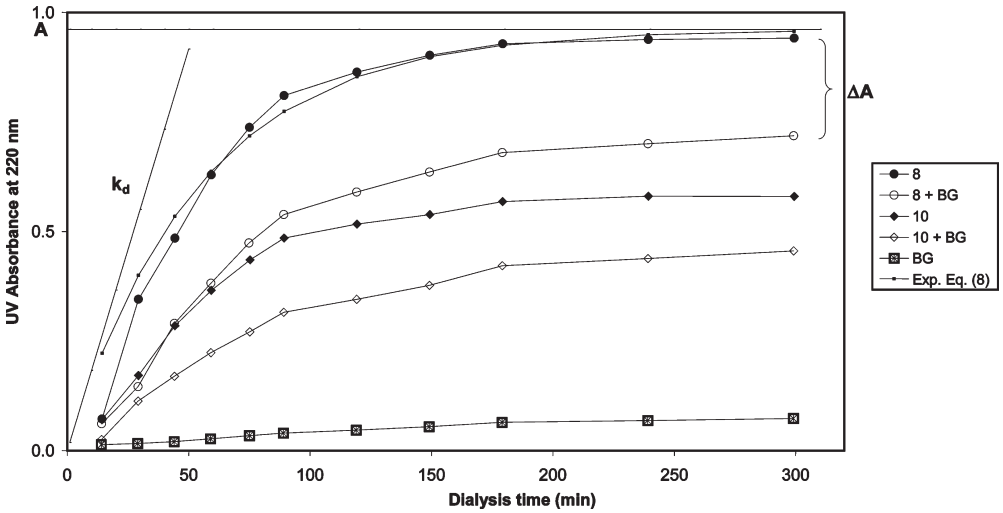


Figure 3. Dialysis curve of compounds **8** and **10** in the presence of 2.5% (w/v) Glucagel (BG). The line of k_d is shown, and the horizontal line illustrates the maximum absorbance to be obtained at equilibrium of the dialysis of the exponential data curve.

Table 3. Percentage Dialysate Retention Based on the Asymptotic UV Absorbance (ΔA) and Dialysis Rate Constants (k_d) from Mixed Solutions of 2.5% (w/v) Glucagel and Four Different Bile Salts and Tryptophan (Means of Minimum Two Replicates)

	dialysate retention (ΔA, %)			dialysis rate constant (k_d)		
compd	day 1	day 2	Δ (2 − 1)	day 1	day 2	Δ (2 − 1)
deoxycholic acid	69	76	7	0.008	0.009	0.001
glycocholic acid	78	86	8	0.008	0.007	−0.001
taurocholic acid	73	81	8	0.013	0.011	−0.002
cholic acid	38	42	4	0.012	0.011	−0.001
tryptophan	23	28	5	0.005	0.004	−0.001

dialysate retention on day 2, that is, positive Δ values. Comparison of the k_d values from days 1 and 2 (Δ 2 − 1) indicated lower dialysis rates at day 2. For the more hydrophilic phenolic derivative glucosides (**Table 2**) no or only very weak retention was observed. Compared with the values obtained for the aglycons, the glucosides are significantly less adsorbed by the fiber.

Four bile salts were tested in the dialysis equilibrium system (**Table 3**). All four bile salts showed good retention in the system with ΔA from 42 to 86. This confirmed the published in vivo evidence for the ability of dietary fibers to adsorb bile salts (*4, 5*). Tryptophan has previously been used as a reference molecule in dialysis tests (*14*) and was also tested in this study and showed a retention of 23 with is compliance with previous studies (*14*).

Comparison of Glucagel and PromOat. To evaluate the ability of a different commercial β-glucan to bind phenolic derivatives, the oat β-glucan product PromOat was investigated and compared to the barley β-glucan product Glucagel. The results (**Table 4**) show no significant differences between the two products with respect to retention of phenolic derivatives under the conditions used in our dialysis equilibrium system. The secondary structure, product composition, and viscosity of these two fiber preparations are not the same (*32*), which may be important for the adsorption. However, this is not captured by our dialysis assay.

Comparison with Commercial Bile Salt Sequestrant Drugs. To evaluate the efficacy of the fiber to adsorb bile salts and other small compounds, a comparative study with the commercial drugs colestyramine and colestipol hydrochloride was carried out. The results presented in **Table 5** show that for

Table 4. Comparison of Glucagel and PromOat (Means of Minimum Two Replicates)[a]

	dialysate retention	
compd (name/no.)	Glucagel	PromOat
3-ethoxy-4-hydroxybenzaldehyde (**7**)	23	26
ethyl 4-hydroxy-3-methoxybenzoate (**20**)	15	19
4-hydroxy-3-methoxybenzaldehyde (**17**)	14	17
4′-hydroxy-3′-methoxyacetophenone (**1**)	10	14

[a] Percentage dialysate retention based on the asymptotic UV absorbance (ΔA) from mixed solutions of 1% (w/v) β-glucan solutions and four different phenolic compounds.

Table 5. Percentage Dialysate Retention Based on the Asymptotic UV Absorbance (ΔA) of Four Phenolic Compounds and Four Bile Salts by Glucagel, Colestyramine, and Colestipol Hydrochloride

	dialysate retention		
compd (name/no.)	Glucagel	colestyramine	colestipol hydrochloride
3-ethoxy-4-hydroxybenzaldehyde (**7**)	23	2	4
ethyl 4-hydroxy-3-methoxybenzoate (**20**)	15	3	6
3,5-dimethoxy-4-hydroxybenzoic acid (**14**)	40	95	93
4-hydroxy-3-methoxybenzoic acid (**18**)	10	96	92
deoxycholic acid	69	89	85
glycocholic acid	78	95	97
taurocholic acid	73	94	93
cholic acid	38	96	95

nonacid compounds (**7**, **20**), the fibers have higher retention capability than the drugs. The drugs are characterized as having an anionic exchange nature. In agreement, the two drugs were able to retain the acidic compounds (**14**, **18**) up to 9 times better than the fiber. This demonstrates that ionic forces are not the main property responsible for the adsorption of small molecules to the fiber. This also confirms that the dialysis assay works with matrices other than the fibers.

Enzymatic Breakdown of the Fibers. To investigate the importance of fiber molecular mass for the adsorption ability, the rerelease of the small compounds from the fiber matrix was measured after partial enzymatic breakdown of the fibers (**Figure 4**). Some, but not all, of the compounds retained in the experiments with intact fibers were released as a result of enzyme breakdown of the fibers and moved freely between the two dialysis compartments. This shows that different parameters contribute to retention of the compounds. The fibers were broken

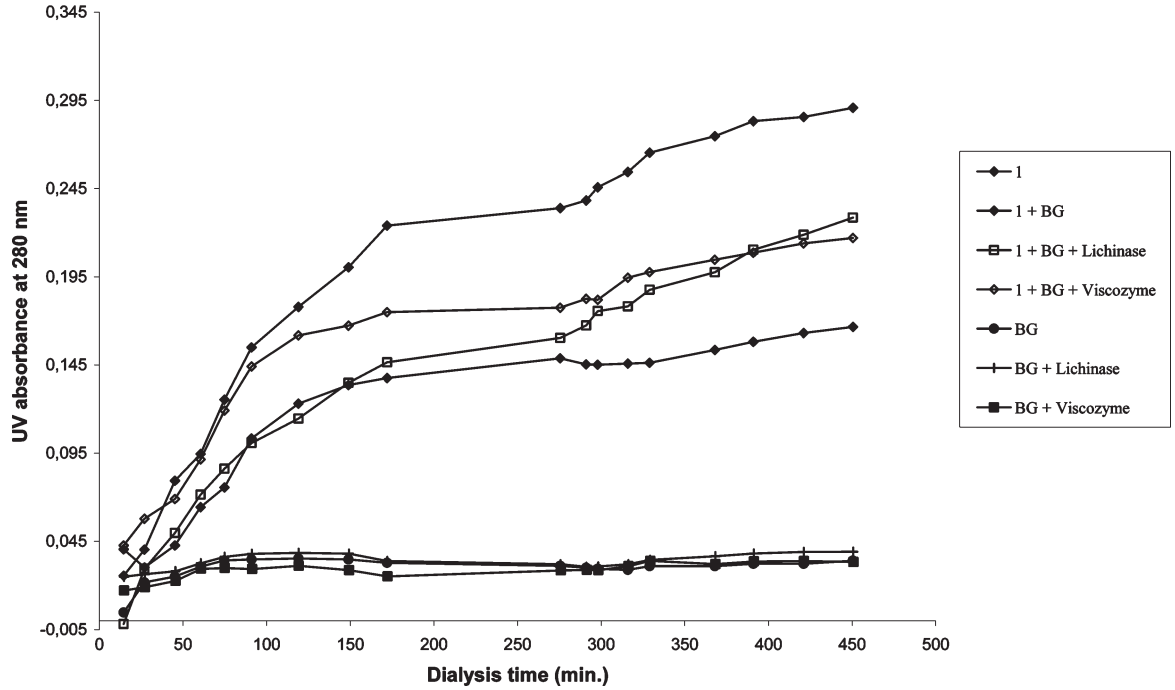


Figure 4. Effect of enzymatic breakdown of the barley β-glucan on its ability to bind phenolic derivatives as determined by dialysis assays. The experiment was carried out using 2.5% (w/v) barley β-glucan Glucagel (BG) as fiber matrix and 4′-hydroxy-3′-methoxyacetophenone (**1**) in the 2.5% (w/v) Lichenase and Viscozyme enzymatic breakdown of the fibers.

down to masses of <5000 Da, which supposedly disrupt the majority of the tertiary structures and some of the secondary structures of glucans. **Figure 4** also shows that even though the fibers were broken down, they were retained in the sample champers, indicating that they could not move freely as the small molecules can.

Multivariate Data Analysis. PCA was conducted using the dialysate retention data sets obtained at day 1 as well as day 2 using the dialysate retentions (ΔA) and rates (k_d). The correlation plot of the free aglycone, glycoside, and of the bile salt data (ΔA and k_d) (**Figure 5**) showed a marked distinction between these three groups of compounds according to ΔA. Additionally, the plot shows a tendency of decreasing k_d values between the data obtained at days 1 and 2 (arrows directed downward) and a weak trend of increasing ΔA values from day 1 to day 2 (arrows directed to the right). No obvious groupings within the 21 tested structures were observed. **Figure 6** shows the PCA score plot for the 21 phenolic derivatives and 6 glucosides represented by 62 descriptors selected using the value of day 1 ΔA as explained under Materials and Methods. The samples have been colored according to the value of ΔA. Although phenolic derivatives with high and low values of ΔA have scores on the extreme left and extreme right of PC1, respectively, the change in ΔA along PC1 is not systematic. Inspection of the remaining PCs did not reveal any improvements in describing the variation of ΔA. The PCA plots based on descriptors selected for the explanation of differences between day 1 and day 2 k_d values did not reveal any structure. The fact that a lowered threshold for correlation with the dependent variable had to be employed in the descriptor selection step supports the notion that a proper explanation of k_d cannot be achieved with the current set of descriptors. PLS regression for the prediction of the day 1 retention value seemed initially to hold some promise, but permutation testing of the model revealed that it was based on a chance correlation. The low quality of the model was also reflected in similar regression coefficients, indicating that no important descriptors could be singled out. The situation did not improve with respect to the remaining independent variables for days 1 and 2. In conclusion, the physical mechanisms involved in fiber retention of the compounds as monitored in the dialysis experiments cannot be explained with our current multivariate data analytical approach. Data analytical exploration of fiber-binding properties has to await the development of a more suitable set of descriptors. At this point the nature of such descriptors is unknown.

DISCUSSION

The metabolic health benefits and viscous properties of β-glucans have been reported by several investigators (*2*−*8*). However, their potential uses, mechanisms of action, and means of incorporation into foods and diets require further exploration. The aim of the present study was to investigate the molecular interactions between β-glucans and selected classes of small molecules. The approach of equilibrium dialysis was chosen because it has been proven useful in interaction studies on bile salts and barley β-glucan (*14*) and because the approach offers a relatively fast method for the analysis of large series of small samples. Dialysis conditions were set to mimic physiological conditions in the sense of continuous movement, a temperature of 37 °C, and pH of 7, which could provide evidence for β-glucan behavior in solutions in the GI tract independent of enzymatic degradation. However, this very simple in vitro study cannot provide an exact description of the physiological actions of β-glucans but rather affords knowledge on their molecular affinities toward small molecules. The effect of hydrocolloids on aroma release from food may be due to numerous mechanisms; one is the physical entrapment of aroma within the food matrix (*15*). Another mechanism involves chemical interaction between the aroma compound and the hydrocolloid components, for example, β-glucan (*16*). In the current study, both types of interactions were studied experimentally.

Along with the test of the 21 phenolic compounds, 6 β-glucosides of these compounds were tested. This would

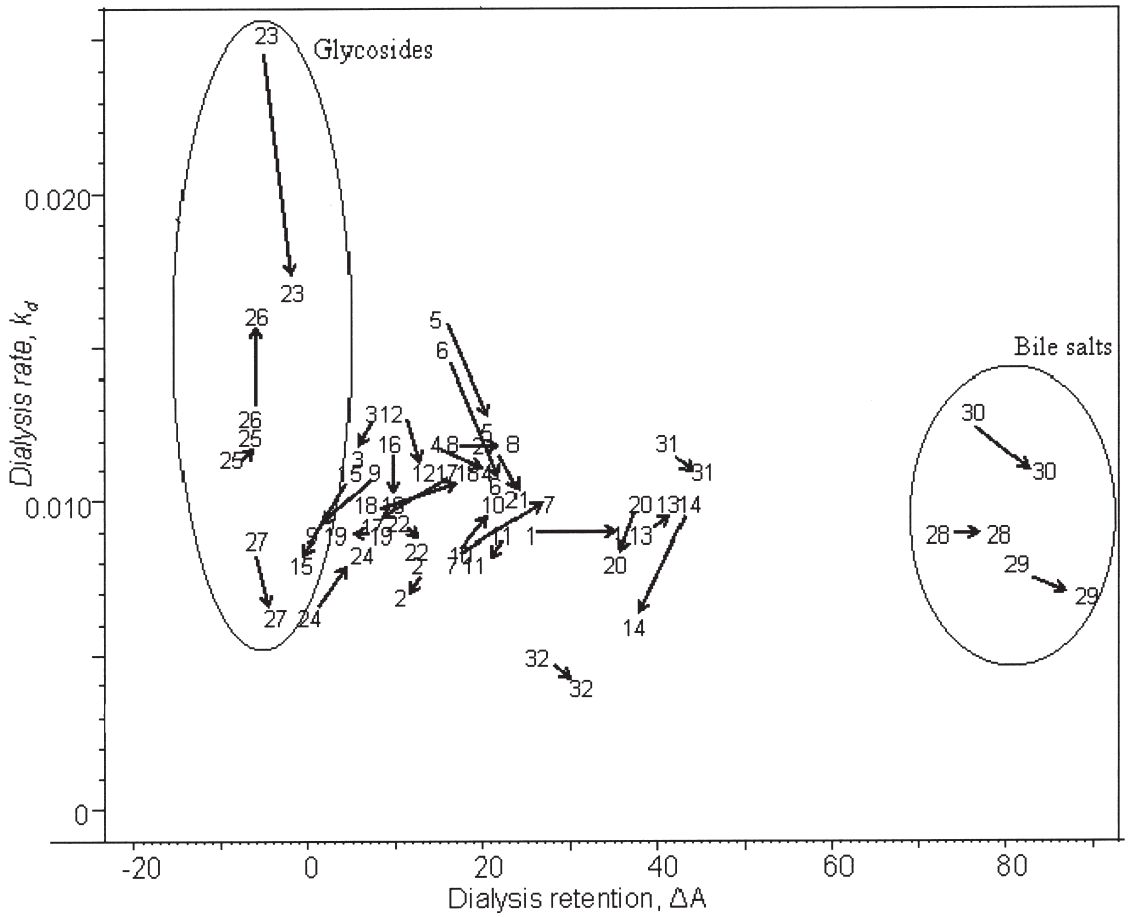


Figure 5. Correlation plot of day 1 dialysate retentions (ΔA) and dialysis rates (k_d). Numbers refer to the compounds listed in **Tables 1**−**3**, and arrows show the direction of movement in the plot between day 1 and day 2 data.

represent the native form in which the aroma compounds are present in the plants, and it was therefore of interest to test whether they are retained in the same fashion as the aglycons. As seen in **Table 2**, the glucosides are not retained or are retained at only a very low degree compared to the retention of the aglycons. This indicates that hydrophobic properties of the small molecules are of importance. Boland et al. (*15*) observed that of 11 flavor compounds tested, the hydrophobic compounds had the significantly lowest partition coefficients, that is, aroma release in a gelatin gel. In this study, an observed increased percentage of retention as a result of increased log *P* values for compounds **3** and **4** would suggest a similar contribution of hydrophobic properties to fiber retention. No such easily envisioned trends were observed in the present study using either simple curve fitting of retention versus log *P* or multivariate data analysis. Story and Kritchevsky (*8*) stated that the hydrophobic properties of the bile salts tested in their publication did not correlate directly with the binding to the tested fibers. This is in good agreement with the observations in our study both for the bile salts and for the phenolic compounds.

The differences observed in percentage of fiber retention of the different phenolic derivatives tested in the day 1 experiment (**Table 1**) do indeed indicate that the ability of the fibers to retain the different derivatives could be of a different nature depending on the physicochemical properties of each individual phenolic derivative. The corresponding glucosides have a significantly lower log *P* value than the aglycons, whereas at the same time containing a sugar moiety that might be expected to be able to interact with the β-glucan via hydrogen bond formation. As seen in **Table 2**, the glucosides are not retained by the fiber. High water solubility could therefore be one of the properties of the small molecule that would reduce possible interactions with the fiber. Among the tested phenolics the four compounds **1**, **13**, **14**, and **20** had significantly higher retentions than the others. The four molecules possess different functional groups at position 1, but all share a hydroxyl group at position 4. The log *P* values for the four molecules range from 0.83 to 1.68 and cannot be correlated to the significantly higher retention of these molecules compared to the remaining 17 compounds. The four molecules do not possess different physical dimensions, nor do they contain other functional groups compared to the rest of compounds tested. This indicates that multiple different binding properties are of importance for the retention of the small molecules to the β-glucan fiber.

The differences between the results obtained at day 1 and day 2 are most likely due to time-dependent changes in the β-glucan matrix, for example, network formation and increased rigidity in the β-glucan solution induced by the increased incubation time prior to dialysis of the day 2 samples (*32*). In a preliminary study on the viscous properties of a 2.5% (w/v) Glucagel solution, the viscosity (at 37 °C and 30 s^{-1}) of the sample increased approximately 6-fold from day 1 to day 2, which would agree with the slower dialysis rates observed for the day 2 samples. The matrix-dependent changes in dialysate retention indicate that some of the retention is due to the hygroscopic and tertiary structure of the β-glucan. The current findings are in agreement with those of Boland et al. (*15*), who found that flavor release was significantly affected by the texture of gelatin, starch, and pectin gels. The most rigid gel showed the lowest flavor release.

Figure 6. PCA score plot. The value of ΔA on day 1 has been used to color the samples (21 phenolic derivatives and 6 phenolic glucosides). The PCA is based on 62 calculated physicochemical descriptors, selected as described under Materials and Methods. Numbers refer to the compounds listed in **Tables 1** and **2**.

Because the term β-glucan is not a uniform definition, this study included a comparative study of the two commercial products Glucagel (barley) and Promoat (oat). The two matrices of these products are chemically quite different, for example, with respect to purity and molecular masses (*32*). Despite these differences, the two β-glucans show nearly identical properties with respect to their ability to retain phenolic derivatives in the dialysis equilibrium assays. This suggests that the retention of the different phenolics is due to several and interacting properties of the fiber matrix and not a few single parameters.

The efficacy of the β-glucan fibers to adsorb bile salts and other small compounds was also evaluated in comparison with the commercial drugs colestyramine and colestipolhydrochloride. This study indicated that ionic interactions are not the main property responsible for the adsorption of small molecules to the fiber, again confirming that the retention is composed of multiple properties. This also confirms that the dialysis assay works with matrices other than the fibers.

The importance of the fiber molecular mass and thereby the physical size of the molecules for the adsorption ability was also evaluated. A clear rerelease of some, but not all, of the retained compounds was observed upon enzymatic degradation of the barley β-glucan-based matrix. Partial rerelease of the small compounds again indicates that various factors contribute to retention of the different phenolic derivatives. The fibers were broken down to an extent that would disrupt the tertiary structure along with some parts of the secondary, which can explain why some phenolic derivatives were retained.

PCA and PLS regression analyses are powerful tools for extraction of important variances in a data matrix consisting of many variables. In the present study, comparison of β-glucan binding of different phenolic derivatives was conducted using molecular modeling and multivariable data analysis. No strong tendency of sample grouping was found by PCA either for day 1 or for day 2. This implicates that the employed set of descriptors is not well suited for explaining the variation in ΔA and k_d. Additionally, no reliable model fit was found using PLS regression to dialysis data. This could be due to the use of insufficient or ineffective descriptors, the need for more data (a larger number of phenolics screened in the dialysis assay) to strengthen the robustness of prediction, or measurement errors within the present data set from the dialysis assay. Even though no reliable prediction model was found, indications of a correlation between the binding data and some of the molecular descriptors were evident. This indicates that multiple parameters are involved in determining the binding of small molecules to fibers and that the multiplicity of parameters involved obscures the correlations of the observed binding to specific descriptors.

The dialysis data presented provide some information on the complex mechanisms controlling the ability of β-glucans to bind low mass compounds. To more accurately determine the nature of the interactions between β-glucans and low mass

compounds, a range of advanced spectroscopy and molecular modeling methodologies will need to be introduced.

In conclusion, the retention of aroma compounds by β-glucan is of great interest from a food composition point of view and from a health perspective. In recent years, the increased health consciousness among consumers has led to extended additional use of hydrocolloids as replacements for fats. The food industry would benefit greatly from an improved understanding of the mechanisms involved in the flavor retention and release from β-glucan matrices along with the nature of the interactions occurring in the GI tract. Knowledge of the action of model compounds in β-glucan matrices can provide general information on the β-glucan affinity toward small molecules, which would be applicable in the studies of, for example, the bioavailability of natural compounds found in association with β-glucans and for health claims on β-glucan.

It was confirmed that β-glucans from barley and oat are able to adsorb bile salts, and for the first time β-glucans' ability to absorb vanillin and other phenolic compounds was demonstrated. The retention in our newly developed dialysis assay depends on numerous and interacting physicochemical properties of the small molecules. The interaction could not be explained by simple correlation to any of the descriptors included in the multivariable data analysis, and the results could not confirm or disprove the previously described hydrophobic binding or micelle capture of the small molecules to the β-glucan.

LITERATURE CITED

(1) Cookson, F. B.; Altschul, R.; Fedoroff, S. The effects of alfalfa on serum cholesterol and in modifying or preventing cholesterol-induced atherosclerosis in rabbits. *J. Atheroscler. Res.* **1967**, *7*, 69–81.

(2) Eastwood, M. A. A molecular biological basis for the nutritional and pharmacological benefits of dietary plants. *Q. J. Med.* **2001**, *94*, 45–48.

(3) Eastwood, M. A.; Kritchevsky, D. Dietary fibre: how did we get where we are? *Annu. Rev. Nutr.* **2005**, *25*, 1–8.

(4) Kritchevsky, D. Cereal fibre and lipidemia. *Cereal Foods World* **1997**, *42*, 81–85.

(5) Story, J. A.; Watterson, J. J.; Matheson, H. B.; Furumoto, E. J. Dietary fibre and bile acid metabolism—an update. *Adv. Exp. Med. Biol.* **1997**, *427*, 259–266.

(6) Beer, M. U.; Arrigoni, E.; Amado, R. Effects of oat gum on blood cholesterol levels in healthy-young men. *Eur. J. Clin. Nut.* **1995**, *49*, 517–22.

(7) Yang, J. L.; Kim, Y. H.; Lee, H. S.; Lee, M. S.; Moon, Y. K. Barley β-glucan lowers serum cholesterol based on the up-regulation of cholesterol 7 α-hydroxylase activity and mRNA abundance in cholesterol-fed rats. *J. Nut. Sci. Vitamin.* **2003**, *49*, 381–387.

(8) Story, J. A.; Kritchevsky, D. Comparison of the binding of various bile acids and bile salts in vitro by several types of fibre. *J. Nutr.* **1976**, *106*, 1292–1294.

(9) Eastwood, M. A.; Anderson, R.; Mitchell, W. D.; Robertson, J.; Podock, S. A method to measure the adsorption of bile salts to vegetable fibre of differing water holding capacity. *J. Nutr.* **1976**, *106*, 1429–1432.

(10) Kahlon, T. S.; Berrios, J. de J.; Smith, G. E.; Pan, J. L. In vitro bile acid binding capacity of wheat bran extruded at five specific mechanical energy levels. *Cereal Chem.* **2006**, *83*, 157–160.

(11) Vahouny, G. V.; Tombes, R.; Cassidy, M. M.; Kritchevsky, D.; Gallo, L. L. Dietary fibres: V. Binding of bile salts, phospholipids and cholesterol from mixed micelles by bile acid sequestrants and dietary fibres. *Lipids* **1980**, *15*, 1012–1018.

(12) Story, J. A.; Watterson, J. J.; Matheson, H. B.; Furumoto, E. J. Dietary fibre and bile acid metabolism. *Adv. Exp. Med. Biol.* **1990**, *270*, 43–48.

(13) Vahouny, G. V.; Khalafi, R.; Satchithanandam, S.; Watkins, D. W.; Story, J. A.; Cassidy, M. M.; Kritchevsky, D. Dietary fibre supplementation and fecal bile acids, neutral steroids and divalent cations in rats. *J. Nutr.* **1987**, *117*, 2009–2015.

(14) Coles, G. D.; Eady, S. L.; Coles, T. K.; Boone, D. J. J.; Furneaux, R. H. β-Glucan binds bile salts. In *Cereals "96", Proceedings of The 46th Australian Cereal Chemistry Conference*; Wringley, C. W., Ed.; RACI Cereal Chemistry Division: Melbourne, Australia, 1996; 180−183.

(15) Boland, A. B.; Buhr, K.; Giannouli, P.; van Ruth, S. M. Influence of gelatin, starch, pectin and artificial saliva on the release of 11 flavour compounds from model gel systems. *Food Chem.* **2004**, *86*, 401–11.

(16) Malkki, Y.; Heinio, R. L.; Autio, K. Influence of oat gum, guar gum and carboxymethyl cellulose on the perception of sweetness and flavor. *Food Hydrocolloids* **1993**, *6*, 525–532.

(17) Pangborn, R. M.; Gibbs, Z. M.; Tassan, C. Effect of hydrocolloids on apparent viscosity and sensory properties of selected beverages. *J. Texture Stud.* **1978**, *9*, 415–436.

(18) Juteau, A.; Cayot, N.; Chabanet, C.; Doublier, J. L.; Guichard, E. Flavour release from polysaccharide gels: different approaches for the determination of kinetic parameters. *Trends Food Sci. Technol.* **2004**, *15*, 394–402.

(19) Yven, C.; Guichard, E.; Giboreau, A.; Roberts, D. D. Assessment of interactions between hydrocolloids and flavour compounds by sensory, headspace, and binding methodologies. *J. Agric. Food Chem.* **1998**, *46*, 1510–1514.

(20) Bowles, R. K.; Morgan, K. R.; Furneaux, R. H.; Coles, G. D. C-13 CP/MAS NMR study of the interaction of bile acids with barley β-D-glucan. *Carbohydr. Polym.* **1996**, *29*, 7–10.

(21) Guichard, E.; Etievant, P. Measurement of interactions between polysaccharides and flavour compounds by exclusion size chromatography: advantages and limits. *Nahrung−Food* **1998**, *42*, 376–379.

(22) Langourieux, S.; Crouzet, J. Study of aroma compounds polysaccharides interactions by dynamic exponential dilution. *Food Sci. Technol.−Lebensm. Wiss. Technol.* **1994**, *27*, 544–549.

(23) Druaux, C.; Voilley, A. Effect of food composition and microstructure on volatile flavour release. *Trends Food Sci. Technol.* **1997**, *8*, 364–368.

(24) Mauthner, F. Über neue synthetische Glucoside. *J. Prakt. Chem. (Neue Folge)* **1912**, *85*, 564–568.

(25) Martens, H.; Næs, T. *Multivariate Calibration*; Wiley: Chichester, U.K., 1989.

(26) *Macromodel*, vers. 9.6; Schrödinger: New York, 2008.

(27) Steinbeck, C.; Han, Y.; Kuhn, S.; Horlacher, O.; Luttmann, E.; Willighagen, E. The Chemistry Development Kit (CDK): an open-source Java library for chemo- and bioinformatics. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 493–500.

(28) *QikProp*, vers. 2.5; Schrödinger: New York, 2005.

(29) *The Unscrambler*, vers. 7.6.; CAMO ASA: Trondheim, Norway.

(30) *Matlab*, vers. 5.3; MathWorks: Natick, MA.

(31) *Latentix*, vers. 1.0; Latent5: Copenhagen, Denmark.

(32) Nielsen, M. S.; Jespersen, B. M.; Karlsson, A.; Simonsen, H. T.; Møller, B. L.; Lærke, H. N.; Larsen, F. H.; Engelsen, S. B. Comparative structural and rheological studies on barley and oat β-glucan products: Glucagel and PromOat. *Cereal Chem.*, **2008**, submitted for publication.

Received for review July 7, 2008. Revised manuscript received October 14, 2008. Accepted October 25, 2008. We thank the programs Build Your Food supported by the Danish Council for Strategic Research, Program Commission of Health, Food and Welfare, and the strategic research initiative BEST funded by the University of Copenhagen for financial support.

JF802057V

Paper IV

Åsmund Rinnan, Niels Johan Christensen, and Søren Balling Engelsen. *How the energy evaluation method used in the geometry optimization step affects the quality of the subsequent QSAR/QSPR models*. Journal of Computer-Aided Molecular Design, **Accepted.**

How the energy evaluation method used in the geometry optimization step affect the quality of the subsequent QSAR/ QSPR models

Åsmund Rinnan*, Niels Johan Christensen, Søren Balling Engelsen

Quality and Technology, Department of Food Science, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, 1958 Frederiksberg C, Denmark

* To receive all correspondence; E-mail: aar@life.ku.dk, Phone: +45 35 33 35 42, Fax: +45 35 33 32 45

Keywords: QSAR, QSPR, energy evaluation, PLS regression, quantum mechanics, semi-empirical, molecular mechanics

Abbreviations: QSPR, Quantitative Structure Property Relationship; QSAR, Quantitative Structure Activity Relationship; MM3* , Allinger's Molecular Mechanics; AM1, Austin Model 1; PM3, Parameterized Model 3, HF, Hartree-Fock; B3LYP, the hybrid exchange-correlation functional based on work from Becke, Lee, Yang and Par; PLS, Partial Least Squares; RMSD, Root Mean Square Distance; MCMM, Monte Carlo Multiple Minimum

Abstract

The quantitative influence of the choice of energy evaluation method used in the geometry optimization step prior to the calculation of molecular descriptors in QSAR and QSPR models was investigated. A total of 11 energy evaluation methods on three molecular datasets (toxicological compounds, aromatic compounds and PPARγ agonists) were studied. The methods employed were: MMFF94s, MM3* with ε_r (relative dielectric constant) = 1, MM3* with ε_r = 80, AM1, PM3, HF/STO-3G, HF/6-31G, HF/6-31G(d,p), B3LYP/STO-3G, B3LYP/6-31G, and B3LYP/6-31G(d,p). The 3D-descriptors used in the QSAR/QSPR models were calculated with commercially available molecular descriptor programs primarily directed toward pharmaceutical research. In order to evaluate the uncertainties involved in the QSAR/QSPR predictions bootstrapping was used to validate all models using 1000 drawings for each data set. The scale free error-term, q^2, was used to compare the relative quality of the models resulting from different optimization methods on the same set of molecules. Depending on the dataset, the average 0.632 bootstrap estimated q^2 varies from 0.55 to 0.57 for the toxicological compounds, from 0.58 to 0.62 for the aromatic compounds, and from 0.69 to 0.75 for the PPARγ agonists. The B3LYP/6-31G(d,p) provided the best overall results, albeit the increase in q^2 was small in all cases. The results clearly indicate that the choice of the energy evaluation method has very limited impact. This study suggests that QSAR or QSPR studies might benefit from the choice of a rapid optimization method with little or no loss in model accuracy.

1 Introduction

QSAR and QSPR methods are becoming more and more attractive for screening purposes in pharmaceutical [1], toxicological [2], nutritional sciences [3] and health sciences [4]. In QSPR and QSAR a wide variety of computational methods for calculating the potential energy surface of the molecules, and thus affecting the geometry optimization of the molecular structures, may be chosen prior to the calculation of molecular descriptors and the subsequent regression step. The computational time increases manifold when going from the more pragmatic to the more sophisticated energy evaluation methods. For typical small drug-sized molecules, molecular mechanics methods can generate an optimized structure in seconds while quantum mechanical methods may require hours or days on present-day computers. Considering that QSPR and QSAR applications often involve datasets with tens, hundreds or even thousands of molecules, the choice of energy evaluation method becomes a real concern when contrasted with the demands for rapid development. While accurate results in computational chemistry often necessitate calculations at a high level of theory, it has also been demonstrated that the geometry obtained using energy evaluations at the highest level of theory does not always lead to the best results in subsequent calculations of molecular properties [5]. This suggests that any presumption about the relationship of the energy evaluation method and the quality of a particular type of results should be carefully examined. Nevertheless, only few attempts have been made to elucidate how the choice of energy evaluation method affects QSAR and QSPR models [6-8]. The most commonly employed energy evaluation methods in QSAR/QSPR studies are AM1 [9], PM3 [10], and HF [11], and B3LYP [12-

15] with the 6-31G(d) basis set [16-20]. However, rarely is any rationale given for using a particular method and almost never are several methods compared within the same study. In a surprisingly high percentage (about 25%) of the published works in literature, any description of the energy evaluation method used during geometry optimization is absent. Such widespread omission of information suggests that the matter in question is either firmly established or irrelevant. In this study the influence of the choice of energy evaluation method used in the geometry optimization step upon the final QSAR/QSPR prediction model is investigated for three diverse datasets selected from the literature. A series of QSAR/QSPR models for the prediction of the measured quantity were built using structures from each of the energy evaluation schemes and the relative quality of these models was compared using q^2.

The authors have chosen to use the phrase "energy evaluation" regarding the methods investigated in this manuscript as we do not investigate the effect of energy minimization algorithms such as conjugate gradients [21], BFGS, DFP [22] or quasi-Newton [23] in conjunction with an energy evaluation method such as B3LYP [12-15], HF [11] or PM3 [10]. The conjunction between the two could be referred to as geometry optimization.

2 Methods

Figure 1 shows how the work of this paper was designed. The grey circle with the molecule depicts the geometry optimization step wherein the energy evaluation method used was changed. These optimizations are then used for two purposes: to calculate the difference in the geometries (RMSD) and to make prediction models. The black squares to the right of the molecules are the data-matrices used for the subsequent prediction models. A total of 1000 bootstrap drawings were performed for each energy evaluation

method and the samples left out by the bootstrap drawing were used as the validation set. The q2-values of these validation sets where subsequently used in order to compare the importance of the choice of the energy evaluation method. This whole process was repeated for each of the energy evaluation methods, i.e. 11 times in total. The number and names of both the compounds and the descriptors were kept constant for all repetitions in order to be able to compare the effect of the energy evaluation methods on the subsequent regression model. Further description of each of the steps follows below.

2.1 Datasets

This study investigates three datasets already published and discussed in earlier papers, all showing good predictive power. Although the datasets used in this study are somewhat reduced from these original studies, the subsets of molecules should still lead to acceptable models.

The three sample sets are diverse: 290 compounds exhibiting acute aquatic toxicity in fathead minnow [24] – named toxicological compounds, 80 compounds with various degrees of penetration of a polydimethylsiloxane membrane [25] – named aromatic compounds, and 12 PPARγ agonists [26] with different pK_i-values for PPARγ. Further details about these datasets are given in Supplementary Material – Part I.

The last dataset contains 12 molecules, but this low number is justified by the goal of this project: investigating the perturbation effect the energy evaluation method has on the subsequent regression analysis,

2.2 Geometry optimization

All structures were built with GaussView [27] and subjected to 5000 steps of MCMM conformational searching in MacroModel [28] using the MMFF94s force-field [29] and the PRCG optimization algorithm [30]. Only the lowest conformer was selected for further use in the study. The conformational search was included to remove human bias from the initially generated structures.

The lowest conformer of each molecule was geometry optimized further using 10 energy evaluation methods: MM3* [31] with $\varepsilon_r = 1$, MM3* with $\varepsilon_r = 80$, AM1, PM3, HF with STO-3G [32], HF/6-31G, HF/6-31G(d,p), B3LYP/STO-3G, B3LYP/6-31G, and B3LYP/6-31G(d,p). For the MM3* method the PRCG optimization algorithm were used, while the Berny optimization algorithm [33] was used for the remaining methods. It is assumed that the choice of the optimization algorithm is insignificant compared to the choice of energy evaluation method. Thus a total of 11 energy evaluation methods were evaluated. In the remainder of the text we use the shorthand notations MM3*-1 and MM3*-80 for MM3* with $\varepsilon_r = 1$ and MM3* with $\varepsilon_r = 80$, respectively.

2.3 Prediction Models

VAMP [34] was employed to calculate the total energy, electronic energy, nuclear energy, surface area, mean polarizability, heat of formation, ionization potential, LUMO/HOMO energies, total dipole, and partial charges for all molecules at the AM1 level. DRAGON [35] provided additional 3D descriptors in the categories: Randic molecular profiles, geometrical descriptors, RDF descriptors, 3D-MoRSE (3D-Molecule

Representation of Structures based on Electron diffraction) descriptors [36], WHIM (Weighted Holistic Invariant Molecular descriptors) descriptors [37], GETAWAY (GEometry, Topology, and Atom-Weights AssemblY) descriptors [38], and charge descriptors (using VAMP charges). Thus all the descriptors used in this study are geometry sensitive descriptors, capturing variations due to the choice of energy evaluation method. Descriptors with no variance in at least one of the molecules of the data set across the optimization methods studied were excluded, in order to focus the study on the geometry sensitive descriptors (the toxicological compounds thus included 293 variables, the aromatic compounds 464 variables and PPARγ agonists included 661 variables, as also can be seen from Figure 1). This was furthermore done because the prediction models including all 681 descriptors are in general worse and less consistent. Furthermore, our focus is not on how to make the best PLS-model, but rather on the differences between comparable PLS-models.Auto-scaling was used in order to give all descriptors the same chance of influencing the model. The prediction models were all PLS-models [39]. In order to estimate the uncertainty of the models, 1000 bootstrap [40] drawings were performed for each dataset (the number of samples in each bootstrap drawing equals the number of samples in the original dataset under investigation). The same set of bootstrap drawings were performed on all energy evaluation methods, ensuring that the selection of which samples goes in the calibration or the validation set does not influence the quality of the final model. The resulting RMSE-values were used to calculate the 0.632 bootstrap estimates of the mean error and of the standard error of the error. The number 0.632 is approximately the probability that any one sample is in a bootstrap drawing [40]. Bootstrapping was performed instead of simple cross-validation,

as bootstrapping also provides an estimate of the uncertainty of the prediction error [40, 41], which leads to options to test for significance between two sets of datasets (two different energy evaluation methods in this particular case). The 0.632 bootstrap RMSE estimates were transformed into q^2-values by the formula shown in Equation 1.

Here y is the reference value, $\hat{y}$ is the predicted value, $\bar{y}$ is the mean reference value, s_y is the standard deviation of the reference values and n is the number of samples in the dataset. All the numbers in the equation refers to the samples in the validation set, i.e. different for each bootstrap drawing, similar to formula 2 in the work by Consonni *et al.* [42].

The difference in the 0.632 bootstrap estimate of the q^2 was tested using a Games-Howell comparison of mean test [43]. This test was performed on the basis of the 0.632 bootstrap estimates of the q^2 values and the uncertainty of these, i.e. using the results from the 1000 bootstrap drawings. The p-values were corrected according the Bonferroni correction, stating that $p_{corr} = 1 - p^{1/k}$, where p is the normal significance value (e.g. 0.05), k is the number of comparisons and p_{corr} is the corrected significance value. For a comparison of n different groups $k = \frac{n \cdot (n-1)}{2}$.

2.4 Variations in geometry

To quantify geometrical differences, the optimized structures of all methods were superimposed on the structure obtained at the highest level of theory (B3LYP/6-31G(d,p)) using all heavy atoms. The RMS distance (RMSD) was calculated for the superposition between equivalent atoms excluding hydrogens, as several of the hydrogens

have rotational freedom, for example in a methyl group. The RMSD is defined as shown in Equation 2.

Here $r_{i,1}$ and $r_{i,2}$ are the positions of atom i in structure 1 and 2, respectively and N is the total number of atoms in the molecule. Furthermore, the number of close contacts were counted; defined as two atoms joined through at least three covalent bonds and separated by less than 75% of the sum of their van der Waals radii. Hydrogen bonds were counted using the following criteria: The H—A-R angle had to be larger than 90 degrees, the H—A distance had to be smaller than 2.5 Å, and the D-H—A angle had to be larger than 120 degrees (H = Hydrogen, A = Acceptor, D = Donor, R = Remainder of molecule).

2.5 Software

Molecular structures were created with GaussView [27]. The MM3* optimizations were carried out with MacroModel [28]. All other optimizations were performed with Gaussian 03 [44]. RMSD values were calculated with Maestro [28]. The prediction models were made in Matlab 7.6 [45] with in-house functions.

3 Results and Discussion

All molecules in this study were geometry optimized in the isolated state neglecting intermolecular interactions which may induce geometry changes important for QSAR/QSPR models.

3.1 Influence of energy evaluation method on geometry

For the toxicological compounds, 42 (about 1%) of the optimized structures had close contacts. These structures were all produced with PM3, HF/STO-3G, or B3LYP/STO-

3G. PM3 and B3LYP/STO-3G gave several bad contacts for the same compound. Ten compounds in this dataset allow for internal hydrogen bonds. All methods reproduced the hydrogen bonds, except in case of MM3*-80 for four structures. The reluctant capability of creating hydrogen bonds in the latter case is a natural consequence of large separation between hydrogen bond donors and acceptors for the five compounds and high relative dielectric constant used for MM3*.

For the aromatic compounds, a close contact was found in 13 structures (11 different compounds) optimized with B3LYP/STO-3G, HF/STO-3G or PM3. All optimization methods reproduced intramolecular hydrogen bonds in the four structures where they were expected. B3LYP/STO-3G and HF/STO-3G, and MM3*-1 produced a false hydrogen bond in one structure.

In the PPARγ agonists a single close contact was found for 13 optimized structures, five structures had two close contacts, and 3 close contacts were found for one compound. The close contacts were in all cases produced with PM3 or B3LYP/STO-3G. One structure is expected to have two hydrogen bonds which were reproduced with most methods, except AM1 and the force field methods, which produce only one hydrogen bond.

A simple comparison of the three datasets indicates that the toxicological compounds and the aromatic compounds generally have low (max median 0.06 Å) and comparable RMSD values, see Figure 2. The RMSDs (max median 1.3 Å) for the PPARγ agonists are larger reflecting the increased flexibility of the larger molecules. A more detailed examination reveals that the median RMSD values for the toxicological compounds and the aromatic compounds follow the same behavior for DFT and Hartree-Fock methods,

although median RMSD values are slightly but consistently lower for the toxicological compounds. For both datasets the lowest median RMSDs are produced with HF/6-31G(d,p), but the difference between this method and B3LYP/6-31G or HF/STO-3G is marginal. B3LYP/STO-3G gives relatively large RMSDs, as evident from medians and 95-percentile values. Visual inspection of the optimized geometries reveals that the energy evaluation methods do not in general introduce any prominent structural distortions when applied to the toxicological compounds or the aromatic compounds. Comparison of the 95-percentiles for these two datasets (Figure 2a and 2b) shows that the evaluation methods give roughly the same pattern of relative deviations from B3LYP/6-31G(d,p). The significant flexibility of the molecules of the PPARγ agonists is reflected in the generally high values of RMSD medians (max 1.26 Å) and other statistics. B3LYP/6-31G gives the lowest median, but not the smallest spread in RMSD.

3.2 Prediction results

The datasets in this study give mediocre models for the toxicological compounds and aromatic compounds ($q^2 \in [0.55, 0.57]$ and $q^2 \in [0.58, 0.62]$ respectively), and satisfactory models for the PPARγ agonists ($q^2 \in [0.69, 0.75]$). The variation in the q^2-values is largest for the PPARγ agonists, and smallest for the toxicological compounds, which is consistent with the RMSD values calculated above, indicating that the differences in the prediction models are due to differences in 3D structure of the molecules. Figure 3 shows the % deviation in the mean q^2 of the 11 energy evaluation methods compared to the best method (see Equation 3), including the standard error of the q^2. The smallest average deviation is found for the most expensive method –

B3LYP/6-31G(d,p). Furthermore, in two of the three datasets this method either has the highest q^2 or there is no significant difference ($p > 0.05$) between B3LYP/6-31G(d,p) and the best model (see Tables 4-6 in Supplementary Material – Part II). For the last dataset (the aromatic compounds) it is only significantly worse than B3LYP/6-31G (best model) and B3LYP/STO-3G ($p < 0.001$), but only by a decrease in q^2 by 0.03 and 0.02 respectively. However, more pragmatic methods such as B3LYP/6-31G, B3LYP/STO-3G, HF/6-31G and even PM3 show similar results. From these observations it becomes evident that the choice of basis set is not so important for the B3LYP method. Taking into account the discussion of the close contacts above, it becomes clear that B3LYP with 6-31G or 6-31G(d,p) are the best choices. From a time perspective B3LYP/6-31G(d,p) is approximately 3, 10 and 100 times more expensive than B3LYP/6-31G, HF/6-31G and PM3 respectively. This study indicates that if a larger QSAR/QSPR screening is to be performed, cheaper methods such as HF/6-31G or PM3 may be employed with success. This is in accordance to the conclusion in the paper by Puzyn et al. [46], where they state that it is better to use the semi-empirical methods PM6 (a new version of PM3) [47] or RM1 (a new version of AM1) [48] instead of the more expensive DFT methods.

4 Conclusions

The influence of the choice of energy evaluation method in the geometry optimization step on the predictive quality of QSAR/QSPR models for three different molecular datasets has been investigated. The lowest energy conformer in each dataset was optimized with 11 different methods and subsequently 3D molecular descriptors were calculated with VAMP and DRAGON. The results show that the energy evaluation methods only to a small extent influence the QSAR/QSPR prediction model. The most

time consuming method – B3LYP/6-31G(d,p) – is the method which in general gives the best prediction models, albeit the increase in q^2 is rather small. This further suggests that the usage of more pragmatic methods such as HF/6-31G and PM3 can be used, especially in larger screening analyses with little or no loss in model accuracy.

Acknowledgements

The authors would like to thank the KU-LIFE strategic research project BEST for the financial funding of Åsmund Rinnan and the project Build Your Food ((FFS05-9) sponsored by the Ministry of Agriculture and Fisheries for the financial funding for Niels Johan Christensen.

References

[1] Agüero-Chapin G, Varona-Santos J, de la Riva G A, Antunes A, Gonzáles-Villa T, Uriarte E, González-Díaz H (2009), J Proteome Res, 8: 2122-2128

[2] Pasha F A, Nez M M, Cho S J, Ansari M, Mishra S K, Tiwari S (2009) Chem Biol Drug Des, 73: 537-544

[3] Jensen B F, Sørensen M D, Kissmeyer A-M, Björkling F, Sonne K, Engelsen S B, Nørgaard L (2003) J Comput Aid Mol Des, 17: 849-859

[4] Thorsteinson N, Ban F, Santos-Filho O, Tabaei S M H, Miguel-Queralt S, Underhill C, Cherkasov A, Hammond G L (2009) Toxicol Appl Pharm, 234: 47-57

[5] Foresman J B, Frisch A (1996) Exploring Chemistry with Electronic Structure Methods, Second Edition, Gaussian, Pittsburgh, 157-158

[6] Hudáky I, Hudáky I, Perczel A (2004) J Comput Chem, 25: 1522-1531

[7] Swart M, Snijders J G (2003) Theor Chem Acc, 110: 34-41

[8] Bayari S, Saglam S, Ustundag H F (2005) J Mol Struct, 726: 225-232

[9] Dewar M J S, Zoebisch E G, Healy E F, Stewart J J P (1985) J Am Chem Soc, 107: 3902-3909

[10] Stewart J J P (1989) J Comput Chem, 10: 209-220

[11] Roothaan C C J (1951) Rev Mod Phys, 23: 69-89

[12] Becke A D (1993) J Chem Phys, 98: 5648-5652

[13] Lee C, Yang W, Parr R G (1988) Phys Rev B, 37: 785-789

[14] Vosko S H, Wilk L, Nusair M (1980) Can J Phys, 58: 1200-1211

[15] Stephens P J, Devlin F J, Chabalowski C F, Frisch M J (1994) J Phys Chem, 98, 11623-11627

[16] Ditchfield R, Hehre W J, Pople J A (1971) J Chem Phys, 54: 724-728

[17] Hehre W J, Ditchfield R, Pople J A (1972) J Chem Phys, 56: 2257-2261

[18] Hariharan P C, Pople J A (1974) Mol Phys, 27: 209-214

[19] Gordon M S (1980) Chem Phys Lett, 76: 163-168

[20] Hariharan P C, Pople J A (1973) Theor Chim Acta, 28: 213-222

[21] Fletcher R, Reeves C M (1964) Computer J, 7: 149-154

[22] Stewart J P P (1989) Comput Chem, 13: 157-158

[23] Dennis J E, More J J (1977) SIAM Rev, 19: 46-89

[24] He L, Jurs P C (2005) J Mol Graphics Modell, 23: 503-523

[25] Agantovic-Kustrin S, Beresford R, Pauzi A, Yusof M (2001) J Pharm Biomed Anal, 26: 241-254

[26] Rücker C, Scarsi M, Meringer M (2006) Bioorg Med Chem, 14: 5178-5195

[27] GaussView, Version 3.09, Dennington II R, Keith T, Milliam J, Eppinnett K, Hovell W L, Gilliland R (2003) Semichem, Inc., Shawnee Mission, KS, http://www.gaussian.com

[28] Schrödinger, LLC. (2007) http://www.schrodinger.com

[29] Halgren T A (1999) J Comput Chem, 20: 720-729

[30] Polak E, Ribiere G (1969) Revenue Francaise Inform Rech Operationelle, 16: 35

[31] Allinger N L, Yuh Y H, Lii J-H (1989) J Am Chem Soc, 111: 8551-8566

[32] Hehre W J, Stewart R F, Pople J A (1969) J Chem Phys, 51: 2657-2664

[33] Peng C Y, Ayala P Y, Schlegel H B, Frisch M J (1996) J Comp Chem, 17: 49-56

[34] Vamp 6.5 (1997) Oxford Software Limited

[35] Talete srl (2009) DRAGON for Windows (Software for Molecular Descriptor Calculations) http://www.talete.mi.it

[36] Schuur J H, Selzer P, Gasteiger J (1996) J Am Chem Soc, 36: 334-344

[37] Todeschini R, Lasagni M (1994) J Chemom, 8: 263-272

[38] Consonni V, Todeschini R, Pavan M, Gramatica P (2002) J Chem Inf Comput Sci, 42: 693-705

[39] Martens H, Martens M, Næs T (2001) Multivariate Analysis of Quality, John Wiley & Sons, Chichester, 111-125

[40] Wehrens R, Putter H, Buydens L M C (2000) Chemom Intell Lab Syst, 54: 35-52

[41] Efron B, Gong G (1983) Am Stat, 37: 36-48

[42] Consonni V, Ballabio D, Todeschini R (2009) J Chem Inf Model, 49: 1669-1678

[43] Howell JF, Games PA (1974) Brit J Math Stat Psy, 27: 72-81

[44] Gaussian 03, Revision C.02, Frisch M J, Trucks G W, Schlegel H B, Scuseria G E, Robb M A, Cheeseman J R, Montgomery J A Jr, Vreven T, Kudin K N, Burant J C, Millam J M, Iyengar S S, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson G A, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox J E, Hratchian H P, Cross J B, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann R E, Yazyev O, Austin A J, Cammi R, Pomelli C, Ochterski J W, Ayala P Y, Morokuma K, Voth G A, Salvador P, Dannenberg J J, Zakrzewski V G, Dapprich S, Daniels A D, Strain M C, Farkas O, Malick D K, Rabuck A D, Raghavachari K, Foresman J B, Ortiz J V, Cui Q, Baboul A G, Clifford S, Cioslowski J, Stefanov B B, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin R L, Fox D J, Keith T, Al-Laham M A, Peng C Y, Nanayakkara A, Challacombe M, Gill P M W, Johnson B, Chen W, Wong M W, Gonzalez C, Pople J A (2009) Gaussian Inc., Wallingford C

[45] The Mathworks ™ (2009) USA, http://www.mathworks.com

[46] Puzyn T, Suzuki N, Haranczyk M, Rak J (2008) J Chem Inf Model, 48: 1174-1180

[47] Rocha G B, Freire R O, Simas A M, Stewart J J P (2006) J Comput Chem, 27: 1101-1111

[48] Stewart J J P (2007) J Mol Model, 13: 1173-1213

5

(1) $$q^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2} = 1 - \left(\frac{RMSE}{s_y} \right)^2 \cdot \frac{n-1}{n}$$

(2) $$\mathrm{RMSD} = \sqrt{\sum_{i=1}^{N} \frac{(r_{i,1} - r_{i,2})^2}{N}}$$

(3) $$deviation = \frac{q^2_{best} - q^2_{current}}{q^2_{best}} \cdot 100\%$$

Fig. 1 Design of the analysis

Fig. 2 RMSD values for the three QSAR/QSPR datasets. (a) Toxicological compounds, (b) aromatic compounds, and (c) PPARγ agonists. Minimum and maximum values are indicated by whiskers. The boxes are vertically limited by the 5 and 95 percentile in (a) and (b) and by the second lowest and second highest value in (c). Median values are indicated with black bars in the interior of the boxes

Fig. 3 Deviation from the maximum 0.632 bootstrap estimate of the mean q^2 given in %, with the standard error showed by whiskers.

Figure 1

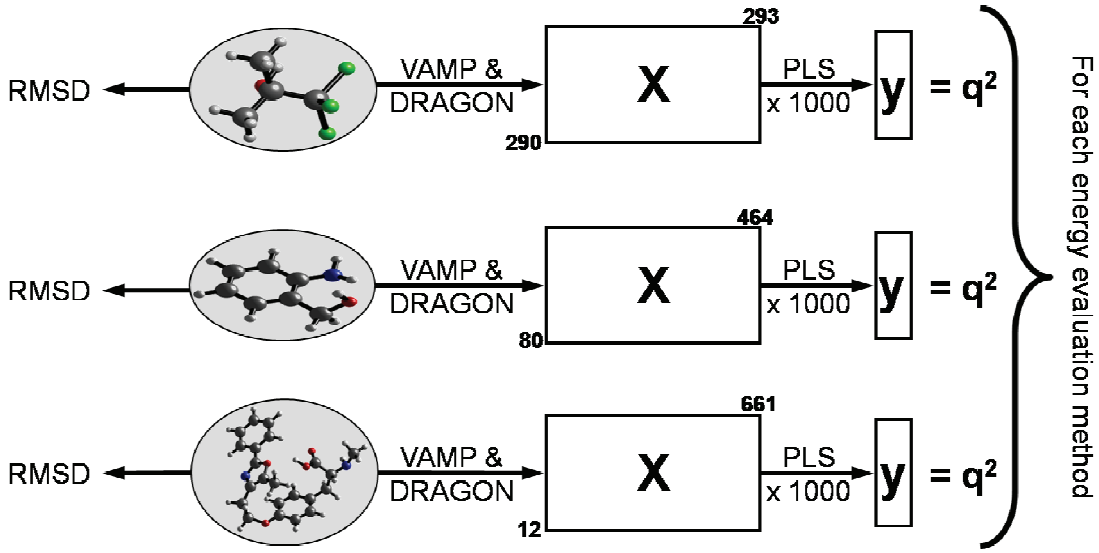

RMSD
VAMP &
DRAGON
293
X
290
PLS
x 1000
y
= q²
RMSD
VAMP &
DRAGON
464
X
80
PLS
x 1000
y
= q²
RMSD
VAMP &
DRAGON
661
X
12
PLS
x 1000
y
= q²
For each energy evaluation method

Figure 2

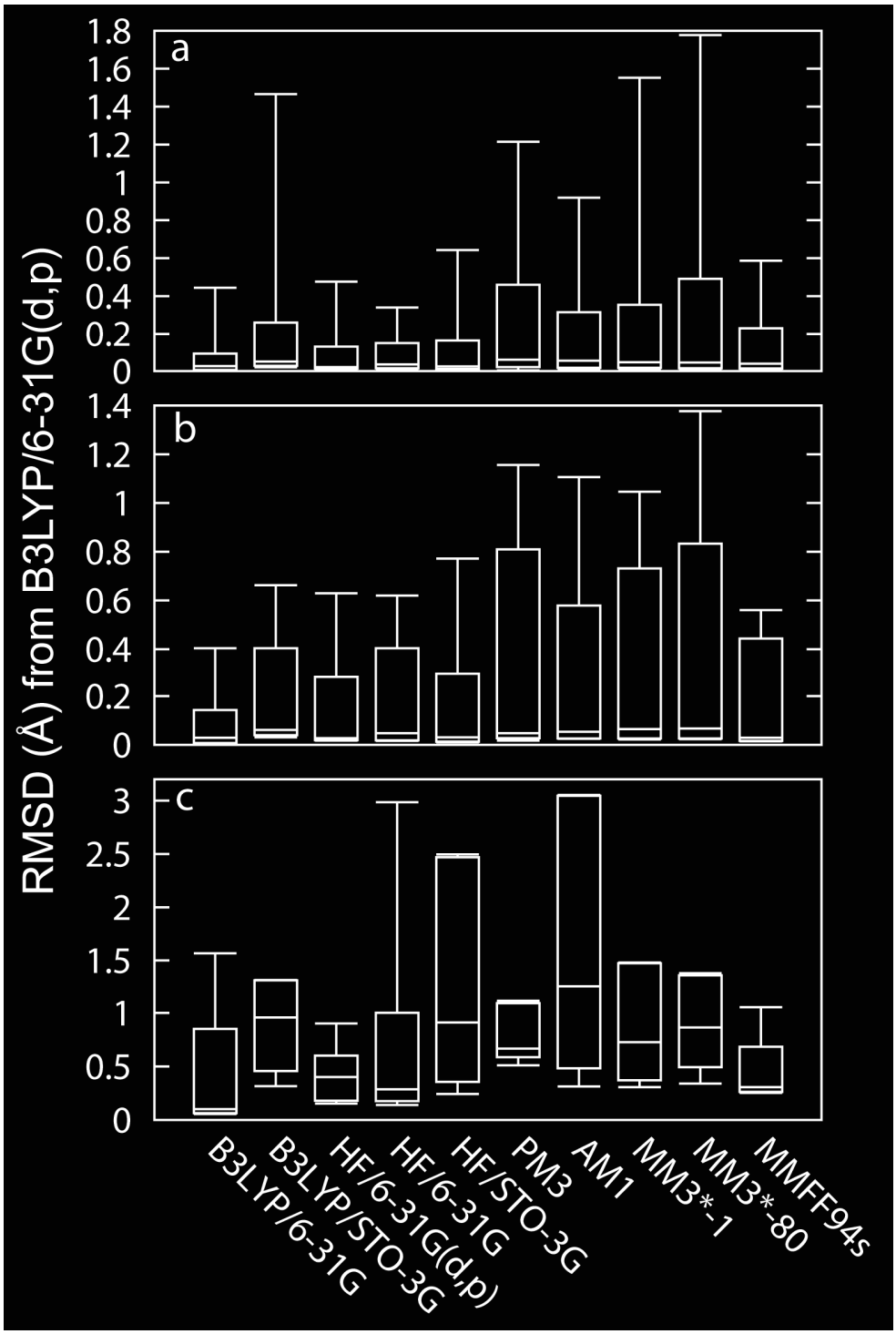
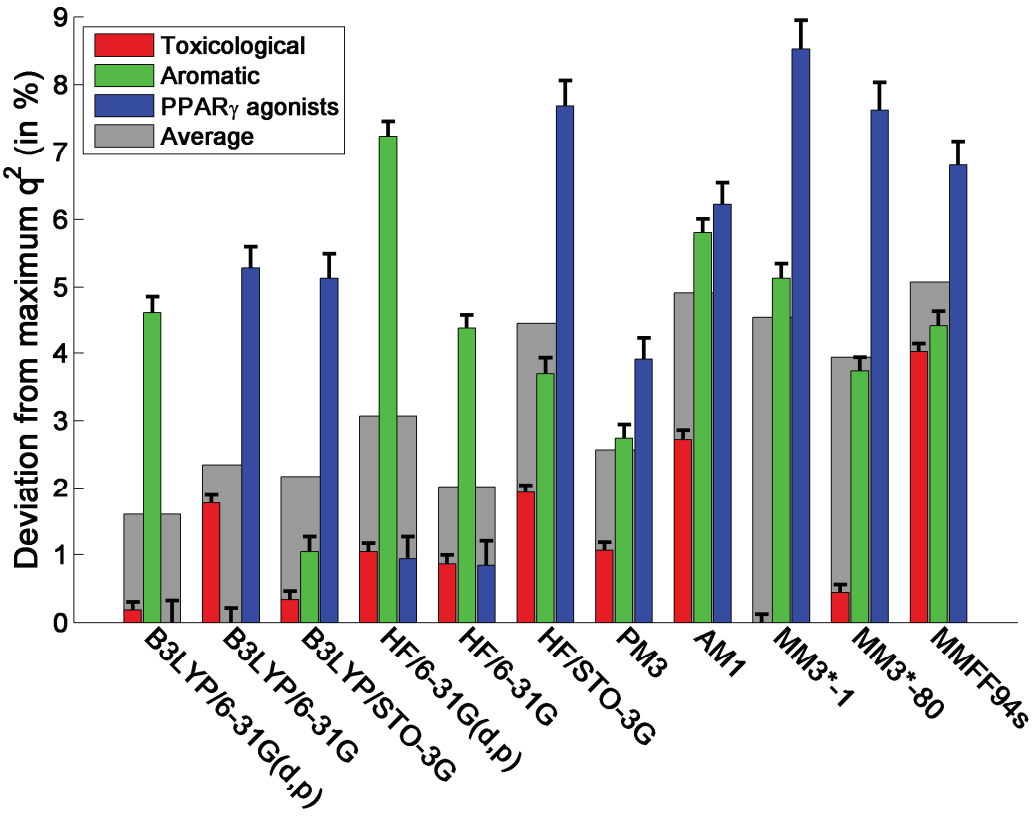

RMSD (Å) from B3LYP/6-31G(d,p)
a
b
c
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
3
2.5
2
1.5
0.5
B3LYP/6-31G
B3LYP/STO-3G
HF/6-31G(d,p)
HF/6-31G
HF/STO-3G
PM3
AM1
MM3*-1
MM3*-80
MMFF94s

Figure 3


Toxicological
Aromatic
PPARγ agonists
Average
Deviation from maximum q2 (in %)
0
1
2
3
4
5
6
7
8
9
B3LYP/6-31G(d,p)
B3LYP/6-31G
B3LYP/STO-3G
HF/6-31G(d,p)
HF/6-31G
HF/STO-3G
PM3
AM1
MM3*-1
MM3*-80
MMFF94s

Supplementary Material – Part I

All the datasets are taken from the literature. However, they have somewhat been reduced in size from the study originally published by removing some compounds. The following information gives lists of the compounds and property/ activity used in the prediction models. Furthermore, the average predicted value of each compound and energy evaluation method is given, together with its standard error, both calculated on the basis of the validation using the 1000 bootstrap drawings. The "#" column is the number of times the compound has been in the validation set, and thus the number of predictions the mean and the standard error is based upon.

Toxicological compounds

These data are all taken from He and Jurs' [21] article. The number of compounds has been reduced from the original 322 to 290 as we had computational difficulties with some of the compounds in one or several of the energy evaluation methods, and some other compounds behaved as outliers in one or several of the PLS-models. The property used in the prediction is the acute aquatic toxicity in fathead minnow.

Table A1: Compound names and their acute aquatic toxicity in fathead minnow [21].

Name	Exp	#	B3LYP	HF	PM3	AM1	MM3*	MMFF94s

			6-31G(d,p)	6-31G	STO-3G	6-31G(d,p)	6-31G	STO-3G			ε = 1	ε = 80	
1,1,1-trichloro-2-methyl-2-propanol	3.12	363	3.75 (0.02)	3.41 (0.01)	3.04 (0.02)	3.67 (0.02)	3.64 (0.02)	3.15 (0.02)	2.95 (0.01)	3.10 (0.02)	3.04 (0.02)	3.18 (0.02)	3.01 (0.02)
1,1,1-trichloroethane	3.4	342	2.77 (0.01)	2.76 (0.01)	2.96 (0.02)	2.60 (0.01)	2.87 (0.02)	3.13 (0.02)	3.19 (0.01)	3.16 (0.02)	3.00 (0.02)	2.97 (0.02)	2.75 (0.02)
1,1,2,2-tetrachloroethane	3.92	354	4.16 (0.01)	4.42 (0.01)	4.27 (0.01)	4.08 (0.02)	4.32 (0.01)	4.23 (0.01)	4.06 (0.01)	4.54 (0.02)	4.06 (0.01)	4.04 (0.01)	3.67 (0.02)
1,1,2-trichloroethane	3.21	375	3.63 (0.01)	3.60 (0.01)	3.54 (0.01)	3.53 (0.01)	3.49 (0.01)	3.81 (0.01)	3.27 (0.01)	3.46 (0.01)	3.65 (0.01)	3.63 (0.01)	3.59 (0.01)
1,1-dichloroethylene	2.84	384	2.24 (0.01)	2.26 (0.01)	2.34 (0.01)	2.31 (0.02)	2.26 (0.01)	2.39 (0.01)	2.66 (0.02)	2.54 (0.01)	2.78 (0.02)	2.88 (0.02)	3.35 (0.02)
1,2,3,4-tetrachlorobenzene	5.29	374	5.59 (0.01)	5.55 (0.01)	5.58 (0.01)	5.52 (0.01)	5.43 (0.01)	5.59 (0.01)	5.53 (0.01)	5.38 (0.01)	5.69 (0.01)	5.64 (0.01)	5.42 (0.01)
1,2,3-trichlorobenzene	4.89	343	5.29 (0.01)	5.01 (0.01)	5.15 (0.01)	5.17 (0.01)	5.02 (0.01)	5.22 (0.01)	5.32 (0.01)	5.26 (0.01)	5.27 (0.01)	5.24 (0.01)	5.15 (0.01)
1,2,3-trichloropropane	3.41	353	3.63 (0.01)	3.62 (0.01)	3.29 (0.01)	3.57 (0.02)	3.61 (0.01)	3.54 (0.01)	3.80 (0.02)	3.57 (0.01)	3.75 (0.02)	3.85 (0.02)	3.92 (0.01)
1,2,4,5-tetrachlorobenzene	5.83	359	5.50 (0.01)	5.40 (0.01)	5.75 (0.01)	5.38 (0.01)	5.77 (0.01)	5.66 (0.01)	5.72 (0.01)	5.67 (0.01)	5.43 (0.01)	5.49 (0.01)	5.22 (0.01)
1,2,4-trichlorobenzene	4.8	374	4.98 (0.01)	5.01 (0.01)	5.09 (0.01)	5.03 (0.01)	5.15 (0.01)	5.06 (0.01)	5.11 (0.01)	5.03 (0.01)	4.97 (0.01)	4.99 (0.01)	4.90 (0.01)
1,2,4-trimethylbenzene	4.19	344	4.35 (0.01)	4.35 (0.01)	4.36 (0.01)	4.05 (0.01)	4.04 (0.01)	4.29 (0.01)	4.38 (0.01)	4.12 (0.01)	4.21 (0.01)	4.20 (0.01)	4.49 (0.01)
1,2-dichloroethane	2.9	385	3.20 (0.01)	3.48 (0.01)	3.33 (0.01)	3.17 (0.01)	3.13 (0.01)	3.37 (0.01)	2.93 (0.01)	3.35 (0.01)	3.26 (0.01)	3.26 (0.01)	3.14 (0.01)
1,2-dimethylpropylamine	2.49	345	2.57 (0.01)	2.36 (0.01)	2.47 (0.01)	2.59 (0.01)	2.39 (0.01)	2.63 (0.01)	2.26 (0.01)	2.43 (0.01)	2.57 (0.01)	2.55 (0.01)	2.79 (0.01)
1-(2-hydroxyethyl)piperazine	1.31	362	1.12 (0.02)	1.24 (0.02)	2.15 (0.03)	1.11 (0.02)	0.70 (0.02)	1.54 (0.02)	1.77 (0.01)	1.23 (0.02)	1.55 (0.02)	1.65 (0.02)	1.49 (0.02)
1,2-propanediamine	1.81	388	2.31 (0.01)	2.41 (0.01)	2.02 (0.01)	2.17 (0.01)	2.05 (0.01)	2.19 (0.01)	2.15 (0.01)	2.01 (0.01)	2.22 (0.01)	2.15 (0.01)	2.18 (0.01)
1,2-propyleneglycol	0.13	368	1.27 (0.01)	1.40 (0.01)	1.27 (0.01)	1.45 (0.01)	1.24 (0.01)	1.41 (0.01)	1.63 (0.01)	1.34 (0.01)	1.30 (0.01)	1.38 (0.01)	1.48 (0.01)
1,3,5-trichlorobenzene	4.74	387	4.79 (0.01)	4.93 (0.01)	4.73 (0.01)	5.17 (0.01)	4.87 (0.01)	4.98 (0.01)	5.08 (0.01)	4.93 (0.01)	4.95 (0.01)	4.96 (0.01)	4.96 (0.01)
1,3-dibromopropane	5.05	385	4.00 (0.02)	4.31 (0.02)	4.52 (0.02)	3.84 (0.02)	4.01 (0.02)	4.00 (0.02)	5.32 (0.02)	4.16 (0.02)	4.14 (0.02)	4.48 (0.02)	4.49 (0.03)
1,3-dichloropropane	2.94	345	3.15 (0.01)	3.01 (0.01)	2.96 (0.01)	3.33 (0.01)	3.36 (0.01)	3.02 (0.01)	3.08 (0.01)	2.87 (0.01)	2.71 (0.01)	2.72 (0.01)	3.09 (0.01)
1,4-bis(chloromethyl)benzene	6.65	388	5.15 (0.01)	5.22 (0.02)	5.14 (0.02)	5.02 (0.01)	5.06 (0.01)	5.26 (0.01)	4.34 (0.02)	5.11 (0.02)	4.72 (0.01)	4.74 (0.01)	4.59 (0.02)
1,4-dibromobenzene	5.28	394	4.46 (0.02)	4.35 (0.03)	4.99 (0.02)	4.77 (0.02)	4.67 (0.03)	4.60 (0.02)	5.17 (0.02)	4.86 (0.02)	4.69 (0.02)	4.38 (0.03)	4.87 (0.02)
1,4-dichlorobutane	3.39	360	3.64 (0.01)	3.45 (0.01)	3.37 (0.01)	3.59 (0.01)	3.60 (0.02)	2.92 (0.02)	3.52 (0.01)	2.94 (0.01)	3.63 (0.02)	3.23 (0.01)	3.12 (0.01)
1,4-dimethoxybenzene	3.07	358	3.47 (0.01)	3.56 (0.01)	3.48 (0.01)	3.29 (0.01)	3.57 (0.01)	3.32 (0.01)	3.51 (0.01)	3.42 (0.01)	3.41 (0.01)	3.46 (0.01)	3.44 (0.01)
1,4-dioxane	0.93	378	0.73 (0.02)	0.83 (0.02)	0.75 (0.02)	0.54 (0.02)	0.72 (0.02)	1.23 (0.02)	0.57 (0.02)	0.96 (0.02)	1.36 (0.02)	1.10 (0.02)	1.01 (0.02)
1,5-dichloropentane	3.75	357	3.83 (0.02)	3.75 (0.01)	3.40 (0.02)	3.77 (0.02)	3.73 (0.02)	3.52 (0.02)	3.68 (0.02)	3.76 (0.01)	3.77 (0.01)	3.86 (0.01)	3.71 (0.02)
1,8-epoxy-p-menthane	3.18	386	3.23 (0.02)	2.73 (0.02)	3.23 (0.02)	3.61 (0.02)	3.48 (0.02)	3.34 (0.02)	4.00 (0.02)	3.56 (0.03)	4.08 (0.02)	4.16 (0.02)	3.49 (0.02)
1,9-decadiene	5.68	386	5.78 (0.01)	5.54 (0.01)	5.79 (0.01)	5.73 (0.01)	5.69 (0.01)	5.87 (0.02)	6.43 (0.02)	6.54 (0.02)	5.65 (0.02)	5.76 (0.02)	5.79 (0.02)

1-amino-2-propanol	1.47	346	0.98 (0.01)	1.04 (0.01)	1.19 (0.01)	1.34 (0.01)	1.23 (0.01)	1.37 (0.01)	1.69 (0.01)	1.55 (0.01)	1.59 (0.01)	1.62 (0.01)	1.25 (0.01)
1-amino-4-fluorobenzene	3.82	386	3.20 (0.01)	3.53 (0.01)	3.18 (0.01)	3.22 (0.01)	3.62 (0.01)	3.14 (0.01)	3.01 (0.01)	3.20 (0.01)	3.34 (0.01)	3.41 (0.01)	3.20 (0.01)
1-bromobutane	3.57	382	3.78 (0.01)	3.81 (0.01)	3.81 (0.01)	3.83 (0.01)	3.77 (0.01)	3.94 (0.01)	3.74 (0.01)	3.86 (0.01)	3.80 (0.01)	3.85 (0.01)	3.86 (0.01)
1-bromoheptane	5.09	385	5.21 (0.01)	5.21 (0.01)	5.11 (0.01)	5.17 (0.01)	5.23 (0.01)	5.14 (0.01)	4.92 (0.01)	4.95 (0.01)	5.16 (0.01)	5.14 (0.01)	5.01 (0.01)
1-bromohexane	4.68	387	4.70 (0.01)	4.80 (0.01)	4.66 (0.01)	4.71 (0.01)	4.70 (0.01)	4.72 (0.01)	4.53 (0.01)	4.65 (0.01)	4.62 (0.01)	4.61 (0.01)	4.69 (0.01)
1-bromoocatne	5.36	343	5.53 (0.01)	5.61 (0.01)	5.41 (0.01)	5.58 (0.01)	5.57 (0.01)	5.46 (0.01)	5.50 (0.01)	5.51 (0.01)	5.39 (0.01)	5.35 (0.01)	5.60 (0.01)
1-bromopropane	3.26	355	3.18 (0.01)	3.33 (0.01)	3.31 (0.01)	3.21 (0.01)	3.21 (0.01)	3.20 (0.01)	3.40 (0.01)	3.28 (0.01)	3.29 (0.01)	3.30 (0.01)	3.18 (0.01)
1-butanol	1.59	377	2.00 (0.01)	1.99 (0.01)	2.18 (0.01)	2.03 (0.01)	1.94 (0.01)	2.02 (0.01)	1.88 (0.01)	1.88 (0.01)	1.83 (0.01)	1.89 (0.01)	1.95 (0.01)
1-chloro-3-methylbenzene	3.84	373	3.76 (0.01)	3.83 (0.01)	3.82 (0.01)	3.88 (0.01)	3.83 (0.01)	3.86 (0.01)	3.97 (0.01)	3.91 (0.01)	3.81 (0.01)	3.80 (0.01)	3.86 (0.01)
1-chloro-4-methylbenzene	4.33	373	4.00 (0.01)	4.03 (0.01)	3.95 (0.01)	3.89 (0.01)	3.98 (0.01)	4.08 (0.01)	4.00 (0.01)	3.90 (0.01)	3.91 (0.01)	3.90 (0.01)	3.85 (0.01)
1-chloronaphthalene	4.85	404	4.30 (0.01)	4.39 (0.01)	4.00 (0.01)	4.44 (0.01)	4.42 (0.01)	4.11 (0.01)	4.17 (0.01)	4.23 (0.01)	4.49 (0.01)	4.41 (0.01)	4.38 (0.01)
1-cyano-2-amino-5-chlorobenzene	3.73	377	4.28 (0.01)	4.21 (0.01)	4.12 (0.01)	4.30 (0.01)	4.24 (0.01)	4.07 (0.01)	4.08 (0.01)	4.22 (0.01)	4.28 (0.01)	4.48 (0.01)	4.78 (0.01)
1-cyano-2-methylbenzene	3.42	394	3.52 (0.01)	3.48 (0.01)	3.47 (0.01)	3.69 (0.01)	3.66 (0.01)	3.40 (0.01)	3.65 (0.01)	3.54 (0.01)	3.75 (0.01)	3.64 (0.01)	4.22 (0.01)
1-cyano-3,5-dibromo-4-hydroxybenzene	4.3	355	5.55 (0.03)	4.97 (0.03)	4.91 (0.03)	5.41 (0.03)	5.11 (0.03)	5.24 (0.02)	5.08 (0.03)	5.30 (0.03)	5.03 (0.03)	5.23 (0.03)	5.72 (0.03)
1-decanol	4.82	366	4.55 (0.01)	4.60 (0.01)	4.92 (0.01)	4.65 (0.01)	4.53 (0.01)	4.70 (0.01)	4.81 (0.01)	4.85 (0.01)	4.72 (0.01)	4.76 (0.01)	4.80 (0.01)
1-dodecanol	5.27	368	5.51 (0.02)	5.35 (0.02)	5.63 (0.02)	5.76 (0.02)	5.77 (0.02)	5.59 (0.02)	5.55 (0.01)	5.54 (0.02)	5.66 (0.02)	5.48 (0.02)	5.76 (0.02)
1-heptanol	3.51	378	3.64 (0.01)	3.59 (0.01)	3.79 (0.01)	3.72 (0.01)	3.64 (0.01)	3.74 (0.01)	3.68 (0.01)	3.72 (0.01)	3.67 (0.01)	3.69 (0.01)	3.73 (0.01)
1-hexanol	2.94	366	2.94 (0.01)	2.95 (0.01)	3.21 (0.01)	2.98 (0.01)	2.92 (0.01)	3.07 (0.01)	3.08 (0.01)	3.08 (0.01)	2.96 (0.01)	2.99 (0.01)	3.01 (0.01)
1-hydroxy-3-methoxybenzene	3.22	372	3.28 (0.01)	3.22 (0.01)	3.17 (0.01)	3.24 (0.01)	3.32 (0.01)	3.18 (0.01)	3.22 (0.01)	3.21 (0.01)	3.25 (0.01)	3.29 (0.01)	3.15 (0.01)
1-methylheptylamine	4.4	351	3.99 (0.01)	3.80 (0.01)	3.86 (0.01)	4.10 (0.01)	3.97 (0.01)	4.09 (0.01)	4.06 (0.01)	4.09 (0.01)	3.82 (0.01)	3.73 (0.01)	4.22 (0.01)
1-methylnaphthalene	4.2	377	3.93 (0.01)	3.93 (0.01)	3.79 (0.01)	4.01 (0.01)	3.92 (0.01)	3.86 (0.01)	3.80 (0.01)	3.98 (0.01)	4.05 (0.01)	3.92 (0.01)	4.14 (0.01)
1-naphthalenol	4.54	366	3.94 (0.01)	3.79 (0.01)	3.87 (0.01)	4.06 (0.01)	3.84 (0.01)	3.72 (0.01)	3.82 (0.01)	3.93 (0.01)	4.20 (0.01)	4.11 (0.01)	3.93 (0.01)
1-nonanol	4.41	354	4.29 (0.01)	4.36 (0.01)	4.61 (0.01)	4.37 (0.01)	4.28 (0.01)	4.48 (0.01)	4.64 (0.01)	4.61 (0.01)	4.38 (0.01)	4.41 (0.01)	4.60 (0.01)
1-octanol	3.98	369	3.82 (0.01)	3.83 (0.01)	4.08 (0.01)	3.91 (0.01)	3.83 (0.01)	3.95 (0.01)	4.13 (0.01)	4.16 (0.01)	3.88 (0.01)	3.91 (0.01)	4.01 (0.01)
1-octyn-3-ol	5.49	376	3.60 (0.01)	3.69 (0.01)	3.61 (0.01)	3.64 (0.01)	3.51 (0.01)	3.41 (0.01)	3.91 (0.01)	3.84 (0.01)	2.92 (0.01)	2.96 (0.01)	3.46 (0.01)
1-propanol	1.12	350	1.68 (0.01)	1.69 (0.01)	1.90 (0.01)	1.76 (0.01)	1.72 (0.01)	1.72 (0.01)	1.48 (0.01)	1.48 (0.01)	1.64 (0.01)	1.65 (0.01)	1.62 (0.01)
1-undecanol	5.22	366	5.11 (0.01)	5.13 (0.01)	5.49 (0.01)	5.20 (0.01)	5.07 (0.01)	5.22 (0.01)	5.24 (0.01)	5.29 (0.01)	5.29 (0.01)	5.30 (0.01)	5.25 (0.01)
2,2,2-trichloroethanol	2.8	387	3.37 (0.01)	3.33 (0.01)	3.21 (0.01)	3.35 (0.01)	3.19 (0.01)	3.59 (0.01)	2.90 (0.01)	3.09 (0.01)	3.53 (0.01)	3.30 (0.01)	3.33 (0.01)
2,2-dimethyl-1-propylamine	2.26	363	2.16 (0.01)	2.16 (0.01)	2.43 (0.01)	2.13 (0.01)	2.28 (0.01)	2.54 (0.01)	2.15 (0.01)	2.58 (0.01)	2.08 (0.01)	2.24 (0.01)	2.15 (0.01)

2,2'-methylenebis(-4-chloro)phenol	5.94	404	5.02 (0.02)	5.03 (0.02)	6.00 (0.02)	4.91 (0.02)	5.14 (0.02)	6.04 (0.02)	5.19 (0.02)	4.61 (0.02)	5.48 (0.02)	5.22 (0.02)	5.44 (0.02)
2,3,4,5-tetrachlorophenate	5.74	371	5.53 (0.01)	5.66 (0.01)	5.52 (0.01)	5.59 (0.01)	5.49 (0.01)	5.57 (0.01)	5.59 (0.01)	5.62 (0.01)	5.53 (0.01)	5.59 (0.01)	5.47 (0.01)
2,3,4,6-tetrachlorophenol	5.35	397	5.43 (0.01)	5.37 (0.01)	5.47 (0.01)	5.22 (0.01)	5.40 (0.01)	5.50 (0.01)	5.51 (0.01)	5.68 (0.01)	5.48 (0.01)	5.60 (0.01)	5.45 (0.01)
2,3,4-trichloroacetophenone	5.05	341	4.53 (0.01)	4.45 (0.01)	5.38 (0.02)	5.12 (0.01)	4.70 (0.02)	4.83 (0.01)	4.86 (0.01)	4.96 (0.01)	5.08 (0.01)	4.96 (0.01)	4.98 (0.01)
2,3,4-trichloroaniline	4.74	369	4.76 (0.01)	5.05 (0.01)	4.36 (0.01)	4.45 (0.01)	4.91 (0.01)	4.65 (0.01)	4.65 (0.01)	4.49 (0.01)	5.22 (0.01)	4.75 (0.01)	4.65 (0.01)
2,3,5,6-tetrachloroaniline	5.93	395	5.74 (0.01)	5.68 (0.01)	5.87 (0.01)	5.37 (0.01)	5.93 (0.01)	5.75 (0.01)	5.15 (0.01)	5.47 (0.01)	5.46 (0.01)	5.57 (0.01)	5.78 (0.01)
2,3,6-trimethylphenol	4.22	351	4.26 (0.01)	4.34 (0.01)	4.19 (0.01)	4.09 (0.01)	4.17 (0.01)	4.50 (0.01)	3.80 (0.01)	3.93 (0.01)	4.25 (0.01)	4.34 (0.01)	4.36 (0.01)
2,3-dimethyl-1,3-butadiene	4.08	343	3.88 (0.01)	3.65 (0.01)	4.00 (0.01)	4.14 (0.01)	4.00 (0.01)	3.80 (0.01)	4.11 (0.02)	4.01 (0.01)	3.33 (0.01)	2.96 (0.01)	3.90 (0.01)
2,4,5-trichlorophenole	5.34	378	5.05 (0.01)	5.09 (0.01)	5.13 (0.01)	5.15 (0.01)	5.06 (0.01)	5.13 (0.01)	5.14 (0.01)	5.14 (0.01)	4.81 (0.01)	4.83 (0.01)	4.90 (0.01)
2,4,6-trichloroaniline	4.59	355	4.60 (0.01)	4.64 (0.01)	4.90 (0.01)	4.51 (0.01)	4.52 (0.01)	4.37 (0.01)	4.32 (0.01)	4.57 (0.01)	4.49 (0.01)	4.61 (0.01)	4.75 (0.01)
2,4,6-trichlorophenol	4.64	359	4.63 (0.01)	4.71 (0.01)	4.75 (0.01)	4.74 (0.01)	4.48 (0.01)	4.79 (0.01)	4.87 (0.01)	4.99 (0.01)	4.69 (0.01)	4.76 (0.01)	4.77 (0.01)
2,4,6-trimethylphenol	4.02	378	4.22 (0.01)	4.15 (0.01)	4.46 (0.01)	4.31 (0.01)	4.10 (0.01)	4.31 (0.01)	4.23 (0.01)	4.37 (0.01)	4.27 (0.01)	4.18 (0.01)	4.41 (0.01)
2,4-dichloroaniline	4.07	354	3.93 (0.01)	4.45 (0.01)	4.00 (0.01)	3.77 (0.01)	4.31 (0.01)	4.08 (0.01)	4.07 (0.01)	3.85 (0.01)	4.47 (0.01)	4.56 (0.01)	3.90 (0.01)
2,4-dichlorobenzaldehyde	4.99	377	4.89 (0.01)	4.93 (0.01)	4.62 (0.01)	5.10 (0.01)	4.97 (0.01)	4.88 (0.01)	4.78 (0.01)	4.72 (0.01)	4.75 (0.01)	4.80 (0.01)	4.80 (0.01)
2,4-dichlorophenol	4.3	368	4.51 (0.01)	4.46 (0.01)	4.56 (0.01)	4.52 (0.01)	4.37 (0.01)	4.58 (0.01)	4.61 (0.01)	4.62 (0.01)	4.36 (0.01)	4.39 (0.01)	4.47 (0.01)
2,4-dichlorotoluene	4.54	352	4.25 (0.01)	4.27 (0.01)	4.23 (0.01)	4.32 (0.01)	4.26 (0.01)	4.34 (0.01)	4.44 (0.01)	4.29 (0.01)	4.25 (0.01)	4.24 (0.01)	4.26 (0.01)
2,4-dimethoxybenzaldehyde	3.92	354	4.31 (0.01)	4.29 (0.01)	4.02 (0.01)	4.14 (0.01)	4.29 (0.01)	4.17 (0.01)	3.91 (0.02)	3.79 (0.01)	4.19 (0.01)	4.12 (0.01)	3.93 (0.02)
2,4-dimethyl-3-pentanol	2.85	364	2.13 (0.02)	2.16 (0.02)	1.83 (0.02)	2.30 (0.02)	2.56 (0.02)	1.73 (0.02)	1.31 (0.02)	2.36 (0.01)	2.05 (0.01)	1.94 (0.02)	1.81 (0.02)
2,4-hexadiene	3.61	337	3.71 (0.01)	3.77 (0.01)	3.73 (0.01)	3.56 (0.01)	3.74 (0.01)	3.85 (0.01)	3.68 (0.01)	3.57 (0.01)	4.05 (0.01)	4.11 (0.01)	3.53 (0.01)
2,4-xylenol	3.86	357	3.76 (0.01)	3.68 (0.01)	3.72 (0.01)	3.72 (0.01)	3.60 (0.01)	3.59 (0.01)	3.73 (0.01)	3.77 (0.01)	3.64 (0.01)	3.58 (0.01)	3.63 (0.01)
2,5-dimethyl-2,4-hexadiene	4.46	358	4.36 (0.02)	4.39 (0.02)	4.61 (0.02)	4.03 (0.02)	4.22 (0.02)	4.53 (0.02)	3.88 (0.02)	3.95 (0.02)	5.35 (0.02)	5.40 (0.02)	3.74 (0.02)
2,6-diisopropylaniline	4.1	348	4.07 (0.02)	3.94 (0.02)	3.75 (0.02)	3.87 (0.02)	4.10 (0.02)	3.29 (0.02)	3.60 (0.02)	3.77 (0.02)	4.19 (0.02)	3.70 (0.01)	4.00 (0.02)
2,6-dimethoxytoluene	3.88	349	3.36 (0.01)	3.31 (0.01)	3.60 (0.01)	3.11 (0.01)	3.17 (0.01)	3.30 (0.01)	3.49 (0.01)	3.16 (0.01)	3.14 (0.01)	3.26 (0.01)	3.12 (0.01)
2,6-di-tert-butyl-4-methylphenol	5.78	374	6.47 (0.02)	6.38 (0.03)	4.12 (0.03)	6.91 (0.03)	6.68 (0.03)	4.56 (0.03)	6.09 (0.03)	5.23 (0.03)	4.88 (0.03)	4.75 (0.03)	5.10 (0.02)
2-allylphenol	3.95	375	4.76 (0.01)	4.53 (0.01)	4.28 (0.01)	4.56 (0.01)	4.37 (0.01)	4.49 (0.01)	4.55 (0.01)	4.52 (0.01)	4.18 (0.01)	4.15 (0.01)	4.35 (0.01)
2-aminoethanol	1.47	392	0.87 (0.01)	1.03 (0.01)	0.82 (0.01)	1.05 (0.01)	1.17 (0.01)	1.05 (0.01)	1.29 (0.01)	1.21 (0.01)	0.88 (0.01)	1.32 (0.01)	0.82 (0.01)
2-butanamine	2.42	351	2.17 (0.01)	2.03 (0.01)	2.14 (0.01)	2.15 (0.01)	2.05 (0.01)	2.21 (0.01)	2.28 (0.01)	2.26 (0.01)	2.07 (0.01)	2.05 (0.01)	2.33 (0.01)
2-butanol	1.31	359	2.06 (0.01)	1.97 (0.01)	2.07 (0.01)	2.11 (0.01)	2.00 (0.01)	2.04 (0.01)	2.25 (0.01)	2.19 (0.01)	1.60 (0.01)	1.57 (0.01)	1.97 (0.01)
2-butanone	1.35	377	2.22 (0.01)	2.21 (0.01)	2.10 (0.01)	2.28 (0.01)	2.27 (0.01)	2.11 (0.01)	2.28 (0.01)	2.17 (0.01)	2.07 (0.01)	2.04 (0.01)	2.32 (0.01)
2-butyn-1-ol	3.84	387	2.66 (0.01)	2.39 (0.01)	2.50 (0.01)	2.57 (0.01)	2.36 (0.02)	2.10 (0.02)	2.41 (0.01)	2.71 (0.01)	2.10 (0.02)	2.07 (0.01)	2.14 (0.02)

2-butyne-1,4-diol	3.21	360	3.07 (0.02)	2.75 (0.03)	2.91 (0.03)	3.00 (0.03)	2.59 (0.03)	2.60 (0.03)	2.76 (0.02)	3.18 (0.02)	2.40 (0.02)	2.64 (0.02)	2.66 (0.03)
2-chloro-4-methylaniline	3.59	369	3.81 (0.01)	3.88 (0.01)	3.97 (0.01)	3.75 (0.01)	3.64 (0.01)	3.69 (0.01)	3.99 (0.01)	3.89 (0.01)	3.97 (0.01)	3.98 (0.01)	4.00 (0.01)
2-chloro-6-fluorobenzaldehyde	4.23	395	4.69 (0.01)	4.41 (0.01)	4.71 (0.01)	4.82 (0.01)	4.64 (0.01)	4.64 (0.01)	4.39 (0.01)	4.67 (0.01)	4.29 (0.01)	4.38 (0.01)	4.44 (0.01)
2-chlorotoluene	4.23	375	3.79 (0.01)	3.75 (0.01)	3.81 (0.01)	3.83 (0.01)	3.77 (0.01)	3.84 (0.01)	4.00 (0.01)	3.90 (0.01)	3.77 (0.01)	3.67 (0.01)	3.88 (0.01)
2-cyano-6-methyl-benzonitrile	4	349	3.90 (0.01)	3.71 (0.01)	4.11 (0.01)	4.10 (0.01)	3.96 (0.01)	3.85 (0.01)	4.15 (0.01)	4.07 (0.01)	4.21 (0.01)	4.17 (0.01)	4.35 (0.01)
2-decanone	4.5	373	4.70 (0.01)	4.72 (0.01)	4.67 (0.01)	4.73 (0.01)	4.73 (0.01)	4.61 (0.01)	4.69 (0.01)	4.49 (0.01)	4.72 (0.01)	4.71 (0.01)	4.93 (0.01)
2-dimethylaminopyridine	2.98	383	3.03 (0.01)	3.37 (0.01)	2.96 (0.01)	2.89 (0.01)	2.79 (0.01)	2.66 (0.01)	2.77 (0.01)	2.66 (0.01)	2.40 (0.01)	2.04 (0.01)	2.59 (0.01)
2-(disopropylamino)-ethanol	2.86	352	3.32 (0.02)	3.38 (0.02)	2.84 (0.02)	3.10 (0.02)	3.17 (0.02)	3.45 (0.02)	3.24 (0.02)	3.75 (0.02)	3.35 (0.01)	3.34 (0.02)	2.79 (0.02)
2-dodecanone	5.19	388	5.64 (0.01)	5.60 (0.01)	5.53 (0.01)	5.66 (0.01)	5.57 (0.01)	5.47 (0.01)	5.29 (0.01)	5.08 (0.01)	5.67 (0.01)	5.64 (0.01)	5.61 (0.01)
2-ethoxyethylacetate	3.5	339	2.73 (0.01)	2.75 (0.02)	3.07 (0.02)	3.10 (0.01)	2.91 (0.01)	2.92 (0.01)	2.99 (0.01)	2.63 (0.01)	2.77 (0.01)	2.89 (0.02)	2.79 (0.02)
2-ethyl-1-hexanol	3.66	352	2.60 (0.01)	2.43 (0.01)	2.79 (0.01)	3.06 (0.01)	3.00 (0.01)	2.54 (0.02)	2.52 (0.02)	2.73 (0.01)	2.65 (0.01)	2.76 (0.01)	2.87 (0.02)
2-(ethylamine)-ethanol	1.78	394	1.76 (0.01)	1.99 (0.01)	1.97 (0.01)	1.69 (0.01)	1.86 (0.01)	2.12 (0.01)	2.30 (0.01)	2.25 (0.01)	2.20 (0.01)	2.12 (0.01)	1.79 (0.01)
2-fluorobenzaldehyde	4.96	380	4.13 (0.01)	4.02 (0.01)	3.99 (0.01)	4.14 (0.01)	3.93 (0.01)	4.08 (0.01)	3.90 (0.01)	4.11 (0.01)	3.77 (0.01)	3.80 (0.01)	3.77 (0.01)
2-heptanone	2.94	377	3.32 (0.01)	3.35 (0.01)	3.21 (0.01)	3.32 (0.01)	3.40 (0.01)	3.19 (0.01)	3.23 (0.01)	3.20 (0.01)	3.36 (0.01)	3.37 (0.01)	3.31 (0.01)
2-hexanone	2.37	343	3.13 (0.01)	3.19 (0.01)	3.04 (0.01)	3.11 (0.01)	3.19 (0.01)	3.08 (0.01)	2.95 (0.01)	2.99 (0.01)	3.13 (0.01)	3.10 (0.01)	3.11 (0.01)
2-hydroxy-4,6-dimethoxybenzaldehyde	4.83	355	4.37 (0.01)	4.23 (0.01)	4.16 (0.01)	3.95 (0.01)	4.50 (0.02)	4.00 (0.01)	3.93 (0.01)	3.96 (0.02)	4.51 (0.01)	4.52 (0.02)	3.83 (0.01)
2-hydroxy-5-bromobenzaldehyde	5.19	368	5.00 (0.02)	4.59 (0.02)	5.15 (0.02)	5.18 (0.02)	4.89 (0.02)	5.01 (0.01)	5.06 (0.01)	4.97 (0.01)	5.18 (0.01)	4.99 (0.01)	5.04 (0.02)
2-hydroxy-5-chlorobenzaldehyde	5.31	338	4.66 (0.01)	4.57 (0.01)	4.86 (0.01)	4.63 (0.01)	4.72 (0.01)	4.64 (0.01)	4.77 (0.01)	4.71 (0.01)	4.77 (0.01)	4.78 (0.01)	4.69 (0.01)
2-hydroxybenzaldehydesalicylaldehyde	4.73	368	4.11 (0.01)	4.01 (0.01)	4.32 (0.01)	4.12 (0.01)	4.19 (0.01)	4.03 (0.01)	4.15 (0.01)	4.15 (0.01)	4.28 (0.01)	4.15 (0.01)	4.13 (0.01)
2-hydroxyethylacrylate	4.38	336	3.13 (0.01)	3.04 (0.01)	2.98 (0.02)	3.09 (0.01)	2.86 (0.02)	2.98 (0.02)	2.99 (0.01)	3.05 (0.02)	3.22 (0.02)	2.79 (0.01)	2.78 (0.02)
2-hydroxyethylmethacrylate	2.76	368	3.51 (0.01)	3.51 (0.01)	3.59 (0.02)	3.13 (0.01)	3.15 (0.01)	3.38 (0.02)	3.54 (0.01)	3.80 (0.01)	3.79 (0.01)	3.02 (0.02)	3.27 (0.01)
2-hydroxypropylacrylate	4.59	353	3.49 (0.02)	3.41 (0.02)	3.44 (0.02)	3.72 (0.02)	3.47 (0.02)	3.48 (0.02)	3.56 (0.02)	3.70 (0.02)	3.63 (0.02)	3.04 (0.01)	3.39 (0.02)
2-methoxyethylamine	2.16	356	1.55 (0.01)	1.43 (0.01)	1.63 (0.01)	1.48 (0.01)	1.56 (0.01)	1.43 (0.01)	1.44 (0.01)	1.58 (0.01)	1.42 (0.01)	1.45 (0.01)	1.88 (0.01)
2-methyl-1-propanol	1.69	385	1.56 (0.01)	1.63 (0.01)	1.86 (0.01)	1.64 (0.01)	1.59 (0.01)	1.65 (0.01)	2.01 (0.01)	2.00 (0.01)	1.83 (0.01)	1.86 (0.01)	1.77 (0.01)
2-methyl-2-propanol	1.06	380	1.22 (0.01)	1.27 (0.01)	1.46 (0.01)	1.16 (0.01)	1.25 (0.01)	1.20 (0.01)	0.97 (0.01)	1.11 (0.01)	0.94 (0.01)	1.01 (0.01)	0.84 (0.01)
2-methylpyridine	2.02	376	2.91 (0.01)	2.78 (0.01)	2.78 (0.01)	2.83 (0.01)	2.82 (0.01)	3.04 (0.01)	2.80 (0.01)	2.82 (0.01)	2.96 (0.01)	2.91 (0.01)	2.66 (0.01)
2-naphthol	4.62	360	4.32 (0.01)	4.19 (0.01)	3.91 (0.01)	4.46 (0.01)	4.36 (0.01)	4.14 (0.01)	4.14 (0.01)	4.18 (0.01)	4.28 (0.01)	4.25 (0.01)	4.19 (0.01)
2-nonanone	3.97	376	4.19 (0.01)	4.23 (0.01)	4.11 (0.01)	4.16 (0.01)	4.20 (0.01)	4.08 (0.01)	4.16 (0.01)	4.08 (0.01)	4.17 (0.01)	4.18 (0.01)	4.28 (0.01)
2-octanone	3.45	376	3.92 (0.01)	3.95 (0.01)	3.85 (0.01)	3.95 (0.01)	4.00 (0.01)	3.84 (0.01)	4.01 (0.01)	3.95 (0.01)	3.98 (0.01)	3.95 (0.01)	4.01 (0.01)
2-pentanone	1.84	370	2.34 (0.01)	2.44 (0.01)	2.28 (0.01)	2.32 (0.01)	2.41 (0.01)	2.31 (0.01)	2.16 (0.01)	2.22 (0.01)	2.28 (0.01)	2.31 (0.01)	2.24 (0.01)

2-phenoxyethanol	2.6	379	3.30 (0.01)	3.07 (0.01)	3.15 (0.01)	3.14 (0.01)	3.05 (0.01)	3.45 (0.01)	3.27 (0.01)	3.19 (0.01)	3.16 (0.01)	3.04 (0.01)	3.14 (0.01)
2-pyridinecarbonitrile	2.16	364	3.27 (0.01)	3.35 (0.01)	3.57 (0.02)	3.20 (0.01)	3.31 (0.01)	3.44 (0.01)	3.55 (0.01)	3.53 (0.01)	3.56 (0.01)	3.80 (0.01)	3.90 (0.01)
3,3'-dichlorobenzidine	5.09	352	5.12 (0.02)	4.85 (0.03)	5.29 (0.02)	5.59 (0.03)	5.29 (0.02)	5.79 (0.02)	5.73 (0.02)	5.52 (0.03)	5.54 (0.02)	5.48 (0.02)	5.37 (0.03)
3,3-dimethyl-2-butanone	3.06	365	2.87 (0.01)	2.86 (0.01)	2.71 (0.01)	2.71 (0.01)	2.69 (0.01)	3.02 (0.01)	2.78 (0.02)	3.21 (0.02)	2.97 (0.01)	3.02 (0.01)	2.93 (0.01)
3,3-dimethylbutylamine	2.23	378	2.18 (0.01)	2.45 (0.01)	2.30 (0.01)	2.32 (0.01)	2.55 (0.01)	2.41 (0.01)	2.91 (0.01)	2.79 (0.01)	2.80 (0.01)	2.75 (0.01)	2.47 (0.01)
3,4-dichloroaniline	4.32	354	4.68 (0.01)	4.68 (0.01)	4.75 (0.01)	4.53 (0.01)	4.64 (0.01)	4.07 (0.01)	4.19 (0.01)	4.78 (0.01)	4.41 (0.01)	4.47 (0.01)	4.44 (0.01)
3,4-dichlorotoluene	4.74	374	4.34 (0.01)	4.34 (0.01)	4.44 (0.01)	4.22 (0.01)	4.20 (0.01)	4.46 (0.01)	4.56 (0.01)	4.48 (0.01)	4.36 (0.01)	4.38 (0.01)	4.35 (0.01)
3,4-xylenol	3.94	364	4.02 (0.01)	3.91 (0.01)	3.80 (0.01)	4.04 (0.01)	4.02 (0.01)	4.09 (0.01)	3.95 (0.01)	3.59 (0.01)	3.91 (0.01)	3.94 (0.01)	4.21 (0.01)
3-chloro-1-propanol	2.07	359	2.61 (0.01)	2.40 (0.01)	2.80 (0.01)	2.60 (0.01)	2.67 (0.01)	2.60 (0.01)	2.31 (0.01)	2.27 (0.01)	2.35 (0.01)	2.51 (0.01)	2.58 (0.01)
3-methoxy-4-hydroxybenzaldehyde	3.12	370	3.47 (0.01)	3.50 (0.01)	3.40 (0.01)	3.72 (0.01)	3.89 (0.01)	3.42 (0.01)	3.40 (0.01)	3.43 (0.01)	3.82 (0.01)	3.92 (0.01)	3.57 (0.01)
3-methyl-2-butanone	2	374	2.65 (0.01)	2.62 (0.01)	2.63 (0.01)	2.66 (0.01)	2.78 (0.01)	2.65 (0.01)	2.38 (0.01)	2.50 (0.01)	3.05 (0.01)	2.99 (0.01)	2.81 (0.01)
3-methyl-3-pentanol	2.18	306	2.85 (0.01)	2.66 (0.01)	2.32 (0.01)	2.85 (0.01)	2.81 (0.01)	2.82 (0.01)	2.25 (0.01)	2.47 (0.01)	2.45 (0.01)	2.50 (0.01)	2.84 (0.01)
3-methylbutanal	4.42	394	2.61 (0.01)	2.66 (0.01)	2.35 (0.01)	2.61 (0.01)	2.73 (0.01)	2.55 (0.01)	2.78 (0.01)	2.74 (0.01)	2.88 (0.01)	2.86 (0.01)	2.88 (0.01)
3-methylpyridine	2.81	375	2.68 (0.01)	2.61 (0.01)	2.61 (0.01)	2.61 (0.01)	2.73 (0.01)	2.68 (0.01)	2.71 (0.01)	2.72 (0.01)	2.78 (0.01)	2.62 (0.01)	2.54 (0.01)
3-pentanone	1.75	368	2.43 (0.01)	2.41 (0.01)	2.14 (0.01)	2.44 (0.01)	2.50 (0.01)	2.16 (0.01)	2.41 (0.01)	2.48 (0.01)	2.29 (0.01)	2.29 (0.01)	2.35 (0.01)
3-pyridinecarboxaldehyde	3.81	366	3.56 (0.01)	3.64 (0.01)	3.59 (0.01)	3.58 (0.01)	3.67 (0.01)	3.72 (0.01)	3.42 (0.01)	3.60 (0.01)	3.58 (0.01)	3.61 (0.01)	3.32 (0.01)
3-pyridinepropanol	2.96	341	2.80 (0.01)	2.83 (0.01)	3.12 (0.01)	2.81 (0.01)	2.72 (0.01)	2.95 (0.01)	2.86 (0.01)	2.97 (0.01)	3.01 (0.01)	3.11 (0.01)	3.20 (0.01)
3-(trifluoromethyl)-benzonitrile	3.56	355	3.65 (0.02)	4.12 (0.02)	3.04 (0.02)	3.67 (0.02)	4.10 (0.02)	4.03 (0.02)	3.44 (0.02)	3.79 (0.02)	3.79 (0.02)	3.66 (0.02)	3.87 (0.02)
4-acetylpyridine	2.86	345	2.91 (0.01)	3.00 (0.01)	3.21 (0.01)	2.94 (0.01)	2.88 (0.01)	3.13 (0.01)	3.06 (0.01)	2.93 (0.01)	2.50 (0.01)	2.57 (0.01)	3.04 (0.01)
4-aminopropiophenone	3.01	396	3.56 (0.01)	3.92 (0.01)	3.94 (0.01)	3.63 (0.01)	3.49 (0.01)	3.58 (0.01)	3.78 (0.01)	3.78 (0.01)	3.56 (0.01)	3.56 (0.01)	3.88 (0.01)
4-benzoylpyridine	3.25	392	4.10 (0.01)	4.13 (0.01)	4.23 (0.01)	4.05 (0.01)	3.94 (0.01)	3.74 (0.01)	4.45 (0.01)	3.94 (0.01)	3.55 (0.01)	3.70 (0.01)	3.67 (0.01)
4-bromoaniline	3.56	356	4.42 (0.01)	4.47 (0.01)	4.21 (0.01)	4.47 (0.01)	4.50 (0.01)	4.19 (0.01)	4.38 (0.01)	4.64 (0.01)	4.40 (0.01)	4.42 (0.01)	4.40 (0.01)
4-butylaniline	4.17	392	4.55 (0.01)	4.53 (0.01)	4.37 (0.01)	4.55 (0.01)	4.64 (0.01)	4.48 (0.01)	4.37 (0.01)	4.51 (0.01)	4.58 (0.01)	4.56 (0.01)	4.59 (0.01)
4-chlorobenzaldehyde	4.81	367	4.53 (0.01)	4.66 (0.01)	4.64 (0.01)	4.61 (0.01)	4.62 (0.01)	4.78 (0.01)	4.40 (0.01)	4.49 (0.01)	4.63 (0.01)	4.62 (0.01)	4.45 (0.01)
4-decylaniline	6.57	369	6.44 (0.02)	6.64 (0.02)	6.38 (0.02)	6.70 (0.02)	6.75 (0.02)	6.58 (0.02)	6.74 (0.02)	6.49 (0.02)	6.60 (0.02)	6.47 (0.02)	6.55 (0.02)
4-dimethylaminobenzaldehyde	3.51	390	4.69 (0.01)	4.47 (0.01)	3.88 (0.01)	4.58 (0.01)	4.21 (0.01)	3.99 (0.01)	4.00 (0.01)	4.11 (0.01)	4.34 (0.01)	4.23 (0.01)	4.10 (0.01)
4-ethoxybenzaldehyde	3.74	392	4.14 (0.01)	4.13 (0.01)	4.20 (0.01)	4.03 (0.01)	4.10 (0.01)	4.32 (0.01)	3.89 (0.01)	3.77 (0.01)	4.30 (0.01)	4.22 (0.01)	4.05 (0.01)
4-ethylaniline	3.22	358	3.58 (0.01)	3.72 (0.01)	3.54 (0.01)	3.50 (0.01)	3.57 (0.01)	3.46 (0.01)	3.48 (0.01)	3.56 (0.01)	3.77 (0.01)	3.82 (0.01)	3.74 (0.01)

4-hexyloxyaniline	4.78	393	4.74 (0.01)	4.92 (0.01)	4.96 (0.01)	4.65 (0.01)	4.75 (0.01)	4.95 (0.01)	4.75 (0.01)	4.91 (0.01)	4.86 (0.01)	4.72 (0.01)	4.80 (0.01)
4-isopropylbenzaldehyde	4.35	368	4.56 (0.01)	4.63 (0.01)	4.54 (0.02)	4.38 (0.01)	4.40 (0.01)	4.68 (0.01)	4.24 (0.01)	4.29 (0.01)	4.66 (0.01)	4.67 (0.01)	4.44 (0.01)
4-methoxyphenol	3.27	373	3.27 (0.01)	3.23 (0.01)	3.34 (0.01)	3.24 (0.01)	3.27 (0.01)	3.14 (0.01)	3.30 (0.01)	3.19 (0.01)	3.37 (0.01)	3.40 (0.01)	3.23 (0.01)
4-methyl-2-pentanone	2.27	374	2.49 (0.01)	2.58 (0.01)	2.43 (0.01)	2.54 (0.01)	2.66 (0.01)	2.41 (0.01)	2.39 (0.01)	2.77 (0.01)	2.61 (0.01)	2.64 (0.01)	2.68 (0.01)
4-methylaniline	2.83	361	3.49 (0.01)	3.57 (0.01)	3.39 (0.01)	3.50 (0.01)	3.49 (0.01)	3.39 (0.01)	3.45 (0.01)	3.54 (0.01)	3.38 (0.01)	3.31 (0.01)	3.46 (0.01)
4-methylpyridine	2.36	362	2.96 (0.01)	2.94 (0.01)	2.84 (0.01)	2.89 (0.01)	3.05 (0.01)	3.03 (0.01)	3.02 (0.01)	3.01 (0.01)	2.81 (0.01)	2.76 (0.01)	2.72 (0.01)
4-octylaniline	6.24	354	5.86 (0.01)	5.84 (0.01)	5.71 (0.01)	6.02 (0.01)	6.13 (0.01)	5.98 (0.01)	6.14 (0.01)	5.96 (0.01)	5.87 (0.01)	5.78 (0.01)	5.91 (0.01)
4-phenoxyphenol	4.58	364	4.24 (0.01)	4.74 (0.01)	4.59 (0.01)	4.70 (0.01)	4.92 (0.01)	4.63 (0.01)	4.89 (0.01)	4.54 (0.01)	4.19 (0.01)	4.59 (0.01)	4.74 (0.01)
4-phenylpyridine	3.98	355	3.76 (0.01)	3.53 (0.02)	4.03 (0.01)	4.30 (0.01)	4.24 (0.01)	3.95 (0.01)	3.99 (0.02)	4.25 (0.01)	4.12 (0.01)	3.88 (0.01)	4.03 (0.01)
5-(diethylamino)-2-pentanone	2.67	348	3.34 (0.02)	3.72 (0.02)	3.83 (0.01)	3.60 (0.02)	3.36 (0.02)	2.97 (0.02)	2.90 (0.02)	3.77 (0.02)	2.96 (0.02)	3.13 (0.02)	3.51 (0.02)
5-ethyl-2-methylpyridine	3.17	352	3.23 (0.01)	3.08 (0.01)	3.20 (0.01)	3.17 (0.01)	3.22 (0.01)	3.52 (0.01)	3.06 (0.01)	3.12 (0.01)	3.31 (0.01)	3.45 (0.01)	2.97 (0.01)
5-methyl-2-hexanone	2.86	364	3.18 (0.01)	3.12 (0.01)	3.21 (0.01)	3.31 (0.01)	3.30 (0.01)	2.95 (0.01)	2.87 (0.01)	3.00 (0.01)	3.09 (0.01)	3.08 (0.01)	3.04 (0.01)
5-nonanone	3.66	388	4.11 (0.01)	4.24 (0.01)	3.87 (0.01)	4.14 (0.01)	4.31 (0.01)	3.86 (0.01)	3.83 (0.01)	3.94 (0.01)	4.20 (0.01)	4.21 (0.01)	3.99 (0.01)
acetaldehyde	3.11	385	1.43 (0.01)	1.06 (0.01)	1.14 (0.01)	1.24 (0.01)	1.30 (0.01)	1.38 (0.01)	1.11 (0.01)	1.16 (0.01)	1.84 (0.01)	1.75 (0.01)	1.72 (0.01)
aceticacid	2.85	364	1.29 (0.01)	1.57 (0.01)	1.74 (0.02)	1.21 (0.01)	1.15 (0.01)	1.32 (0.01)	1.30 (0.01)	1.17 (0.01)	1.56 (0.01)	1.55 (0.01)	1.52 (0.01)
acetone	0.85	391	1.23 (0.01)	1.41 (0.01)	1.33 (0.01)	1.35 (0.01)	1.36 (0.01)	1.22 (0.01)	1.18 (0.01)	1.19 (0.01)	1.20 (0.01)	1.24 (0.01)	1.27 (0.01)
acetonitrile	1.49	372	0.95 (0.02)	1.19 (0.02)	1.10 (0.02)	1.21 (0.02)	1.21 (0.02)	0.87 (0.02)	1.19 (0.02)	1.10 (0.02)	1.93 (0.02)	1.86 (0.02)	1.48 (0.02)
acetophenone	2.87	370	3.08 (0.01)	3.19 (0.01)	3.40 (0.01)	3.17 (0.01)	3.00 (0.01)	3.27 (0.01)	3.32 (0.01)	3.33 (0.01)	2.97 (0.01)	2.90 (0.01)	3.38 (0.01)
acridine	4.89	375	5.39 (0.02)	4.96 (0.01)	4.52 (0.01)	5.51 (0.02)	5.28 (0.02)	5.57 (0.02)	4.76 (0.02)	5.01 (0.02)	4.91 (0.02)	4.73 (0.02)	4.90 (0.02)
adipicacid	3.18	394	3.98 (0.02)	3.88 (0.02)	4.30 (0.02)	3.97 (0.02)	3.94 (0.01)	3.63 (0.02)	4.06 (0.02)	4.43 (0.02)	4.00 (0.02)	4.29 (0.02)	3.82 (0.02)
aniline	3.03	367	2.92 (0.01)	3.38 (0.01)	2.77 (0.01)	2.93 (0.01)	3.39 (0.01)	2.70 (0.01)	2.83 (0.01)	2.96 (0.01)	3.24 (0.01)	3.19 (0.01)	3.01 (0.01)
benzaldehyde	3.93	371	3.90 (0.01)	3.99 (0.01)	3.93 (0.01)	4.02 (0.01)	3.98 (0.01)	4.00 (0.01)	3.74 (0.01)	3.88 (0.01)	3.95 (0.01)	3.88 (0.01)	3.86 (0.01)
benzene	3.5	388	3.48 (0.01)	3.50 (0.01)	3.48 (0.01)	3.43 (0.01)	3.40 (0.01)	3.39 (0.01)	3.43 (0.01)	3.44 (0.01)	3.27 (0.01)	3.20 (0.01)	3.44 (0.01)
benzonitrile	2.98	365	3.27 (0.01)	3.22 (0.01)	3.36 (0.02)	3.45 (0.01)	3.39 (0.01)	3.15 (0.01)	3.49 (0.01)	3.57 (0.02)	3.81 (0.01)	3.80 (0.01)	4.14 (0.01)
benzophenone	4.11	365	4.34 (0.02)	4.38 (0.02)	4.10 (0.02)	4.20 (0.02)	3.98 (0.02)	3.71 (0.02)	4.77 (0.02)	4.14 (0.02)	3.70 (0.02)	3.74 (0.02)	3.80 (0.02)
benzylalcohol	2.37	351	3.39 (0.01)	3.17 (0.01)	3.68 (0.01)	3.32 (0.01)	3.34 (0.01)	3.34 (0.01)	2.78 (0.01)	3.28 (0.01)	2.54 (0.01)	2.60 (0.01)	2.94 (0.01)
benzylamine	3.02	361	3.24 (0.01)	3.54 (0.01)	3.52 (0.01)	3.25 (0.01)	3.59 (0.01)	3.47 (0.01)	3.32 (0.01)	3.07 (0.01)	3.21 (0.01)	3.39 (0.01)	3.37 (0.01)
benzylchloride	4.4	382	4.27 (0.02)	4.27 (0.02)	4.20 (0.02)	4.28 (0.02)	4.34 (0.02)	4.22 (0.01)	4.02 (0.02)	4.39 (0.02)	4.07 (0.01)	4.04 (0.01)	4.17 (0.02)
biphenyl	4.8	373	4.41 (0.01)	4.05 (0.01)	4.68 (0.01)	4.91 (0.01)	4.62 (0.01)	4.73 (0.01)	4.73 (0.01)	5.09 (0.01)	4.83 (0.01)	4.47 (0.01)	4.81 (0.01)

bromobenzene	4.45	381	4.34 (0.01)	4.01 (0.01)	4.18 (0.01)	4.49 (0.01)	4.04 (0.01)	4.42 (0.01)	4.38 (0.01)	4.49 (0.01)	4.18 (0.01)	4.12 (0.01)	4.43 (0.01)
butanal	3.65	384	2.55 (0.01)	2.54 (0.01)	2.26 (0.01)	2.55 (0.01)	2.63 (0.01)	2.45 (0.01)	2.28 (0.01)	2.35 (0.01)	2.55 (0.01)	2.53 (0.01)	2.52 (0.01)
camphor	3.14	353	2.71 (0.04)	3.12 (0.03)	3.00 (0.03)	2.96 (0.04)	2.42 (0.04)	2.77 (0.03)	3.35 (0.04)	2.56 (0.04)	3.25 (0.03)	3.19 (0.03)	3.38 (0.03)
carbontetrachloride	3.56	376	3.12 (0.02)	3.34 (0.03)	3.04 (0.03)	3.22 (0.03)	3.07 (0.03)	2.74 (0.03)	3.58 (0.02)	3.13 (0.03)	2.85 (0.03)	3.05 (0.03)	3.07 (0.03)
chlorobenzene	3.7	364	4.02 (0.01)	4.03 (0.01)	4.04 (0.01)	4.07 (0.01)	3.98 (0.01)	4.03 (0.01)	4.08 (0.01)	4.04 (0.01)	3.98 (0.01)	3.99 (0.01)	4.05 (0.01)
chloroform	3.06	362	3.11 (0.02)	3.14 (0.01)	2.93 (0.01)	3.20 (0.02)	3.25 (0.02)	3.06 (0.02)	3.25 (0.01)	3.60 (0.02)	2.63 (0.02)	2.86 (0.02)	2.97 (0.02)
cumene	4.23	371	3.71 (0.01)	3.73 (0.01)	3.50 (0.01)	3.53 (0.01)	3.59 (0.01)	3.58 (0.01)	3.40 (0.01)	3.54 (0.01)	3.47 (0.01)	3.51 (0.01)	3.54 (0.01)
cyclohexane	2.96	335	2.57 (0.02)	2.54 (0.02)	2.34 (0.02)	2.33 (0.02)	2.71 (0.02)	2.57 (0.02)	2.51 (0.02)	2.48 (0.02)	2.90 (0.02)	2.87 (0.02)	2.55 (0.02)
cyclohexanol	2.06	369	2.40 (0.01)	2.39 (0.01)	2.49 (0.01)	2.30 (0.02)	2.41 (0.01)	2.45 (0.01)	2.08 (0.01)	2.09 (0.02)	2.43 (0.01)	2.43 (0.01)	2.27 (0.01)
cyclohexanone	2.19	365	3.37 (0.02)	3.27 (0.01)	3.03 (0.02)	3.41 (0.02)	3.38 (0.02)	3.23 (0.02)	3.39 (0.02)	3.80 (0.02)	3.25 (0.02)	3.32 (0.02)	3.61 (0.02)
dibutyladipate	4.85	377	5.74 (0.04)	5.76 (0.03)	5.65 (0.03)	5.55 (0.03)	5.67 (0.03)	5.44 (0.03)	4.78 (0.02)	5.58 (0.04)	4.96 (0.02)	5.21 (0.03)	5.49 (0.03)
dibutylphthalate	5.33	367	5.60 (0.04)	5.41 (0.03)	6.23 (0.03)	5.29 (0.03)	4.61 (0.05)	5.83 (0.03)	4.92 (0.03)	5.03 (0.03)	4.97 (0.02)	5.16 (0.02)	6.05 (0.03)
dichloromethane	2.42	354	2.64 (0.02)	2.54 (0.02)	2.18 (0.02)	2.58 (0.02)	2.82 (0.02)	2.65 (0.02)	2.65 (0.02)	3.19 (0.02)	2.07 (0.02)	2.19 (0.02)	2.42 (0.02)
dicyclopentadiene	3.63	370	1.83 (0.04)	2.61 (0.04)	2.14 (0.04)	2.27 (0.04)	2.35 (0.04)	2.08 (0.04)	1.99 (0.04)	1.11 (0.05)	2.44 (0.04)	2.62 (0.04)	2.39 (0.03)
diethylamine	1.93	354	2.21 (0.01)	2.50 (0.01)	2.35 (0.01)	1.92 (0.01)	2.17 (0.01)	2.57 (0.01)	2.53 (0.01)	2.01 (0.01)	2.51 (0.01)	2.51 (0.01)	2.04 (0.01)
diethyleneglycol	0.15	353	1.89 (0.01)	1.67 (0.01)	1.71 (0.01)	1.81 (0.01)	1.80 (0.02)	2.09 (0.01)	1.88 (0.02)	1.90 (0.01)	1.57 (0.01)	1.37 (0.01)	1.44 (0.01)
diethylether	1.46	401	1.71 (0.01)	1.63 (0.01)	1.99 (0.01)	1.91 (0.01)	2.02 (0.01)	1.90 (0.01)	1.69 (0.01)	1.79 (0.01)	1.58 (0.01)	1.66 (0.01)	1.86 (0.01)
diethylmalonate	4.01	391	3.72 (0.01)	3.76 (0.02)	3.67 (0.02)	3.67 (0.01)	3.65 (0.02)	3.58 (0.01)	3.75 (0.02)	3.37 (0.02)	4.36 (0.02)	4.53 (0.02)	3.66 (0.02)
diethylsebacate	4.98	351	5.72 (0.04)	5.77 (0.04)	5.62 (0.03)	5.51 (0.04)	5.68 (0.04)	5.41 (0.03)	4.73 (0.02)	5.57 (0.04)	4.89 (0.03)	5.56 (0.03)	5.47 (0.03)
diethylsuccinate	3.09	376	4.51 (0.01)	4.43 (0.01)	4.07 (0.02)	4.56 (0.01)	4.37 (0.01)	4.22 (0.01)	4.55 (0.02)	4.17 (0.02)	5.18 (0.02)	4.29 (0.02)	4.88 (0.02)
diisobutylphthalate	5.49	376	4.59 (0.03)	5.01 (0.03)	5.73 (0.04)	4.94 (0.03)	4.38 (0.04)	5.85 (0.03)	5.84 (0.02)	5.67 (0.02)	5.51 (0.03)	5.32 (0.03)	6.13 (0.03)
diisopropylether	2.11	377	2.09 (0.01)	2.00 (0.01)	1.99 (0.01)	1.59 (0.01)	1.72 (0.01)	2.50 (0.01)	2.15 (0.01)	1.60 (0.01)	2.02 (0.01)	2.12 (0.01)	2.09 (0.01)
dimethoxymethane	1.04	389	1.22 (0.01)	1.49 (0.01)	1.04 (0.01)	1.52 (0.02)	1.61 (0.02)	0.79 (0.02)	1.45 (0.01)	1.36 (0.01)	0.44 (0.01)	0.84 (0.01)	1.21 (0.01)
dimethylphthalate	3.7	341	3.55 (0.02)	3.59 (0.02)	4.32 (0.01)	3.99 (0.02)	3.93 (0.02)	4.54 (0.02)	4.21 (0.01)	4.51 (0.02)	3.76 (0.01)	4.05 (0.01)	4.36 (0.02)
dinbutyletherdibutylether	3.61	361	3.50 (0.01)	3.48 (0.01)	3.81 (0.01)	3.78 (0.01)	3.74 (0.01)	3.64 (0.01)	3.80 (0.01)	4.03 (0.01)	3.25 (0.02)	3.27 (0.02)	3.58 (0.02)
di-n-hexylamine	5.38	358	5.37 (0.02)	5.87 (0.02)	5.33 (0.02)	4.80 (0.02)	5.21 (0.02)	6.39 (0.02)	5.75 (0.02)	5.41 (0.02)	6.10 (0.02)	6.00 (0.02)	5.06 (0.02)
di-n-pentylether	4.71	369	4.02 (0.01)	4.10 (0.01)	4.63 (0.01)	4.04 (0.01)	3.96 (0.01)	4.27 (0.01)	4.79 (0.01)	5.09 (0.01)	3.77 (0.02)	3.75 (0.02)	4.01 (0.02)
diphenylamine	4.65	386	4.24 (0.01)	4.04 (0.01)	4.15 (0.01)	4.42 (0.01)	4.36 (0.01)	3.96 (0.01)	4.40 (0.01)	4.83 (0.01)	5.10 (0.01)	4.93 (0.01)	4.35 (0.01)
diphenylether	4.63	371	3.96 (0.02)	4.72 (0.01)	4.12 (0.01)	4.04 (0.01)	3.77 (0.01)	4.27 (0.01)	4.28 (0.01)	4.28 (0.01)	4.19 (0.01)	4.35 (0.01)	4.45 (0.01)

diphenylphthalate	6.6	377	6.75 (0.03)	6.75 (0.03)	6.38 (0.04)	5.82 (0.03)	7.49 (0.04)	5.29 (0.03)	6.63 (0.05)	6.76 (0.04)	6.66 (0.04)	6.79 (0.03)	5.40 (0.03)
dlimonene	5.29	393	4.63 (0.01)	4.43 (0.01)	4.68 (0.02)	4.59 (0.02)	4.95 (0.02)	4.66 (0.01)	5.10 (0.02)	4.59 (0.02)	4.52 (0.02)	4.64 (0.02)	4.63 (0.02)
ethanol	0.52	381	1.01 (0.01)	0.95 (0.02)	1.26 (0.02)	1.06 (0.02)	0.99 (0.02)	1.04 (0.01)	0.90 (0.01)	0.89 (0.01)	0.92 (0.01)	0.94 (0.01)	0.99 (0.01)
ethyl-4-aminobenzoate	3.67	345	3.86 (0.01)	3.88 (0.01)	3.80 (0.01)	3.90 (0.01)	3.96 (0.01)	3.94 (0.01)	3.66 (0.01)	3.96 (0.01)	4.11 (0.01)	4.28 (0.01)	4.21 (0.01)
ethylacetate	2.58	375	2.27 (0.01)	2.22 (0.01)	2.34 (0.01)	2.41 (0.01)	2.44 (0.01)	2.32 (0.01)	2.31 (0.01)	2.20 (0.01)	2.39 (0.01)	2.51 (0.01)	2.27 (0.01)
ethylacrylate	4.6	342	3.21 (0.01)	3.16 (0.01)	3.36 (0.01)	3.24 (0.01)	3.33 (0.01)	3.40 (0.01)	3.37 (0.01)	3.21 (0.01)	3.31 (0.01)	3.54 (0.01)	3.07 (0.01)
ethylbenzene	3.59	358	3.30 (0.01)	3.32 (0.01)	3.39 (0.01)	3.22 (0.01)	3.27 (0.01)	3.33 (0.01)	3.32 (0.01)	3.32 (0.01)	3.43 (0.01)	3.51 (0.01)	3.45 (0.01)
ethylbenzoate	4.23	371	3.89 (0.01)	3.76 (0.01)	3.72 (0.01)	3.91 (0.01)	3.82 (0.01)	3.99 (0.01)	3.79 (0.01)	3.77 (0.01)	4.02 (0.01)	4.17 (0.01)	4.09 (0.01)
ethylenediamine	2.55	369	1.48 (0.01)	1.23 (0.01)	1.79 (0.01)	1.51 (0.01)	1.27 (0.01)	1.75 (0.01)	1.62 (0.01)	1.54 (0.01)	1.56 (0.01)	1.42 (0.01)	1.61 (0.01)
ethyleneglycol	0.04	354	1.22 (0.01)	1.25 (0.01)	1.09 (0.01)	1.24 (0.01)	1.16 (0.01)	1.29 (0.01)	1.14 (0.01)	1.09 (0.01)	1.09 (0.01)	1.08 (0.01)	1.14 (0.01)
ethylhexanoate	4.21	353	4.05 (0.01)	3.89 (0.01)	3.89 (0.01)	4.13 (0.01)	4.13 (0.01)	3.95 (0.01)	4.05 (0.01)	4.00 (0.01)	4.05 (0.01)	3.99 (0.01)	4.14 (0.01)
eugenol	3.84	362	4.19 (0.01)	4.38 (0.01)	4.42 (0.01)	4.38 (0.01)	4.62 (0.01)	4.52 (0.01)	4.40 (0.01)	3.96 (0.01)	4.39 (0.01)	4.83 (0.01)	4.38 (0.01)
glutaraldehyde	3.94	351	3.13 (0.02)	2.99 (0.02)	2.25 (0.02)	3.01 (0.01)	3.14 (0.01)	2.80 (0.02)	2.68 (0.02)	2.70 (0.02)	2.87 (0.02)	3.07 (0.02)	2.79 (0.02)
glyoxal	2.43	380	2.41 (0.02)	2.48 (0.02)	2.27 (0.02)	2.21 (0.02)	2.15 (0.02)	2.28 (0.02)	2.15 (0.02)	2.28 (0.02)	3.05 (0.02)	3.01 (0.02)	3.04 (0.02)
hexachloroethane	5.23	370	4.85 (0.03)	4.87 (0.03)	5.40 (0.03)	5.25 (0.03)	5.33 (0.02)	4.07 (0.02)	4.55 (0.02)	4.09 (0.04)	5.31 (0.04)	4.91 (0.04)	4.57 (0.02)
hexanal	3.66	393	3.63 (0.01)	3.61 (0.01)	3.34 (0.01)	3.56 (0.01)	3.69 (0.01)	3.50 (0.01)	3.33 (0.01)	3.47 (0.01)	3.76 (0.01)	3.74 (0.01)	3.60 (0.01)
hexyleneglycol	1.13	365	1.85 (0.01)	2.04 (0.01)	2.01 (0.01)	1.93 (0.01)	2.01 (0.01)	1.99 (0.01)	1.92 (0.02)	2.12 (0.01)	1.93 (0.01)	2.06 (0.01)	2.33 (0.01)
isoprene	2.95	357	3.51 (0.01)	3.34 (0.01)	3.61 (0.01)	3.47 (0.01)	3.42 (0.01)	3.78 (0.01)	4.01 (0.01)	3.63 (0.01)	3.45 (0.01)	3.48 (0.01)	3.32 (0.01)
isopropanolol	0.78	380	1.05 (0.01)	1.04 (0.01)	1.29 (0.01)	1.08 (0.01)	1.03 (0.01)	1.01 (0.01)	1.07 (0.01)	0.99 (0.01)	0.71 (0.01)	0.73 (0.01)	0.90 (0.01)
m-cresol	3.29	346	3.48 (0.01)	3.40 (0.01)	3.56 (0.01)	3.56 (0.01)	3.38 (0.01)	3.34 (0.01)	3.67 (0.01)	3.71 (0.01)	3.52 (0.01)	3.51 (0.01)	3.61 (0.01)
m-dichlorobenzene	4.27	352	4.51 (0.01)	4.50 (0.01)	4.54 (0.01)	4.65 (0.01)	4.48 (0.01)	4.64 (0.01)	4.67 (0.01)	4.60 (0.01)	4.52 (0.01)	4.55 (0.01)	4.60 (0.01)
methanol	0.05	367	1.19 (0.02)	1.38 (0.02)	1.18 (0.02)	1.21 (0.02)	1.40 (0.02)	0.99 (0.02)	0.89 (0.02)	0.88 (0.02)	0.94 (0.02)	0.88 (0.02)	0.96 (0.02)
methyl-4-chlorobenzoate	4.15	332	4.07 (0.01)	4.10 (0.01)	4.30 (0.01)	4.24 (0.01)	4.24 (0.01)	4.17 (0.01)	4.26 (0.01)	4.07 (0.01)	4.15 (0.01)	4.33 (0.01)	4.09 (0.01)
methyl-4-cyanobenzoate	3.54	352	4.20 (0.01)	4.13 (0.01)	4.33 (0.01)	4.45 (0.02)	4.29 (0.01)	4.22 (0.02)	4.31 (0.01)	4.23 (0.02)	4.18 (0.01)	4.38 (0.01)	4.43 (0.02)
methylacetate	2.27	376	1.96 (0.01)	2.00 (0.01)	2.04 (0.01)	2.08 (0.01)	2.09 (0.01)	1.91 (0.01)	2.19 (0.01)	2.13 (0.01)	2.03 (0.01)	2.13 (0.01)	1.95 (0.01)
methylmethacrylate	2.5	367	2.75 (0.01)	2.95 (0.01)	3.01 (0.01)	2.74 (0.01)	2.85 (0.01)	2.77 (0.01)	3.29 (0.01)	3.12 (0.01)	3.08 (0.01)	3.09 (0.01)	2.69 (0.01)
methyl-tert-butylether	2.12	368	1.89 (0.01)	1.77 (0.01)	1.96 (0.01)	1.86 (0.01)	1.81 (0.01)	2.28 (0.01)	1.75 (0.01)	1.55 (0.01)	2.05 (0.01)	1.99 (0.01)	1.82 (0.02)
m-xylene	3.82	367	3.48 (0.01)	3.50 (0.01)	3.64 (0.01)	3.45 (0.01)	3.38 (0.01)	3.66 (0.01)	3.68 (0.01)	3.72 (0.01)	3.35 (0.01)	3.32 (0.01)	3.48 (0.01)
n-aminoethylpiperazine	1.82	345	1.32 (0.02)	1.60 (0.02)	2.63 (0.03)	1.00 (0.02)	0.84 (0.02)	1.79 (0.02)	2.18 (0.02)	1.54 (0.02)	2.05 (0.02)	1.85 (0.02)	1.76 (0.02)
naphthalene	4.32	358	4.32 (0.01)	4.30 (0.01)	4.18 (0.01)	4.28 (0.01)	4.23 (0.01)	4.14 (0.01)	4.21 (0.01)	4.42 (0.01)	4.37 (0.01)	4.37 (0.01)	4.40 (0.01)

n-butylacetate	3.81	346	3.17 (0.01)	3.09 (0.01)	3.12 (0.01)	3.20 (0.01)	3.16 (0.01)	3.11 (0.01)	3.20 (0.01)	2.97 (0.01)	3.21 (0.01)	3.28 (0.01)	3.04 (0.01)
n-butylamine	2.44	362	2.30 (0.01)	2.24 (0.01)	2.25 (0.01)	2.23 (0.01)	2.26 (0.01)	2.17 (0.01)	2.24 (0.01)	2.15 (0.01)	2.21 (0.01)	2.23 (0.01)	2.13 (0.01)
n-butylnbutyrate	4.09	371	4.02 (0.01)	3.79 (0.01)	3.63 (0.01)	4.06 (0.01)	3.92 (0.01)	3.68 (0.01)	4.04 (0.01)	3.78 (0.01)	3.74 (0.01)	3.70 (0.01)	3.86 (0.01)
n-butylphenylether	4.42	385	4.18 (0.01)	4.35 (0.01)	4.33 (0.01)	3.94 (0.01)	4.06 (0.01)	4.36 (0.01)	4.03 (0.01)	4.11 (0.01)	4.30 (0.01)	4.18 (0.01)	4.13 (0.01)
n-decyclamine	5.18	405	4.99 (0.01)	4.82 (0.01)	4.93 (0.01)	4.86 (0.01)	4.97 (0.01)	4.81 (0.01)	4.83 (0.01)	4.81 (0.01)	5.09 (0.01)	5.12 (0.01)	4.79 (0.01)
n-dodecylamine	6.26	377	5.73 (0.01)	5.49 (0.01)	5.55 (0.01)	5.58 (0.01)	5.53 (0.01)	5.43 (0.01)	5.31 (0.01)	5.34 (0.01)	5.82 (0.01)	5.81 (0.01)	5.41 (0.01)
n-ethyl-m-toluidine	3.44	381	3.45 (0.01)	3.63 (0.01)	3.21 (0.01)	3.60 (0.01)	3.55 (0.01)	3.55 (0.01)	3.75 (0.01)	3.74 (0.01)	3.85 (0.01)	3.84 (0.01)	4.04 (0.01)
n-heptylamine	3.72	390	3.98 (0.01)	3.87 (0.01)	3.89 (0.01)	3.92 (0.01)	3.97 (0.01)	3.89 (0.01)	4.05 (0.01)	3.91 (0.01)	3.93 (0.01)	3.95 (0.01)	3.97 (0.01)
n-hexanoicacid	2.76	361	3.66 (0.01)	3.69 (0.01)	3.58 (0.01)	3.69 (0.01)	3.59 (0.01)	3.60 (0.01)	3.55 (0.01)	3.59 (0.01)	3.56 (0.01)	3.47 (0.01)	3.56 (0.01)
n-hexylacetate	4.56	402	4.17 (0.01)	4.12 (0.01)	4.07 (0.01)	4.10 (0.01)	4.05 (0.01)	4.20 (0.01)	4.61 (0.01)	4.26 (0.01)	4.07 (0.01)	4.07 (0.01)	4.18 (0.01)
n-hexylamine	3.25	335	3.34 (0.01)	3.33 (0.01)	3.29 (0.01)	3.23 (0.01)	3.33 (0.01)	3.26 (0.01)	3.34 (0.01)	3.23 (0.01)	3.31 (0.01)	3.36 (0.01)	3.22 (0.01)
n-methylaniline	3.03	360	2.84 (0.01)	2.72 (0.01)	2.94 (0.01)	2.96 (0.01)	2.78 (0.01)	2.77 (0.01)	3.02 (0.01)	2.71 (0.01)	2.97 (0.01)	2.92 (0.01)	2.98 (0.01)
n,n-dibutylformamide	3.25	386	4.01 (0.02)	4.26 (0.02)	3.60 (0.02)	4.04 (0.02)	4.12 (0.02)	4.19 (0.02)	4.02 (0.01)	4.46 (0.02)	4.03 (0.02)	4.22 (0.02)	4.83 (0.02)
n,n-diethylaniline	3.96	337	4.30 (0.01)	4.25 (0.02)	3.54 (0.01)	4.32 (0.01)	4.06 (0.01)	4.28 (0.02)	3.84 (0.01)	3.66 (0.01)	3.79 (0.02)	3.98 (0.02)	4.26 (0.02)
n,n-diethylcyclohexylamine	3.86	360	3.51 (0.02)	3.70 (0.02)	3.21 (0.02)	3.30 (0.02)	3.67 (0.02)	3.79 (0.02)	2.89 (0.02)	3.49 (0.02)	3.58 (0.02)	3.50 (0.02)	2.88 (0.02)
n,n-diethylethanolamine	1.82	372	2.70 (0.01)	2.99 (0.02)	2.38 (0.01)	2.50 (0.01)	2.39 (0.02)	2.65 (0.01)	2.30 (0.01)	2.32 (0.02)	1.98 (0.01)	2.09 (0.01)	2.05 (0.01)
n-propylacetate	3.23	373	2.95 (0.01)	2.87 (0.01)	2.90 (0.01)	2.98 (0.01)	2.99 (0.01)	2.93 (0.01)	2.91 (0.01)	2.70 (0.01)	3.00 (0.01)	3.13 (0.01)	2.71 (0.01)
n-propylamine	2.28	358	2.05 (0.01)	2.06 (0.01)	1.97 (0.01)	1.95 (0.01)	2.04 (0.01)	1.98 (0.01)	1.93 (0.01)	1.98 (0.01)	2.06 (0.01)	2.04 (0.01)	1.87 (0.01)
n-tridecylamine	6.45	383	6.32 (0.02)	6.03 (0.02)	6.06 (0.02)	6.15 (0.02)	5.99 (0.02)	5.96 (0.02)	5.70 (0.02)	5.77 (0.02)	6.34 (0.02)	6.31 (0.02)	5.88 (0.02)
o-chloroaniline	4.35	406	3.48 (0.01)	3.94 (0.01)	3.37 (0.01)	3.38 (0.01)	3.90 (0.01)	3.43 (0.01)	3.51 (0.01)	3.37 (0.01)	3.98 (0.01)	3.99 (0.01)	3.44 (0.01)
o-chlorophenol	4	354	4.18 (0.01)	4.13 (0.01)	4.16 (0.01)	4.18 (0.01)	4.07 (0.01)	4.18 (0.01)	4.21 (0.01)	4.28 (0.01)	3.96 (0.01)	3.91 (0.01)	4.13 (0.01)
o-cresol	3.9	362	3.30 (0.01)	3.22 (0.01)	3.27 (0.01)	3.27 (0.01)	3.22 (0.01)	3.13 (0.01)	3.20 (0.01)	3.24 (0.01)	3.38 (0.01)	3.23 (0.01)	3.14 (0.01)
o-dichlorobenzene	4.19	388	4.70 (0.01)	4.56 (0.01)	4.74 (0.01)	4.61 (0.01)	4.57 (0.01)	4.67 (0.01)	4.76 (0.01)	4.75 (0.01)	4.53 (0.01)	4.52 (0.01)	4.55 (0.01)
o-phenylphenol	4.5	366	5.03 (0.01)	5.00 (0.01)	4.25 (0.01)	4.98 (0.01)	4.85 (0.01)	5.15 (0.01)	4.70 (0.01)	4.44 (0.01)	4.26 (0.01)	4.14 (0.01)	4.88 (0.01)
o-xylene	3.81	366	3.85 (0.01)	3.85 (0.01)	3.75 (0.01)	3.73 (0.01)	3.80 (0.01)	3.88 (0.01)	3.75 (0.01)	3.50 (0.01)	3.79 (0.01)	3.76 (0.01)	4.04 (0.01)
p-chloroaniline	3.62	355	3.74 (0.01)	4.26 (0.01)	3.69 (0.01)	3.59 (0.01)	4.13 (0.01)	3.68 (0.01)	3.73 (0.01)	3.75 (0.01)	4.07 (0.01)	4.12 (0.01)	3.74 (0.01)
p-chloro-m-cresol	4.4	366	4.29 (0.01)	4.06 (0.01)	4.17 (0.01)	4.29 (0.01)	4.13 (0.01)	4.36 (0.01)	4.34 (0.01)	4.24 (0.01)	4.11 (0.01)	4.13 (0.01)	4.26 (0.01)
p-chlorophenol	4.46	378	4.16 (0.01)	4.00 (0.01)	4.27 (0.01)	4.07 (0.01)	3.96 (0.01)	4.15 (0.01)	4.17 (0.01)	4.17 (0.01)	3.96 (0.01)	4.05 (0.01)	4.03 (0.01)
p-cresol	3.76	380	3.23 (0.01)	3.23 (0.01)	3.29 (0.01)	3.29 (0.01)	3.26 (0.01)	3.10 (0.01)	3.42 (0.01)	3.34 (0.01)	3.34 (0.01)	3.35 (0.01)	3.34 (0.01)
p-dichlorobenzene	4.27	387	4.55 (0.01)	4.77 (0.01)	4.74 (0.01)	4.58 (0.01)	4.73 (0.01)	4.67 (0.01)	4.69 (0.01)	4.52 (0.01)	4.49 (0.01)	4.52 (0.01)	4.47 (0.01)

pentachloroethane	4.43	370	4.54 (0.01)	4.56 (0.02)	4.70 (0.01)	4.66 (0.02)	4.43 (0.01)	4.63 (0.02)	4.45 (0.02)	4.14 (0.02)	4.76 (0.02)	4.66 (0.02)	4.43 (0.02)
pentachlorophenol	6.04	381	6.13 (0.01)	6.49 (0.01)	6.33 (0.01)	6.00 (0.02)	6.06 (0.01)	6.39 (0.01)	6.24 (0.01)	6.23 (0.02)	6.39 (0.02)	6.54 (0.01)	6.46 (0.01)
pentachloropyridine	5.73	377	6.25 (0.02)	6.02 (0.01)	5.33 (0.02)	6.22 (0.02)	6.06 (0.02)	6.02 (0.02)	6.26 (0.01)	6.04 (0.01)	5.93 (0.02)	6.00 (0.02)	5.95 (0.02)
pentafluoroaniline	3.69	368	4.36 (0.02)	4.08 (0.03)	4.15 (0.02)	3.99 (0.03)	3.92 (0.03)	4.23 (0.02)	4.32 (0.02)	4.30 (0.02)	4.12 (0.03)	4.08 (0.03)	4.50 (0.03)
pentafluorobenzaldehyde	5.25	366	4.41 (0.02)	4.02 (0.02)	5.00 (0.02)	4.20 (0.02)	3.71 (0.02)	4.76 (0.02)	4.99 (0.02)	4.42 (0.02)	3.85 (0.02)	3.88 (0.02)	4.07 (0.02)
pentanal	3.82	341	2.80 (0.01)	2.85 (0.01)	2.52 (0.01)	2.73 (0.01)	2.83 (0.01)	2.66 (0.01)	2.41 (0.01)	2.55 (0.01)	2.95 (0.01)	2.98 (0.01)	2.66 (0.01)
p-ethylphenol	4.07	364	3.74 (0.01)	3.64 (0.01)	3.73 (0.01)	3.64 (0.01)	3.64 (0.01)	3.72 (0.01)	3.64 (0.01)	3.42 (0.01)	3.74 (0.01)	3.86 (0.01)	3.75 (0.01)
phenol	3.5	348	3.50 (0.01)	3.37 (0.01)	3.54 (0.01)	3.50 (0.01)	3.38 (0.01)	3.37 (0.01)	3.41 (0.01)	3.45 (0.01)	3.25 (0.01)	3.23 (0.01)	3.40 (0.01)
phenylsalicylate	5.27	364	5.11 (0.02)	4.59 (0.01)	5.35 (0.01)	5.08 (0.02)	4.97 (0.02)	5.02 (0.01)	4.59 (0.01)	5.14 (0.02)	4.87 (0.01)	5.06 (0.02)	4.82 (0.02)
phthalazine	3.11	324	4.58 (0.01)	4.17 (0.01)	4.09 (0.01)	4.76 (0.01)	4.60 (0.01)	4.27 (0.01)	4.32 (0.01)	4.04 (0.01)	4.55 (0.01)	4.51 (0.01)	4.27 (0.01)
propionitrile	1.56	378	2.69 (0.02)	2.51 (0.01)	2.03 (0.01)	2.50 (0.02)	2.40 (0.02)	2.56 (0.02)	1.95 (0.02)	1.44 (0.01)	2.29 (0.01)	2.29 (0.01)	2.03 (0.01)
p-t-butylstyrene	3.51	371	4.78 (0.01)	4.81 (0.01)	5.09 (0.01)	4.80 (0.01)	4.54 (0.01)	4.86 (0.01)	5.15 (0.01)	5.28 (0.01)	4.85 (0.01)	4.99 (0.01)	4.89 (0.01)
p-tert-amylphenol	4.8	354	4.41 (0.01)	4.29 (0.01)	4.05 (0.01)	4.42 (0.01)	4.34 (0.01)	4.41 (0.01)	4.56 (0.01)	4.35 (0.02)	4.22 (0.01)	4.25 (0.01)	4.39 (0.01)
p-tert-butylphenol	4.47	353	4.17 (0.01)	4.09 (0.01)	4.05 (0.02)	4.01 (0.01)	3.83 (0.01)	4.17 (0.01)	4.04 (0.02)	4.21 (0.02)	3.88 (0.02)	3.97 (0.02)	3.86 (0.02)
p-xylene	4.21	361	3.85 (0.01)	3.91 (0.01)	3.98 (0.01)	3.66 (0.01)	3.69 (0.01)	4.01 (0.01)	3.91 (0.01)	3.83 (0.01)	3.67 (0.01)	3.66 (0.01)	3.71 (0.01)
pyridine	2.9	348	2.96 (0.01)	2.91 (0.01)	2.87 (0.01)	2.90 (0.01)	2.96 (0.01)	2.98 (0.01)	2.79 (0.01)	2.83 (0.01)	2.73 (0.01)	2.72 (0.01)	2.78 (0.01)
quinoline	3.45	377	4.15 (0.01)	3.97 (0.01)	3.73 (0.01)	4.01 (0.01)	3.91 (0.01)	4.17 (0.01)	3.70 (0.01)	4.04 (0.01)	3.84 (0.01)	3.78 (0.01)	3.85 (0.01)
secbutylamine	2.42	356	2.17 (0.01)	2.04 (0.01)	2.14 (0.01)	2.14 (0.01)	2.05 (0.01)	2.20 (0.01)	2.28 (0.01)	2.26 (0.01)	2.07 (0.01)	2.05 (0.01)	2.33 (0.01)
styrene	3.51	363	3.60 (0.01)	3.71 (0.01)	4.11 (0.01)	3.46 (0.01)	3.35 (0.01)	3.70 (0.01)	3.92 (0.01)	3.55 (0.01)	3.35 (0.01)	3.77 (0.01)	3.52 (0.01)
tert-butylacetate	2.55	329	2.15 (0.01)	2.46 (0.01)	2.55 (0.01)	2.23 (0.01)	2.18 (0.01)	2.60 (0.01)	2.68 (0.01)	2.28 (0.01)	2.60 (0.01)	2.59 (0.01)	2.55 (0.01)
tetrachloroethylene	3.91	389	3.97 (0.02)	3.78 (0.02)	3.97 (0.02)	3.71 (0.02)	3.55 (0.02)	3.70 (0.02)	3.67 (0.02)	3.21 (0.02)	4.39 (0.02)	4.24 (0.02)	4.91 (0.03)
tetrachlorophenol	6.13	367	5.47 (0.01)	5.60 (0.01)	5.45 (0.01)	5.54 (0.01)	5.41 (0.01)	5.52 (0.01)	5.53 (0.01)	5.55 (0.01)	5.44 (0.01)	5.52 (0.01)	5.40 (0.01)
tetrahydrofuran	1.52	362	0.56 (0.02)	0.68 (0.02)	1.47 (0.02)	0.58 (0.02)	0.65 (0.02)	1.00 (0.02)	1.40 (0.02)	1.81 (0.02)	0.62 (0.02)	0.45 (0.01)	0.49 (0.02)
thymol	4.67	381	3.72 (0.01)	3.88 (0.01)	4.07 (0.01)	3.82 (0.01)	3.93 (0.01)	3.93 (0.01)	3.58 (0.01)	3.92 (0.01)	4.14 (0.01)	4.13 (0.01)	3.89 (0.01)
toluene	3.42	386	3.25 (0.01)	3.27 (0.01)	3.27 (0.01)	3.17 (0.01)	3.20 (0.01)	3.27 (0.01)	3.33 (0.01)	3.27 (0.01)	3.16 (0.01)	3.09 (0.01)	3.24 (0.01)
trichloroethylene	3.47	362	3.37 (0.02)	3.43 (0.01)	3.67 (0.02)	3.42 (0.02)	3.23 (0.02)	3.27 (0.02)	3.31 (0.02)	2.96 (0.02)	3.68 (0.02)	3.66 (0.02)	4.17 (0.02)
trioxanestrioxane	1.18	357	1.19 (0.02)	0.36 (0.02)	0.30 (0.02)	1.69 (0.03)	1.10 (0.02)	0.19 (0.02)	1.22 (0.02)	1.78 (0.02)	1.58 (0.02)	1.28 (0.02)	1.71 (0.02)
tripropylamine	3.45	354	2.34 (0.02)	2.00 (0.02)	3.28 (0.03)	3.30 (0.02)	3.20 (0.02)	2.42 (0.03)	2.67 (0.02)	2.83 (0.03)	2.64 (0.02)	2.49 (0.02)	2.20 (0.03)

Aromatic compounds

These data are all taken from Agatonovic-Kustrin et al.'s [22] article. The property used in the prediction is the degrees of penetration of a polydimethylsiloxane membrane. The number of compounds was reduced from the original 254 down to 80 to focus on the compounds which had more than one rotational bond.

Table A2: The compound names and their degree of penetration of a polydimethylsiloxane membrane [22].

Name	log J	#	B3LYP			HF			PM3	AM1	MM3*		MMFF94s
			6-31G(d,p)	6-31G	STO-3G	6-31G(d,p)	6-31G	STO-3G			ε = 1	ε = 80	
1,3,5-Triethylbenzene	-1.08	351	-1.16 (0.03)	-0.85 (0.03)	-1.63 (0.04)	-1.33 (0.04)	-1.09 (0.04)	-1.20 (0.04)	-1.47 (0.04)	-1.45 (0.04)	-1.84 (0.04)	-1.95 (0.04)	-0.63 (0.04)
1,3-Diisopropylbenzene	-1.06	384	-0.65 (0.04)	-0.96 (0.04)	-1.03 (0.03)	-0.97 (0.03)	-0.96 (0.04)	-0.87 (0.04)	-0.84 (0.03)	-0.45 (0.03)	-0.70 (0.03)	-0.93 (0.03)	-0.75 (0.04)
1-Butylbenzene	-0.75	393	-1.10 (0.01)	-0.99 (0.01)	-1.14 (0.01)	-1.28 (0.01)	-1.08 (0.01)	-1.16 (0.01)	-1.09 (0.01)	-1.48 (0.01)	-1.01 (0.01)	-0.98 (0.01)	-1.03 (0.01)
1-Ethoxynaphthalene	-2.79	378	-1.75 (0.02)	-1.70 (0.01)	-1.46 (0.01)	-1.82 (0.01)	-1.71 (0.01)	-1.76 (0.02)	-1.72 (0.02)	-1.87 (0.02)	-2.04 (0.02)	-1.96 (0.02)	-1.85 (0.02)
1-Isoquinolinecarboxylicacid	-4.13	348	-4.05 (0.02)	-4.08 (0.02)	-4.19 (0.02)	-4.37 (0.02)	-4.09 (0.02)	-4.02 (0.02)	-4.73 (0.02)	-4.02 (0.02)	-4.15 (0.02)	-4.22 (0.02)	-3.81 (0.02)
1-Methyl-2-phenoxyethylamine	-1.63	359	-1.90 (0.01)	-1.61 (0.01)	-1.66 (0.01)	-2.17 (0.01)	-1.91 (0.01)	-1.73 (0.01)	-2.02 (0.01)	-1.87 (0.01)	-2.15 (0.02)	-2.27 (0.01)	-1.78 (0.01)
1-Naphthoicacid	-2.98	360	-3.83 (0.02)	-3.76 (0.01)	-3.83 (0.01)	-3.38 (0.02)	-3.68 (0.01)	-3.82 (0.01)	-3.96 (0.01)	-3.99 (0.01)	-3.98 (0.02)	-3.96 (0.02)	-3.88 (0.02)
1-Phenyl-2-propanol	-2.02	340	-1.29 (0.01)	-1.44 (0.01)	-1.42 (0.01)	-1.32 (0.01)	-1.30 (0.01)	-1.34 (0.01)	-1.39 (0.01)	-1.43 (0.01)	-1.22 (0.01)	-1.34 (0.01)	-1.40 (0.01)
2,5-Pyridinedicarboxylicacid	-5.21	367	-5.02 (0.02)	-5.03 (0.02)	-5.20 (0.02)	-5.16 (0.02)	-5.27 (0.02)	-5.01 (0.02)	-5.22 (0.02)	-5.24 (0.02)	-5.26 (0.02)	-5.25 (0.02)	-5.20 (0.02)
2-Aminobenzylalcohol	-2.63	375	-2.57 (0.02)	-2.58 (0.02)	-2.77 (0.02)	-2.20 (0.02)	-2.58 (0.02)	-2.68 (0.02)	-2.39 (0.02)	-2.34 (0.02)	-2.59 (0.02)	-2.48 (0.02)	-2.41 (0.02)
2-Anisaldehyde	-2.03	364	-1.82 (0.01)	-1.66 (0.01)	-1.54 (0.01)	-1.87 (0.01)	-1.79 (0.01)	-1.81 (0.01)	-1.88 (0.01)	-1.77 (0.01)	-1.98 (0.01)	-2.01 (0.01)	-1.84 (0.01)
2-Butoxypyridine	-1.16	371	-1.33 (0.01)	-1.53 (0.01)	-1.59 (0.02)	-1.30 (0.01)	-1.38 (0.01)	-1.25 (0.01)	-1.22 (0.01)	-1.26 (0.01)	-1.47 (0.02)	-1.42 (0.02)	-1.57 (0.02)
2-Chlorophenoxyaceticacid	-2.93	344	-3.96 (0.03)	-3.71 (0.02)	-4.28 (0.02)	-4.07 (0.03)	-4.25 (0.03)	-3.94 (0.03)	-3.50 (0.03)	-3.68 (0.03)	-3.60 (0.03)	-2.76 (0.03)	-3.99 (0.03)
2-Fluorobenzoicacid	-2.29	380	-2.65 (0.02)	-2.75 (0.01)	-2.86 (0.01)	-2.67 (0.02)	-2.68 (0.02)	-2.74 (0.01)	-3.09 (0.01)	-2.96 (0.01)	-2.99 (0.01)	-3.15 (0.01)	-2.63 (0.02)
2-Fluoropropiophenone	-1.44	372	-1.83 (0.01)	-1.74 (0.01)	-1.69 (0.01)	-1.75 (0.01)	-1.74 (0.01)	-1.63 (0.01)	-1.75 (0.02)	-1.78 (0.02)	-1.38 (0.01)	-1.43 (0.02)	-1.60 (0.01)

2-Furoicacid	-2.48	369	-2.26 (0.02)	-2.40 (0.02)	-2.69 (0.02)	-2.25 (0.02)	-2.38 (0.02)	-2.61 (0.02)	-2.36 (0.02)	-2.26 (0.03)	-2.32 (0.02)	-2.20 (0.02)	-2.30 (0.02)
2-Hydroxyacetophenone	-1.78	389	-2.21 (0.02)	-2.15 (0.02)	-1.64 (0.01)	-2.13 (0.02)	-2.24 (0.02)	-2.08 (0.02)	-1.81 (0.02)	-2.04 (0.01)	-2.13 (0.02)	-2.10 (0.02)	-2.18 (0.02)
2-Methyl-5-butylpyridine	-1.11	372	-1.07 (0.02)	-1.10 (0.02)	-1.14 (0.01)	-1.19 (0.02)	-1.13 (0.02)	-0.99 (0.02)	-1.41 (0.02)	-1.32 (0.02)	-1.33 (0.02)	-1.26 (0.02)	-1.33 (0.02)
2-Naphtalenaceticacid	-3.57	353	-3.82 (0.02)	-3.51 (0.02)	-3.69 (0.02)	-3.89 (0.02)	-3.45 (0.02)	-3.69 (0.02)	-3.45 (0.02)	-3.59 (0.02)	-3.63 (0.02)	-3.60 (0.02)	-3.64 (0.02)
2-Phenylpropionaldehyde	-1.69	371	-1.16 (0.01)	-1.42 (0.01)	-1.35 (0.01)	-1.26 (0.01)	-1.48 (0.01)	-1.29 (0.01)	-1.60 (0.01)	-1.50 (0.01)	-1.56 (0.01)	-1.63 (0.01)	-1.19 (0.01)
2-Pyrazinecarboxylicacid	-4.07	361	-3.00 (0.02)	-2.96 (0.01)	-3.02 (0.01)	-2.79 (0.02)	-2.78 (0.02)	-3.14 (0.02)	-2.95 (0.01)	-2.96 (0.01)	-3.01 (0.01)	-3.17 (0.02)	-2.94 (0.02)
2-Quinolinecarboxylicacid	-3.55	371	-4.27 (0.01)	-4.13 (0.01)	-4.17 (0.01)	-4.24 (0.01)	-4.32 (0.01)	-4.25 (0.01)	-4.28 (0.01)	-4.59 (0.01)	-4.38 (0.01)	-4.45 (0.01)	-4.18 (0.01)
2-Thiophenaceticacid	-2.48	367	-2.24 (0.02)	-2.38 (0.02)	-2.32 (0.01)	-2.21 (0.02)	-2.33 (0.02)	-1.72 (0.02)	-1.98 (0.02)	-2.36 (0.01)	-2.52 (0.02)	-2.27 (0.02)	-2.00 (0.02)
2-Thiophenmethanol	-2.18	369	-1.62 (0.02)	-2.11 (0.02)	-1.81 (0.01)	-1.79 (0.02)	-1.91 (0.02)	-2.06 (0.02)	-2.03 (0.02)	-1.99 (0.02)	-1.35 (0.02)	-1.85 (0.01)	-1.76 (0.02)
2-Thiophenmethylamine	-1.4	350	-1.59 (0.02)	-1.49 (0.01)	-1.59 (0.02)	-1.32 (0.02)	-1.24 (0.02)	-1.56 (0.02)	-1.53 (0.02)	-1.44 (0.02)	-1.61 (0.02)	-1.69 (0.01)	-1.64 (0.02)
3-Aminobenzoicacid	-3.73	346	-3.20 (0.01)	-2.97 (0.01)	-3.17 (0.01)	-3.31 (0.01)	-3.09 (0.01)	-3.21 (0.01)	-3.10 (0.01)	-3.30 (0.01)	-3.07 (0.01)	-2.98 (0.01)	-3.26 (0.01)
3-Anisaldehyde	-2.09	363	-2.11 (0.01)	-2.07 (0.01)	-1.97 (0.01)	-2.07 (0.01)	-2.04 (0.01)	-2.01 (0.01)	-1.99 (0.01)	-2.01 (0.01)	-2.02 (0.01)	-2.04 (0.01)	-2.00 (0.01)
3-Anisicacid	-2.58	377	-2.46 (0.02)	-2.42 (0.02)	-2.60 (0.02)	-2.51 (0.02)	-2.49 (0.02)	-2.43 (0.01)	-2.34 (0.01)	-2.38 (0.01)	-2.47 (0.02)	-2.48 (0.02)	-2.43 (0.02)
3-Chlorobenzoicacid	-2.37	368	-3.14 (0.01)	-2.96 (0.01)	-2.87 (0.01)	-3.18 (0.01)	-2.94 (0.02)	-3.00 (0.02)	-2.97 (0.01)	-3.13 (0.01)	-3.08 (0.02)	-3.13 (0.01)	-3.16 (0.02)
3-Hydroxy-4-methoxybenzoicacid	-4.37	370	-3.24 (0.01)	-3.47 (0.01)	-3.50 (0.01)	-3.14 (0.01)	-3.23 (0.01)	-3.36 (0.01)	-3.26 (0.01)	-3.32 (0.01)	-3.26 (0.01)	-3.33 (0.01)	-3.16 (0.01)
3-Hydroxybenzoicacid	-3.31	339	-3.25 (0.01)	-3.33 (0.01)	-3.16 (0.01)	-3.21 (0.01)	-3.33 (0.01)	-3.23 (0.01)	-3.22 (0.01)	-3.16 (0.01)	-3.34 (0.01)	-3.35 (0.01)	-3.34 (0.01)
3-Methoxyacetophenone	-1.99	342	-2.72 (0.01)	-2.69 (0.01)	-2.87 (0.01)	-2.78 (0.01)	-2.76 (0.01)	-2.69 (0.01)	-2.59 (0.01)	-2.64 (0.01)	-2.73 (0.01)	-2.74 (0.01)	-2.68 (0.01)
3-Methoxybenzylacetate	-2.13	386	-2.62 (0.02)	-2.64 (0.02)	-2.11 (0.02)	-2.58 (0.02)	-2.77 (0.02)	-2.55 (0.02)	-2.59 (0.02)	-2.18 (0.02)	-2.55 (0.02)	-2.63 (0.02)	-2.52 (0.02)
3-Phenoxytoluene	-2.01	369	-2.15 (0.02)	-2.44 (0.02)	-2.42 (0.02)	-1.59 (0.03)	-2.18 (0.02)	-1.87 (0.02)	-2.02 (0.03)	-1.84 (0.02)	-2.50 (0.02)	-2.45 (0.02)	-2.32 (0.02)
3-Phenyl-1-propanol	-2.32	375	-1.77 (0.01)	-2.12 (0.01)	-1.87 (0.01)	-1.85 (0.01)	-1.99 (0.01)	-1.72 (0.01)	-1.71 (0.01)	-1.73 (0.01)	-1.81 (0.01)	-1.89 (0.01)	-1.85 (0.01)
3-Phenyl-1-propylamine	-1.46	363	-1.70 (0.01)	-1.55 (0.01)	-1.58 (0.01)	-1.74 (0.01)	-1.62 (0.01)	-1.76 (0.02)	-1.58 (0.02)	-1.67 (0.02)	-1.63 (0.02)	-1.66 (0.02)	-1.73 (0.01)
3-Phenylbutyraldehyde	-1.96	381	-1.72 (0.02)	-1.61 (0.01)	-1.34 (0.01)	-1.95 (0.02)	-1.78 (0.01)	-1.78 (0.02)	-1.86 (0.02)	-1.76 (0.02)	-1.62 (0.02)	-1.69 (0.02)	-1.64 (0.02)
3-Quinolinecarboxylicacid	-4.41	377	-3.84 (0.02)	-4.09 (0.01)	-4.09 (0.01)	-3.63 (0.02)	-3.92 (0.02)	-3.89 (0.01)	-3.95 (0.01)	-3.88 (0.02)	-4.04 (0.01)	-4.09 (0.01)	-3.93 (0.01)
3-Thiopheneaceticacid	-2.41	381	-2.55 (0.02)	-2.61 (0.02)	-2.27 (0.02)	-2.72 (0.02)	-2.85 (0.02)	-2.61 (0.02)	-2.06 (0.02)	-2.39 (0.02)	-2.26 (0.02)	-2.10 (0.02)	-2.36 (0.02)
3-Toluicacid	-2.31	363	-2.27 (0.01)	-2.24 (0.01)	-2.49 (0.01)	-2.59 (0.01)	-2.53 (0.01)	-2.46 (0.01)	-2.30 (0.01)	-2.40 (0.01)	-2.51 (0.01)	-2.51 (0.01)	-2.30 (0.01)
4-Acetoxybenzoicacid	-3.11	352	-3.92 (0.02)	-3.52 (0.02)	-3.78 (0.01)	-4.11 (0.02)	-3.48 (0.02)	-4.17 (0.02)	-4.00 (0.02)	-4.16 (0.02)	-3.98 (0.02)	-4.15 (0.02)	-4.04 (0.02)
4-Aminobenzoicacid	-3.49	351	-3.53 (0.02)	-3.53 (0.02)	-3.40 (0.02)	-3.48 (0.02)	-3.52 (0.01)	-3.54 (0.01)	-3.40 (0.01)	-3.77 (0.01)	-3.51 (0.01)	-3.50 (0.01)	-3.80 (0.01)
4-Anisaldehyde	-2.07	381	-2.43 (0.01)	-2.32 (0.01)	-2.23 (0.01)	-2.42 (0.01)	-2.25 (0.01)	-2.38 (0.01)	-2.38 (0.01)	-2.32 (0.01)	-2.25 (0.01)	-2.28 (0.01)	-2.37 (0.01)
4-Anisicacid	-3.23	340	-2.79 (0.01)	-2.90 (0.01)	-2.91 (0.01)	-2.79 (0.02)	-2.85 (0.01)	-2.82 (0.01)	-2.96 (0.01)	-2.96 (0.01)	-2.93 (0.01)	-2.93 (0.01)	-2.91 (0.01)
4-Carboxybenzaldehyde	-3.44	360	-4.03 (0.01)	-3.96 (0.01)	-3.86 (0.01)	-4.08 (0.01)	-4.00 (0.01)	-3.98 (0.01)	-4.02 (0.01)	-3.86 (0.01)	-3.85 (0.01)	-3.78 (0.01)	-3.92 (0.01)

4-Chlorobenzoicacid	-3.09	363	-2.89 (0.02)	-2.94 (0.01)	-2.81 (0.01)	-2.93 (0.02)	-2.80 (0.02)	-2.70 (0.02)	-2.60 (0.02)	-2.75 (0.02)	-2.84 (0.02)	-2.84 (0.02)	-2.83 (0.02)
4-Hydroxybenzoicacid	-3.53	367	-3.69 (0.01)	-3.94 (0.01)	-3.61 (0.01)	-3.56 (0.01)	-3.77 (0.01)	-3.70 (0.01)	-3.77 (0.01)	-3.69 (0.01)	-3.81 (0.01)	-3.79 (0.01)	-3.74 (0.01)
4-Isopropylbenzaldehyde	-1.64	365	-1.93 (0.01)	-1.92 (0.01)	-2.16 (0.01)	-2.11 (0.01)	-2.06 (0.01)	-2.00 (0.01)	-2.04 (0.01)	-1.90 (0.01)	-1.73 (0.01)	-1.73 (0.01)	-2.03 (0.01)
4-Methoxy-2-quinolinicacid	-4.62	347	-3.98 (0.02)	-3.87 (0.02)	-3.97 (0.02)	-3.61 (0.02)	-3.89 (0.02)	-3.85 (0.02)	-3.64 (0.02)	-3.86 (0.02)	-3.76 (0.02)	-3.78 (0.02)	-3.60 (0.01)
4Quinolinecarboxylicacid	-4.52	368	-3.36 (0.02)	-3.70 (0.01)	-3.62 (0.02)	-3.91 (0.01)	-3.76 (0.01)	-3.52 (0.02)	-4.02 (0.02)	-3.81 (0.02)	-3.53 (0.02)	-3.64 (0.02)	-3.57 (0.02)
4-t-Butylbenzoicacid	-2.76	353	-2.91 (0.04)	-2.56 (0.03)	-2.73 (0.03)	-2.62 (0.03)	-2.48 (0.03)	-2.73 (0.03)	-2.73 (0.03)	-2.47 (0.04)	-2.63 (0.03)	-2.58 (0.03)	-2.79 (0.04)
6-Chloronicotinicacid	-3.1	383	-3.17 (0.02)	-3.31 (0.02)	-3.33 (0.02)	-2.94 (0.02)	-3.28 (0.02)	-3.46 (0.02)	-3.16 (0.02)	-3.07 (0.02)	-3.24 (0.02)	-3.34 (0.02)	-3.13 (0.02)
6-Methoxyquinoline	-2.1	330	-2.83 (0.02)	-2.85 (0.02)	-2.84 (0.01)	-2.85 (0.02)	-2.94 (0.02)	-2.92 (0.02)	-2.91 (0.02)	-2.87 (0.02)	-2.81 (0.02)	-2.85 (0.02)	-2.81 (0.02)
6-Quinolinecarboxylicacid	-4.67	365	-4.16 (0.02)	-4.24 (0.01)	-4.06 (0.01)	-3.93 (0.01)	-4.16 (0.01)	-4.24 (0.01)	-4.06 (0.01)	-3.97 (0.01)	-4.06 (0.01)	-3.96 (0.01)	-4.18 (0.01)
8-Quinolinecarboxylicacid	-4.21	367	-4.18 (0.02)	-4.15 (0.03)	-4.36 (0.02)	-4.50 (0.02)	-4.22 (0.03)	-4.35 (0.02)	-3.85 (0.02)	-4.15 (0.02)	-3.67 (0.01)	-3.63 (0.01)	-4.41 (0.02)
Benzoicacid	-2.32	376	-2.36 (0.01)	-2.48 (0.01)	-2.47 (0.01)	-2.37 (0.01)	-2.46 (0.01)	-2.32 (0.01)	-2.42 (0.01)	-2.37 (0.01)	-2.44 (0.01)	-2.36 (0.01)	-2.30 (0.01)
Butylphenylether	-1.25	371	-1.00 (0.01)	-1.17 (0.01)	-1.17 (0.01)	-1.06 (0.01)	-1.26 (0.01)	-1.05 (0.01)	-1.12 (0.01)	-1.10 (0.01)	-1.24 (0.01)	-1.33 (0.01)	-1.05 (0.01)
Butyrophenone	-1.72	378	-1.35 (0.02)	-1.25 (0.01)	-1.35 (0.01)	-1.09 (0.02)	-1.28 (0.02)	-1.37 (0.02)	-1.05 (0.01)	-1.16 (0.02)	-1.28 (0.01)	-1.09 (0.01)	-1.12 (0.02)
Dibenzyl	-1.98	345	-2.15 (0.02)	-2.48 (0.02)	-2.52 (0.02)	-2.06 (0.03)	-2.08 (0.03)	-1.83 (0.02)	-1.84 (0.03)	-1.89 (0.03)	-2.08 (0.03)	-1.98 (0.03)	-2.08 (0.02)
Ethyl-2-methylbenzoat	-1.48	380	-1.02 (0.02)	-1.17 (0.02)	-1.59 (0.01)	-0.97 (0.01)	-1.04 (0.01)	-1.17 (0.01)	-1.17 (0.02)	-1.18 (0.01)	-1.21 (0.02)	-1.18 (0.01)	-1.37 (0.01)
Ethylcinnamate	-1.95	346	-1.45 (0.02)	-1.40 (0.02)	-1.31 (0.02)	-1.52 (0.02)	-1.58 (0.02)	-1.64 (0.02)	-1.38 (0.02)	-1.13 (0.02)	-1.31 (0.02)	-1.18 (0.02)	-1.51 (0.02)
Ethylnicotinate	-1.53	357	-2.07 (0.01)	-2.02 (0.01)	-2.30 (0.01)	-2.01 (0.02)	-1.98 (0.02)	-2.20 (0.01)	-2.20 (0.02)	-2.17 (0.01)	-1.87 (0.01)	-1.84 (0.01)	-1.91 (0.01)
Ethylsalicylate	-1.61	347	-1.87 (0.02)	-1.93 (0.02)	-1.51 (0.01)	-1.93 (0.02)	-1.88 (0.02)	-1.76 (0.02)	-1.83 (0.02)	-1.75 (0.02)	-1.74 (0.02)	-1.80 (0.02)	-1.80 (0.01)
Furfurylalcohol	-1.86	374	-2.58 (0.02)	-2.33 (0.02)	-2.46 (0.02)	-2.73 (0.02)	-2.39 (0.02)	-2.61 (0.02)	-2.46 (0.02)	-2.86 (0.02)	-2.24 (0.02)	-2.05 (0.02)	-2.33 (0.02)
Isophthalicacid	-3.99	363	-4.45 (0.01)	-4.59 (0.01)	-4.41 (0.01)	-4.26 (0.02)	-4.52 (0.02)	-4.56 (0.01)	-4.53 (0.02)	-4.64 (0.01)	-4.50 (0.02)	-4.47 (0.02)	-4.53 (0.02)
Methoxymethylphenylsulphide	-1.68	410	-1.30 (0.02)	-1.23 (0.02)	-1.37 (0.02)	-1.01 (0.02)	-1.06 (0.02)	-1.55 (0.02)	-1.59 (0.02)	-0.93 (0.02)	-0.79 (0.02)	-1.06 (0.02)	-1.02 (0.02)
Methyl-2-methoxybenzoate	-2.19	391	-2.16 (0.02)	-2.50 (0.02)	-1.82 (0.01)	-1.70 (0.02)	-2.10 (0.02)	-1.73 (0.01)	-1.72 (0.02)	-1.98 (0.01)	-1.92 (0.01)	-1.80 (0.01)	-2.02 (0.01)
Methyl-4-t-butylbenzoate	-1.71	372	-2.69 (0.02)	-2.49 (0.02)	-2.34 (0.02)	-2.61 (0.02)	-2.63 (0.02)	-2.56 (0.02)	-2.45 (0.02)	-2.76 (0.02)	-2.65 (0.02)	-2.51 (0.02)	-2.70 (0.03)
Methylbenzoate	-1.46	363	-1.38 (0.01)	-1.24 (0.01)	-1.55 (0.01)	-1.42 (0.01)	-1.34 (0.01)	-1.37 (0.01)	-1.43 (0.01)	-1.46 (0.01)	-1.37 (0.01)	-1.31 (0.01)	-1.26 (0.01)
Methylparaben	-2.74	376	-2.61 (0.01)	-2.67 (0.01)	-2.58 (0.01)	-2.48 (0.01)	-2.48 (0.01)	-2.63 (0.01)	-2.69 (0.01)	-2.60 (0.01)	-2.65 (0.01)	-2.60 (0.01)	-2.59 (0.01)
Methylsalicylate	-1.67	363	-2.05 (0.01)	-2.07 (0.01)	-1.99 (0.01)	-2.18 (0.01)	-2.11 (0.01)	-2.04 (0.01)	-1.98 (0.01)	-1.94 (0.01)	-2.14 (0.01)	-2.16 (0.01)	-2.13 (0.01)
Nicotinicacid	-3.76	360	-3.24 (0.02)	-2.87 (0.02)	-3.75 (0.02)	-3.32 (0.02)	-2.78 (0.02)	-3.04 (0.02)	-3.60 (0.02)	-2.87 (0.02)	-3.05 (0.02)	-3.11 (0.02)	-3.35 (0.02)
Phenethylamine	-1.26	369	-1.82 (0.02)	-1.52 (0.01)	-1.65 (0.01)	-1.50 (0.01)	-1.33 (0.01)	-1.72 (0.01)	-1.58 (0.01)	-1.55 (0.01)	-1.49 (0.01)	-1.29 (0.01)	-1.64 (0.01)
Phenetole	-1.11	371	-0.85 (0.01)	-1.11 (0.02)	-1.09 (0.02)	-0.99 (0.01)	-1.32 (0.02)	-0.85 (0.01)	-1.00 (0.02)	-0.96 (0.01)	-1.04 (0.01)	-0.95 (0.01)	-0.90 (0.01)
Phenoxyaceticacid	-2.46	367	-2.58 (0.01)	-2.96 (0.01)	-2.58 (0.01)	-2.66 (0.01)	-2.85 (0.01)	-2.57 (0.01)	-2.79 (0.01)	-2.43 (0.01)	-2.91 (0.01)	-2.84 (0.01)	-2.97 (0.01)

Phenylacetate	-1.65	366	-1.76 (0.02)	-1.04 (0.01)	-1.41 (0.01)	-1.45 (0.02)	-1.00 (0.01)	-1.67 (0.02)	-1.53 (0.03)	-1.80 (0.02)	-1.48 (0.02)	-1.64 (0.03)	-1.81 (0.02)
Picolinicacid	-3.28	377	-3.11 (0.02)	-3.11 (0.01)	-3.15 (0.02)	-3.09 (0.02)	-3.19 (0.02)	-3.07 (0.01)	-3.13 (0.02)	-3.07 (0.02)	-3.16 (0.01)	-3.12 (0.01)	-3.17 (0.02)
Propylparaben	-2.72	395	-2.41 (0.01)	-2.51 (0.01)	-2.38 (0.01)	-2.32 (0.01)	-2.27 (0.01)	-2.46 (0.01)	-2.48 (0.01)	-2.40 (0.01)	-2.44 (0.01)	-2.62 (0.01)	-2.40 (0.01)
Salicylicacid	-2.57	373	-3.04 (0.01)	-3.13 (0.01)	-2.89 (0.01)	-3.14 (0.01)	-3.11 (0.01)	-3.02 (0.01)	-3.00 (0.01)	-2.97 (0.01)	-3.20 (0.01)	-3.27 (0.01)	-3.07 (0.01)
Terephthalicacid	-5.14	355	-4.29 (0.01)	-4.39 (0.01)	-4.46 (0.01)	-4.41 (0.01)	-4.41 (0.01)	-4.32 (0.01)	-4.40 (0.01)	-4.45 (0.01)	-4.35 (0.01)	-4.34 (0.01)	-4.29 (0.01)

PPARγ agonists

These compounds and their reference values are all taken from Rücker et al.'s [23] article. The property used in the prediction is pK_i for PPARγ. The numbers given in the table are the ones used by Rücker et al. in their article. The total number of compounds has been reduced to 12 to focus on a sub-group of all the compounds with a similar molecular skeleton.

Table A3: The compound number according to Rücerk et al. [23], and their pK_i for PPARγ.

No.	pk_i for PPARγ	#	B3LYP			HF			PM3	AM1	MM3*		MMFF94s
			6-31G(d,p)	6-31G	STO-3G	6-31G(d,p)	6-31G	STO-3G			ε = 1	ε = 80	
78	8.68	366	9.29 (0.02)	9.09 (0.03)	9.02 (0.02)	9.20 (0.02)	9.15 (0.02)	9.40 (0.03)	9.26 (0.02)	9.38 (0.03)	9.79 (0.03)	9.88 (0.03)	9.43 (0.03)
87	8.74	325	8.19 (0.02)	7.64 (0.02)	7.78 (0.02)	8.08 (0.02)	8.55 (0.02)	8.31 (0.02)	8.32 (0.02)	8.42 (0.02)	8.38 (0.03)	8.26 (0.02)	8.01 (0.02)
93	8.62	347	8.13 (0.02)	8.84 (0.03)	8.53 (0.03)	8.47 (0.02)	8.58 (0.02)	8.38 (0.03)	8.40 (0.02)	8.45 (0.02)	8.32 (0.03)	8.43 (0.03)	8.27 (0.03)
94	9.01	344	8.35 (0.02)	8.23 (0.01)	8.28 (0.02)	8.33 (0.02)	8.26 (0.02)	8.18 (0.02)	8.17 (0.02)	7.93 (0.01)	8.28 (0.02)	8.29 (0.02)	8.22 (0.02)
149	5.68	358	5.63 (0.01)	5.56 (0.01)	5.41 (0.01)	5.49 (0.02)	5.78 (0.02)	5.44 (0.02)	5.49 (0.01)	5.46 (0.01)	5.84 (0.02)	5.72 (0.01)	5.74 (0.01)
151	5.59	354	6.26 (0.01)	6.20 (0.01)	6.34 (0.01)	6.33 (0.02)	6.34 (0.01)	6.57 (0.01)	6.24 (0.01)	6.37 (0.01)	6.50 (0.01)	6.47 (0.01)	6.37 (0.02)
153	6.55	357	6.95 (0.01)	7.00 (0.01)	6.84 (0.01)	6.88 (0.01)	6.65 (0.01)	6.66 (0.01)	7.23 (0.01)	6.91 (0.01)	6.89 (0.01)	6.99 (0.01)	7.09 (0.01)

154	6.32	348	6.87 (0.01)	6.86 (0.01)	6.91 (0.01)	6.94 (0.01)	6.81 (0.01)	6.91 (0.01)	6.92 (0.01)	6.89 (0.01)	6.80 (0.01)	6.79 (0.01)	6.87 (0.01)
155	6.8	351	6.67 (0.01)	6.53 (0.01)	6.52 (0.01)	6.60 (0.01)	6.51 (0.01)	6.53 (0.01)	6.59 (0.01)	6.46 (0.01)	6.42 (0.01)	6.50 (0.01)	6.56 (0.01)
156	6.07	346	5.74 (0.01)	5.81 (0.01)	5.86 (0.01)	5.75 (0.01)	5.80 (0.01)	5.95 (0.01)	5.74 (0.01)	6.00 (0.01)	5.73 (0.01)	5.79 (0.01)	5.73 (0.01)
157	6.44	363	6.37 (0.01)	6.30 (0.01)	6.40 (0.01)	6.41 (0.01)	6.53 (0.01)	6.44 (0.01)	6.29 (0.01)	6.59 (0.01)	6.44 (0.01)	6.33 (0.01)	6.35 (0.01)
158	6.01	363	6.66 (0.01)	6.72 (0.01)	6.96 (0.01)	6.66 (0.01)	7.05 (0.02)	7.01 (0.01)	6.86 (0.01)	6.93 (0.01)	6.91 (0.01)	6.83 (0.01)	6.83 (0.01)

Supplementary Material – Part II

The following tables are the results from the Games-Howell test which was performed on the bootstrap samples between the 11 different energy evaluation methods. This means that 1000 bootstrap drawings for each energy evaluation have been compared, with the p-value corrected by the Bonferroni correction. The Games-Howell test is a pairwise test investigating if the q^2-values derived from two energy evaluation methods are significantly different or not. Table A4 shows the results for the toxicological compounds, Table A5 for the aromatic compounds and Table A6 for the PPARγ agonists.

Table A4: Results from the Games-Howell test on the bootstrap samples for the toxicological data. (***: $p < 0.001$, **: $0.001 < p < 0.01$, *: $0.01 < p < 0.05$, (*): $0.05 < p < 0.1$)

	B3LYP/6-31G	B3LYP/STO-3G	HF/6-31G(d,p)	HF/6-31G	HF/STO-3G	PM3	AM1	MM3*-1	MM3*-80	MMFF94s
B3LYP/6-31G(d,p)	(*)	-	(*)	-	***	***	***	-	-	***
B3LYP/6-31G		*	-	-	-	***	***	**	***	***
B3LYP/STO-3G	*		*	-	***	***	***	-	-	***
HF/6-31G(d,p)	-	*		-	-	***	***	*	***	***
HF/6-31G	-	-	-		-	***	***	-	*	***
HF/STO-3G	-	***	-	-		***	**	***	***	-
PM3	***	***	***	***	***		-	***	***	***
AM1	***	***	***	***	**	-		***	***	-
MM3*-1	**	-	*	-	***	***	***		-	***
MM3*-80	***	-	***	**	***	***	***	-		***
MMFF94s	***	***	***	***	-	***	-	***	***	

Table A5: Results from the Games-Howell test on the bootstrap samples for the aromatic data. (***: $p < 0.001$, **: $0.001 < p < 0.01$, *: $0.01 < p < 0.05$, (*): $0.05 < p < 0.1$)

	B3LYP/6-31G	B3LYP/STO-3G	HF/6-31G(d,p)	HF/6-31G	HF/STO-3G	PM3	AM1	MM3*-1	MM3*-80	MMFF94s
B3LYP/6-31G(d,p)	***	***	***	-	-	-	*	-	-	-
B3LYP/6-31G		-	***	***	***	***	***	***	***	***
B3LYP/STO-3G	-		***	***	***	***	***	***	***	***
HF/6-31G(d,p)	***	***		***	***	***	-	***	***	(*)
HF/6-31G	***	***	***		-	-	*	-	-	-
HF/STO-3G	***	***	***	-		-	***	-	*	***
PM3	***	***	***	-	-		-	-	-	-
AM1	***	***	-	*	***	-		***	-	-
MM3*-1	***	***	***	-	-	-	***		*	***
MM3*-80	***	***	***	-	*	-	-	*		-
MMFF94s	***	***	(*)	-	***	-	-	***	-	

Table A6: Results from the Games-Howell test on the bootstrap samples for the PPARγ agonists. (***: $p < 0.001$, **: $0.001 < p < 0.01$, *: $0.01 < p < 0.05$, (*): $0.05 < p < 0.1$)

	B3LYP/6-31G	B3LYP/STO-3G	HF/6-31G(d,p)	HF/6-31G	HF/STO-3G	PM3	AM1	MM3*-1	MM3*-80	MMFF94s
B3LYP/6-31G(d,p)	***	***	-	-	***	***	***	***	***	***
B3LYP/6-31G		-	***	***	*	-	-	***	***	-
B3LYP/STO-3G	-		***	***	-	-	-	***	***	-
HF/6-31G(d,p)	***	***		-	***	***	***	***	***	***
HF/6-31G	***	***	-		***	***	***	***	***	***
HF/STO-3G	*	-	***	***		-	-	-	-	-
PM3	-	-	***	***	-		-	-	-	-
AM1	-	-	***	***	-	-		-	***	-
MM3*-1	***	***	***	***	-	-	-		-	-
MM3*-80	***	***	***	***	-	-	***	-		-
MMFF94s	-	-	***	***	-	-	-	-	-	

Supplementary Material – Part III

The total calculation times for the 11 different energy evaluation methods are given in Table A7.

Table A7: Calculating times for the 11 different energy evaluation methods for the three datasets given in days (d), hours (h), minutes (m) and seconds (s).

Method	Toxicological	Aromatic	PPARγ agonists
B3LYP/6-31G(d,p) [1)]	20d 5h 37m 44.7s	4d 11h 12m 2.1s	58d 5h 47m 4.9s
B3LYP/6-31G [1)]	6d 1h 31m 36.2s	1d 11h 55m 4.1s	19d 23h 34m 3.0s
B3LYP/STO-3G [1)]	3d 12h 26m 23.9s	0d 23h 13m 54.5s	8d 14h 59m 45.1s
HF/6-31G(d,p) [1)]	9d 22h 33m 20.4s	2d 20h 2h 1.6s	27d 3h 39m 7.8s
HF/6-31G [1)]	1d 13h 19m 46.3s	0d 8h 53m 4.7s	6d 12h 6m 0.1s
HF/STO-3G [1)]	0d 10h 3m 41.3s	0d 3h 36m 31.4s	1d 14h 8m 2.0s
PM3 [1)]	0d 3h 50m 49.0s	0d 1h 35m 42.1s	0d 1h 5m 55.1s
AM1 [1)]	0d 3h 32m 56.9s	0d 1h 31m 12.5s	0d 1h 12m 1.5s
MM3*-1 [2)]	0d 0h 0m 9.7s	0d 0h 0m 4.6s	0d 0h 0m 26.7s
MM3*-80 [2)]	0d 0h 0m 11.5s	0d 0h 0m 3.3s	0d 0h 0m 29.8s
MMFF94s [2)]	0d 0h 0m 12.3s	0d 0h 0m 4.3s	0d 0h 0m 31.2s

1. Calculated using Gaussian 98 on a P4/ 3.2GHz computer running under Ubuntu 6.10 Edgy Eft.
2. Calculated using Macromodel 9.11 on a P4/ 3.8GHz computer running under Ubuntu 5.10 Breezy Badger.

Paper V

Niels H. Andersen, Niels Johan Christensen, Peter Rygaard Lassen, Teresa B. Freedman, Laurence A. Nafie, Kristian Strømgaard, and Lars Hemmingsen. *Structure and absolute configuration of ginkgolide B characterized by IR- and VCD spectroscopy*. Chirality, 2009. Published (web).

Structure and Absolute Configuration of Ginkgolide B Characterized by IR- and VCD Spectroscopy

NIELS H. ANDERSEN,[1] NIELS JOHAN CHRISTENSEN,[2] PETER R. LASSEN,[1] TERESA B.N. FREEDMAN,[3] LAURENCE A. NAFIE,[3] KRISTIAN STRØMGAARD,[4] AND LARS HEMMINGSEN[5]*

[1]*Department of Physics, Quantum Protein Center (QuP), Technical University of Denmark, Kgs. Lyngby, Denmark*
[2]*Department of Food Science, Quality and Technology, Faculty of Life Sciences, University of Copenhagen, Frederiksberg, Denmark*
[3]*Department of Chemistry, Syracuse University, New York*
[4]*Department of Medicinal Chemistry, Faculty of Pharmaceutical Sciences, University of Copenhagen, Copenhagen, Denmark*
[5]*Department of Basic Sciences and Environment, Faculty of Life Sciences, University of Copenhagen, Frederiksberg, Denmark*

ABSTRACT Experimental and calculated (B3LYP/6-31G(d)) vibrational circular dichroism (VCD) and IR spectra are compared, illustrating that the structure and absolute configuration of ginkgolide B (GB) may be characterized directly in solution. A conformational search for GB using MacroModel and subsequent DFT optimizations (B3LYP/6-31G(d)) provides a structure for the lowest energy conformer which agrees well with the structure determined by X-ray diffraction. In addition, a conformer at an energy of 7 kJ mol^{-1} (B3LYP/6-311+G(2d,2p)) with respect to the lowest energy conformer is predicted, displaying different intramolecular hydrogen bonding. Differences between measured and calculated IR and VCD spectra for GB at certain wavenumbers are rationalized in terms of interactions with solvent, intermolecular GB-GB interactions, and the potential presence of more than one conformer. This is the first detailed investigation of the spectroscopic fingerprint region (850−1300 cm^{-1}) of the natural product GB employing infrared absorption and VCD spectroscopy. *Chirality 00:000–000, 2009.*

KEY WORDS: ginkgolides; vibrational circular dichroism; DFT calculations; conformational analysis; fingerprint region

INTRODUCTION

The interest in the medical effects of the Ginkgo tree has persisted since ancient time in Eastern Asia. A standardized extract from the Ginkgo tree (Egb761) contains ginkgolides GA, GB, GC, GJ, and GM. It appears that GM is only found in the root bark and GJ only in the leaves whereas the other ginkgolides can be isolated from all parts of the tree.[1] The ginkgo tree itself is remarkable in being a specimen that has survived since the Jurassic era.[1,2] In 1932 ginkgolides were discovered as the bitter chemicals produced by *Ginkgo biloba*, and later structure elucidation revealed that these compounds are complex terpene trilactones. Discovery of their ability to interact with the platelet activating factor (PAF) receptor in 1985 markedly increased the attention on ginkgolides. Among the ginkgolides, ginkgolide B (GB) is found to be the most potent PAF receptor antagonist.[1] Recently, a new target for ginkgolides was discovered, as it was found that they bind to and block glycine receptors in the brain.[3,4] The overall IR spectral features of ginkgolides comprise characteristics of alcohols, ethers, esters and aliphatic hydrocarbons, which are all strongly entangled due to sharing of many atoms in the ring, see Figure 1.[5]

The fingerprint spectral region (850–1300 cm^{-1}) is therefore highly complex, and many spectral bands should rather be assigned to the vibrational character of the whole molecular frame than to specific groups. Addition or removal of OH— than to specific groups or inversions of chiral centers are therefore expected to have a considerable effect on both IR and vibrational circular dichroism (VCD) spectra. The number of chiral centers for GA, GB, GC, GJ, and GM are 10, 11, 12, 11, and 12, respectively. However, three of these (C_4, C_5, and C_9) cannot contribute to the number of possible diastereomers without rupture of the 5-ring structure, and four (C_2, C_3, C_6, and C_{12}) will severely alter the overall shape of the molecular frame. Inversion of C_8 or C_{14} implies change of the *t*-butyl or methyl group orientation. Thus, we have limited our investigations of GB only to include diastereomers with inversions at C_1 and C_{10} (denoted GB-C_1i and GB-C_{10}i, respectively). These isomers are relevant for total synthesis of

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: The Lundbeck Foundation

*Correspondence to: Lars Hemmingsen, Department of Basic Sciences and Environment (IGM), Faculty of Life Sciences, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark.
E-mail: lhe@life.ku.dk

Received for publication 27 November 2008; Accepted 25 February 2009
DOI: 10.1002/chir.20730
Published online in Wiley InterScience (www.interscience.wiley.com).

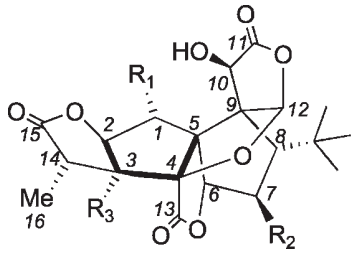


Fig. 1. Molecular structure of the ginkgolides GA, GB, GC, GJ, and GM.

GB where stereospecific addition to double bonds at these two carbons occurs.[1,6] GB ($C_{20}H_{24}O_{10}$) that has 224 electrons and no symmetry is among the larger molecules studied by this combination of experimental and theoretical methods, albeit similar work on even larger molecular systems have been published.[7–9] In recent work by He et al.[10] the interaction between ginkgolides and the amyloid peptide Aβ(25,35) was investigated, with a particular focus on the carbonyl stretching vibrations, i.e. the 1795–1826 cm^{-1} part of the infrared spectrum, and how these vibrations may be affected by such intermolecular interactions. A recent review on VCD spectroscopy refers to the few VCD studies of ginkgolides in the literature (including this work).[11]

In this work we focus on the application of IR and VCD spectroscopy in the fingerprint region combined with conformational searching and DFT calculations of these spectroscopic properties, as a case study to establish if IR and VCD spectroscopy may be applied to determine the absolute configuration of synthetic or naturally occurring ginkgolides directly in solution, and to elucidate potential GB-solvent interactions, intermolecular GB-GB interactions or the presence of more than one conformer.

MATERIALS AND METHODS

IR and VCD Spectroscopy

Ginkgolides are very soluble in highly polar and hydrogen bonding solvents such as water/acetone mixtures, methanol or DMSO but they are insoluble in low dielectric constant solvents as CCl_4 and $CHCl_3$.[2] DMSO-d_6, CD_3CN and KBr were chosen as media for sample preparation. DMSO-d_6 dissolves GB most efficiently, evaporates slowly but suffers from a limited transparency range (1100–2100 cm^{-1}). CD_3CN has a much better transparency range (500–2500 cm^{-1}) but dissolves less GB and evaporates faster while KBr pellets might cause artifacts in VCD spectra due to mechanical stress that is difficult to reproduce. In all cases care should be taken to avoid water because of hygroscopicity, and the risk of lactone ring opening, which occurs at pH above 6.[12] We used two different instruments, both of the FT-type with PEM modulation of the IR light. The instrument at the Quantum Protein Centre, Copenhagen, Denmark, is a Thermo Nicolet Nexus 870 infrared spectrometer equipped with a single PEM and improved with a filter that only allowed light in the range between 1000 and 1750 cm^{-1} to pass. Thus the C = O stretch region was not recorded and in Figures 2 and 3 this region is not presented for the experiment with GB dissolved in DMSO-d_6 recorded under these circumstances. The second instrument is a modified Chiral/IR FT-VCD instrument at Syracuse University, New York. Here, an extra PEM is added after the sample, eliminating most of the linear birefringence (LB) artifacts that often occur. As a consequence this instrument produces a much flatter baseline but background spectra still have to be recorded.[13] GB was isolated and purified as earlier described.[1] 3.9 mg GB (424.4 g mol^{-1}) was dissolved in 20 μl DMSO-d_6 and placed in a 48 μm CaF_2 sandwich-cell in the Thermo Nicolet instrument with the PEM set to 1500 cm^{-1}. Assuming volume-additivity of solvent and GB (ρ = 1.377 g cm^{-3})[14] the concentration was c = 400 mM. Sample and solvent were scanned for 57 h and the latter spectrum was subtracted. The cell was solvent-tight for several days. A calibration of the instrument was performed immediately after scanning of the sample. The efficiency of the PEM is almost independent of the frequency in the range 1000–1750 cm^{-1} and the raw VCD data were scaled by multiplication with the factor 0.565. This factor was derived from instrument calibration using α-pinene. Two measurements were carried out with the modified ChiralIR instrument in Syracuse, New York. 1.8 mg GB was dissolved in 50 μl CD_3CN and transferred to a 54 μm BaF_2 cell. The concentration was c = 85 mM. Ca. 0.5 mg GB was recovered from the CD_3CN experiment and was ground with 250 mg KBr (ρ = 2.75 g cm^{-3}, CRC) and pressed to a 13 mm diameter disk for 4 min at 13,000 lbs. The GB concentration and disk path-length were calculated to be c = 13 mM and l = 0.69 mm, respectively. In both cases the PEMs were set to 1400 cm^{-1}. The software corrected the VCD spectra automatically using previous calibration measurements.

Theoretical Methods—DFT Calculations and Conformational Search

All DFT calculations were carried out with Gaussian 03.[15] Conformational analysis with MacroModel[16] was performed on all ginkgolides, including the two stereoisomers of GB, GBC_1i, and $GBC_{10}i$. The input files for MacroModel were produced from published crystal structures of GA, GB, and GC and modifications of these.[17,18] The conformational analysis was carried out with the Monte Carlo Multiple Minimum method, employing the MMFFs forcefield. Conformers within an energy window of 50 kJ mol^{-1} were saved. All conformers with relative energies within 20 kJ mol^{-1} of the lowest energy conformer for each ginkgolide type were geometry optimized at the B3LYP/6-31G(d) level of theory. At this level of theory, the two lowest energy conformers of GA, GJ, GBC_1i and $GBC_{10}i$ were separated by at least 17 kJ mol^{-1}. For GB, GC, and GM the two lowest energy conformers were found within approximately 8 kJ mol^{-1}, see Table 1.

For all subsequent calculations, only the first conformer was used for GA, GJ, GBC_1i, and $GBC_{10}i$ whereas the two lowest energy conformers were included for GB, GC, and GM. The optimized structures for the conformers are

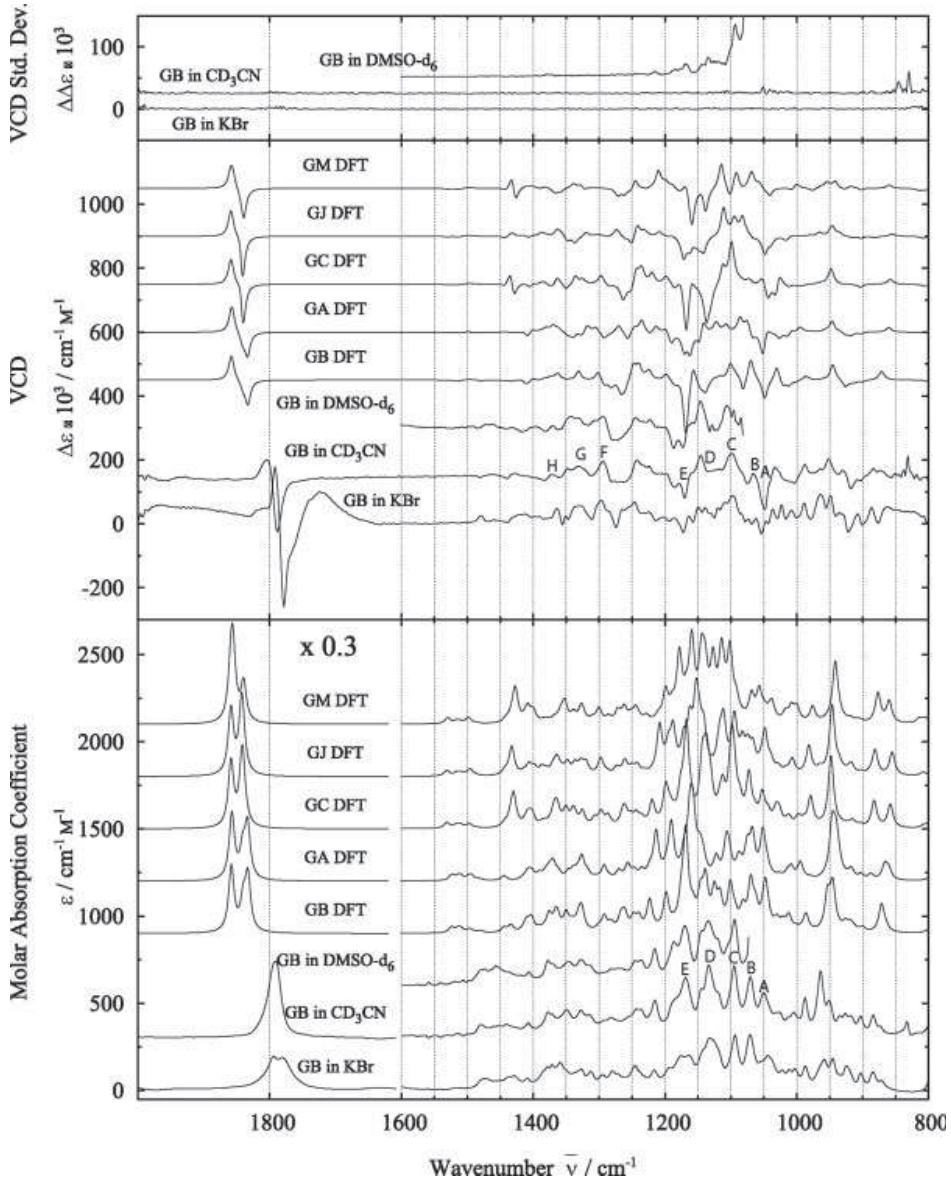


Fig. 2. Comparison of experimental IR and VCD spectra for GB with calculated (B3LYP/6-31G(d)) spectra for GB, GA, GC, GJ and GM. The wavenumbers of the calculated transitions at B3LYP/6-31G(d) level are scaled by a factor 0.982. The intensity of the IR bands in the range between 1600 and 2000 cm^{-1} is multiplied by a factor 0.3. The upper panel illustrates VCD-noise, the middle panel VCD and the lower panel IR absorption. Prominent peaks for GB are indicated with the letters A, B, C, D, E, F, G, and H.

included as supporting information for this article. The VCD and IR spectra for all relevant conformers were calculated at the B3LYP/6-31G(d) level of theory. The adequacy of B3LYP/6-31G(d) for the optimization of geometry and the calculation of IR and VCD spectra has been demonstrated previously.[19] For graphical presentation of the VCD and IR spectra Lorentzian band profiles of 5 cm^{-1} bandwidth (HWHM) were applied. The calculated wavenumbers were multiplied with a scaling factor of 0.982. This scaling factor deviates from the one used in an earlier study.[14] However, the spectra presented in this work are of a higher resolution, and the chosen scaling factor ensured the best fit to the observed data in the fingerprint region. To achieve better accuracy of the relative energies, a single point energy calculation was undertaken for each geometry optimized conformer with the larger basis set 6-311+G(2d,2p).[20] The relative energies of the conformers are summarized in Table 1. The predicted VCD and IR spectra for GB, GC, and GM presented in Figures 2 and 3 are composite spectra produced by addition of the calculated spectra for each conformer weighted by a Boltzmann factor based on Gibbs free energy. The Gibbs free energies were calculated by addition of the electronic energies from the higher quality single point calculations to the thermal correction to energy for the corresponding 6-31G(d) frequency calculations. The DFT calculations in

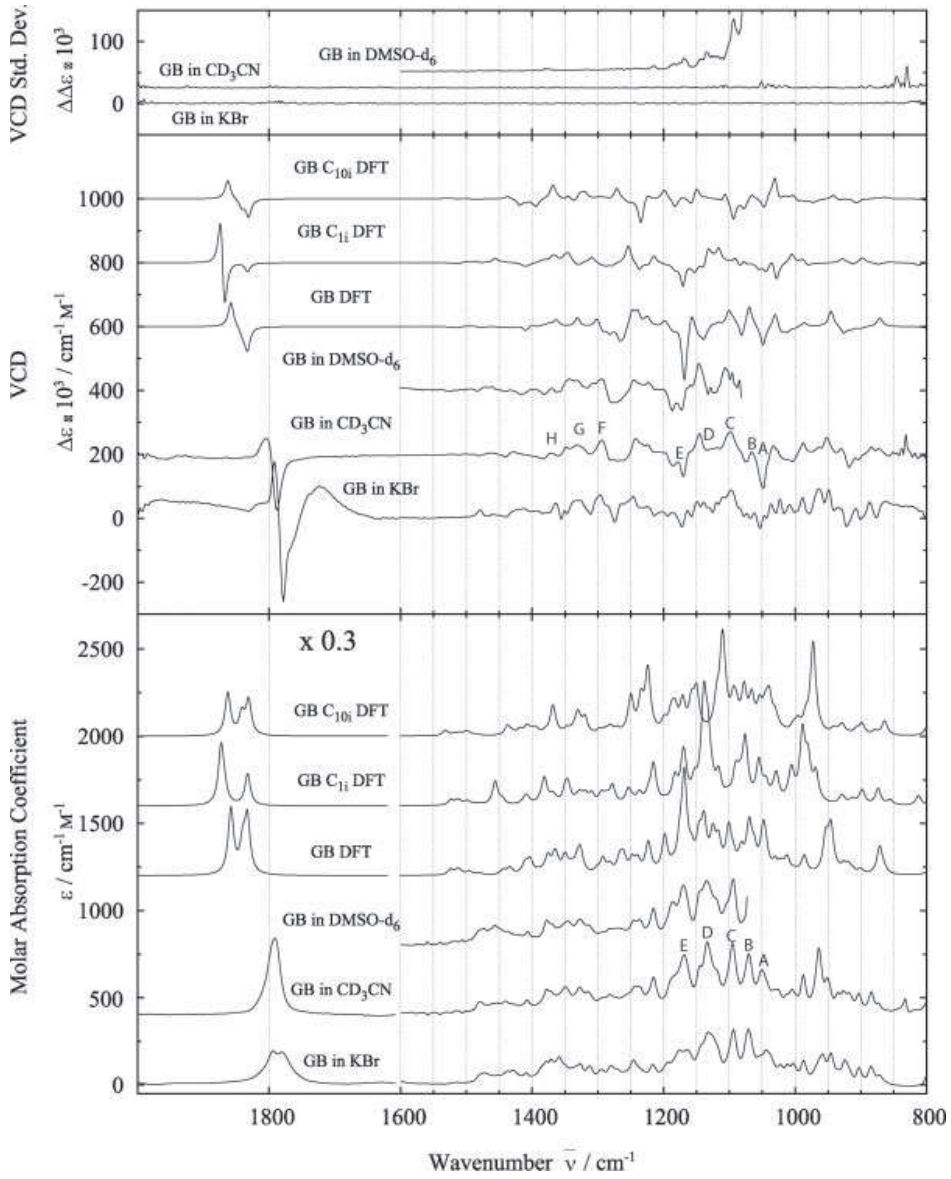


Fig. 3. Comparison of experimental IR and VCD spectra for GB with calculated (B3LYP/6-31G(d)) spectra for GB, GB-C_1i and GB-C_{10}i. The wavenumbers of the calculated transitions at B3LYP/6-31G(d) level are scaled by a factor 0.982. The intensity of the IR bands in the range between 1600 and 2000 cm^{-1} is multiplied by a factor 0.3. The upper panel illustrates VCD-noise, the middle panel VCD and the lower panel IR absorption. Prominent peaks for GB are indicated with the letters A, B, C, D, E, F, G, and H.

this study were carried out at the Quantum Protein Centre, the Technical University of Denmark, and at HPC, the University of Oslo, Norway.

RESULTS AND DISCUSSION

The structure of the lowest energy conformer of GB optimized at the B3LYP/6-31G(d) level of theory agrees well with the crystal structure[17] and with previous theoretical work.[14] Although there are discrepancies for certain bands, the overall agreements between the calculated and experimental VCD spectra of GB in the fingerprint region are striking, see Figure 2.

This suggests, as expected, that VCD spectroscopy can be applied in determination of absolute configuration of ginkgolides. Differences between the experimental spectra recorded for GB in KBr and CD_3CN and the theoretical spectrum in the C = O stretching region are observed, but the spectral pattern of VCD-signs remains similar. The IR spectrum of GB in KBr and Nujol have previously been recorded and partially assigned.[2,5,21,22] In the fingerprint region (850–1300 cm^{-1}), five broad and intense IR bands are observed at 1045, 1065, 1098, 1134, and 1170 cm^{-1}. These have been labeled A, B, C, D, and E in Figure 2 on the measured VCD and IR spectra for GB in CD_3CN. The first two of these bands are obscured by solvent absorp-

TABLE 1. The energy difference ΔE (kJ mol^{-1}) at different levels of theory between the two lowest conformers for ginkgolides with conformers within an energy window of 20 kJ mol^{-1}[a]

	ΔE (kJ mol^{-1})		
Ginkgolide	MMFF94s	B3LYP/ 6-31G(d)	B3LYP/6-311+ G(2d,2p)
GB	8	7	7
GC	9	7	8
GM	10	8	8

[a]For the lowest energy conformer of GA, GB, GC, GJ, GM, GB-C_1i, and GB-C_{10}i the total B3LYP/6-31G(d) energies are −3816425, −4013896, −4211357, −4013886, −4013889, −4013887, and −4013864 kJ mol^{-1}, respectively. In the same order the B3LYP/6-311+G(2d,2p) energies are −3817644, −4015200, −4212747, −4015191, −4015194, −4015186, and −4015172 kJ mol^{-1}.

tion in the experiment with DMSO-d_6 as solvent. Overarchingly, these bands are accompanied by negative VCD at 1170 and 1045 cm^{-1} and positive VCD at 1098 and 1065 cm^{-1}. Comparing with the other ginkgolides, the differences between the calculated spectra are prominent at certain wavenumbers. At around 1430 cm^{-1} GC, GJ and GM display IR peaks that are accompanied by a positive-negative VCD feature for GC and GM. Around 1340 cm^{-1} GA and GJ show similar VCD patterns, but this spectral motif appears to be inverted in case of GM. At 1300, 1330, and 1370 cm^{-1} (Labeled F, G, and H in Fig. 2) GB has a characteristic pattern consisting of three positive VCD peaks that seems to occur also for GC. The negative VCD features for GB between 1300 and 1260 cm^{-1} are similar to those for GC that also has a negative VCD shoulder at 1250, whereas GA and GJ have positive VCD peaks at around 1270 cm^{-1}. For GM this negative VCD pattern is limited to the range from 1280 to 1250 cm^{-1}. The strong negative VCD band at 1170 cm^{-1} is calculated for all the ginkgolides except GM where the peak is shifted to 1160 cm^{-1}. Except for GA, all ginkgolides have negative VCD signals near 1140 cm^{-1}. The observations from Figure 2 suggest that IR and VCD spectroscopy might allow for discrimination between GB and the other types of ginkgolide. However, this notion rests on the attractive, yet unverified, assumption that the measured spectrum is closely approximated by the calculated spectrum for each type of ginkgolide.

This work also includes calculations on the two GB stereoisomers GB-C_1i and GB-C_{10}i with single inversions at C_1 and C_{10}. These configurations are of particular relevance in connection with our future plans for synthesis. It is therefore of interest to test if the calculated spectra of these stereoisomers differ significantly from the measured GB spectra. From Figure 3 it is clear that the calculated IR and VCD spectra for GB do indeed differ markedly from those of GB-C_1i and GB-C_{10}i. Although several IR and VCD patterns appear alike, significant differences are found at 872, 926, 946, 1030–1130, 1157, and 1260–1290 cm^{-1} in the VCD spectra. Experimental verification is required, but the pronounced differences in calculated spectra suggest that discrimination between the corresponding experimental spectra might be possible.

It should be noted that the theoretical calculations presented here are performed on the isolated ginkgolide molecules, and therefore interactions between the ginkgolides and solvent (or other molecules) are not represented in the calculated spectra. The most conspicuous deviations between calculated and experimental GB VCD spectra are found at the band 1065 cm^{-1}, which is more positive in the calculated spectrum and at 1170 cm^{-1} which is more negative in the calculated spectrum. In the corresponding IR spectra, no significant difference between calculation and measurement is observed at 1065 cm^{-1}, but the calculated absorption intensity at 1170 cm^{-1} is markedly increased when compared with the measured spectrum. For the latter mode, C-O-H bending is prominent and the discrepancy may reflect interactions with solvent or neighboring GB molecules. By visualization of the calculated eigenmodes, C_{10}OH and C_3OH bending could be assigned to the band at 1170 cm^{-1} whereas C_1OH shows bending motion at 1134 cm^{-1}. This agrees with the observation that the calculated spectra for GA and GB-C_1i (see Fig. 3) display no negative VCD peak at the latter frequency, whereas the sign of VCD at the first frequency is inverted in case of GM that has no C_3OH.

The presence of more than one conformer is another potential origin of differences between measured and calculated spectra. As described in the experimental section the predicted spectra for GB, GC, and GM are composite VCD and IR spectra containing Boltzmann weighted contributions from the two lowest energy conformers. At room temperature the spectra for GB, GC, and GM are dominated by contributions from the lowest energy conformers by 94, 94, and 96%, respectively, assuming that

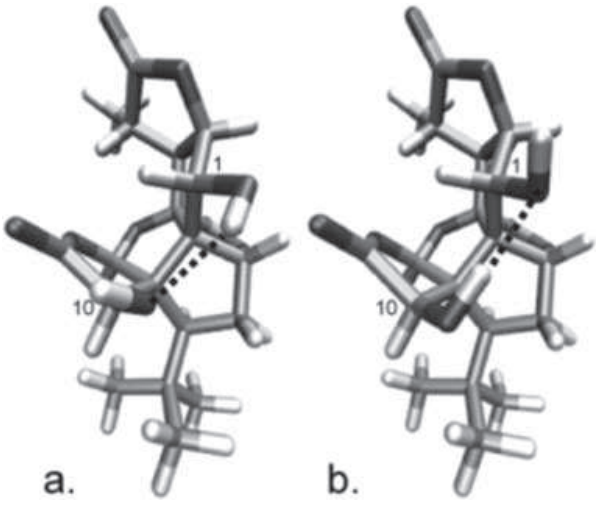


Fig. 4. The lowest (**a**) and the second lowest (**b**) energy conformers of ginkgolide B are separated by 7 kJ mol^{-1} according to DFT calculations (see the text for details). For both conformers, an intramolecular hydrogen bond is formed between the OH groups on C_{10} and C_1. The roles of hydrogen bond donor/acceptor changes from one conformer to the other. The hydrogen bond distance is 2.05 and 1.90 Å in conformer a and b, respectively.

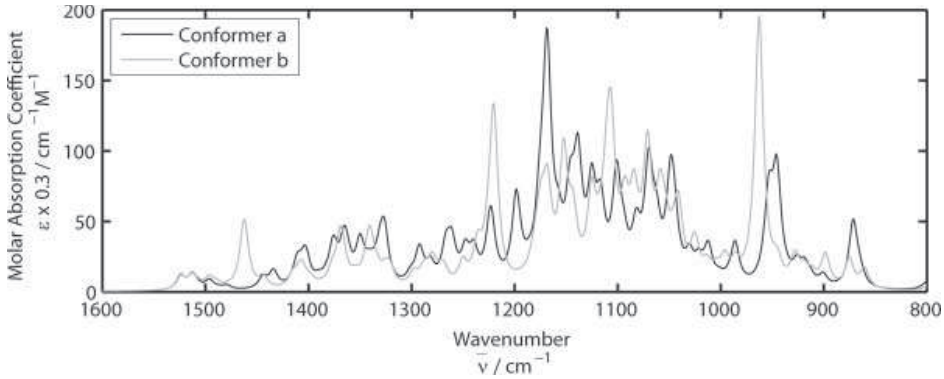


Fig. 5. Calculated IR spectra for conformer a and b (Fig. 4) of ginkgolide B. The peak in the spectrum for conformer b at approximately 1460 cm^{-1} suggests that it may be partly responsible for the absorption in the range 1450–1500 cm^{-1} observed experimentally. Similarly, conformer b displays lower absorption at 1170 and 1050 cm^{-1}, and a prominent peak at about 970 cm^{-1}, which may account for the diminished intensity or line-broadening in these regions of the experimental spectra.

the relative energies are calculated correctly with the theoretical method applied. Therefore only minor changes in the calculated spectra are seen when including the second conformers. However, the actual population of conformers may differ from the Boltzmann weighting scheme used, since the energies calculated at the B3LYP/6-311+G(2d,2p) level of theory may be uncertain with up to 13 kJ mol^{-1}.[23] Thus, it is interesting to investigate if contributions from the two conformers of GB (Fig. 4) can be discerned in the experimental spectra.

The experimental IR spectra for GB displays absorption in the range 1450–1500 cm^{-1}, whereas the corresponding calculated spectrum does not indicate absorption in this region. However, the calculated spectrum of the conformer b displays absorption at approximately 1460 cm^{-1}, see Figure 5. Thus, we hypothesize that presence of this conformer may be a least partly responsible for the absorption in the range 1450–1500 cm^{-1} observed experimentally. The notion of a significant contribution of the conformer b can also be invoked to account for the attenuated amplitudes at 1170 and 1050 cm^{-1} and the line broadening at 970 cm^{-1} observed in the experimental IR spectra with respect to the Boltzmann weighted spectra. Investigating the influence of conformers on the VCD spectra (see Fig. 6), it is evident that both conformers of GB have negative signals at 1170 cm^{-1}. The signal from conformer b is, however, significantly less negative than for conformer a. Thus inclusion of the conformer b might therefore account for the less negative VCD signal measured at 1170 cm^{-1}.

Although this suggests the presence of conformer b at room temperature, the discrepancies between calculated and experimental data may also originate from intermolecular interactions that are indeed likely to include the same two —OH groups (vide supra) that change structure between the two conformers. Alternatively, the discrepancies might originate from inadequacies of the applied theoretical method, but it is striking that the discrepancies are observed exactly for the bands where they are expected as a consequence of the presence of conformer b or intermolecular interactions, indicating that the either one or both of the two latter occur. The coexistence of conformer a and b may have implications for affinity and structure based selectivity of GB binding to receptors and other biomolecules.

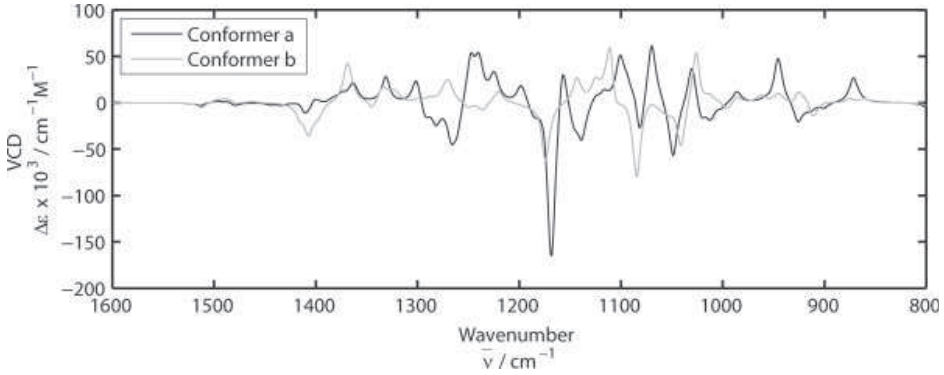


Fig. 6. Calculated VCD spectra for the conformer a and b (Fig. 4) for ginkgolide B. the presence of conformer b might account for the discrepancy between the experimental VCD and calculated signal at 1170 cm^{-1} (Fig. 2).

CONCLUSIONS

VCD and IR spectra were recorded for GB and predicted VCD and IR spectra were determined using density functional theory for GB and other ginkgolides, including stereoisomers of GB with inversions at a single chiral center. A striking similarity between the calculated and measured VCD and IR spectra for GB was found in the fingerprint region. This illustrates, as expected, the applicability of these methods to determine the absolute configuration of GB. The minor discrepancies between the experimental and calculated spectra for GB provide interesting information consistent with the presence of more than one GB conformer or intermolecular interactions.

ACKNOWLEDGMENTS

The authors thank the University of Oslo, Norway, for providing the superdome calculation facility at HPC and for collaboration with IMBV and the Department of Chemistry. The Danish National Research Foundation is acknowledged for financial support to the QUP center.

LITERATURE CITED

1. Strømgaard K, Nakanishi K. Chemistry and biology of terpene trilactones from Ginkgo biloba. Angew Chem Int Ed 2004;43:1640–1658.
2. van Beek TA. Pharmacological action and mechanisms of ginkgolide B. Bioorg Med Chem 2005;13:5001–5012.
3. Kondratskaya EK, Krishtal EL. Effects of Ginkgo biloba extract constituents on glycine-activated strychnine-sensitive receptors in hippocampal pyramidal neurons of the rat. Neurophysiology 2002;34:155–157.
4. Ivic L, Sands TTJ, Fishkin N, Nakanishi K, Kriegstein AR, Strømgaard K. Terpene trilactones from Ginkgo biloba are antagonists of cortical glycine and GABAA receptors. J Biol Chem 2003;278:49279–49285.
5. Braquet P, editors. Ginkgolides—chemistry, biology, pharmacology and clinical perspectives. Barcelona: JR Prous Science Publishers; 1988. 794 p.
6. Corey EJ, Gavai AV. Enantioselective route to a key intermediate in the total synthesis of ginkgolide B. Tetrahedron Lett 1988;29:3201–3204.
7. Freedman TB, Cao X, Rajca A, Wang H, Nafie LA. Determination of absolute configuration in molecules with chiral axes by vibrational circular dichroism: a C2-symmetric annelated heptathiophene and a D2-symmetric dimer of 1,1′-binaphthyl. J Phys Chem A 2003;107:7692–7696.
8. Brotin T, Cavagnat D, Dutasta J-P, Buffeteau T. Vibrational circular dichroism study of optically pure cryptophane-A. J Am Chem Soc 2006;128:5533–5540.
9. Bürgi T, Urakawa A, Behzadi B, Ernst KH, Baiker A. The absolute configuration of heptahelicene: a VCD spectroscopy study. New J Chem 2004;3:332–334.
10. He J, Petrovic AG, Dzyuba SV, Berova N, Nakanishi K, Polavarapu PL. Spectroscopic investigation of Ginkgo biloba terpene trilactones and their interaction with amyloid peptide Aβ(25-35). Spectrochim Acta Part A Mol Biomol Spectrosc 2008;69:1213–1222.
11. Nafie LA. Vibrational Circular Dichroism: a new tool for the solution-state determination of the structure and absolute configuration of chiral natural product molecules. Nat Prod Commun 2008;3:451–466.
12. Zekri O, Boudeville P, Genay P, Perly B, Braquet P, Jouenne P, Burgot J-L. Ionization constants of ginkgolide B in aqueous solution. Anal Chem 1996;68:2598–2604.
13. Nafie LA. Dual polarization modulation: real-time, spectral multiplex separation of circular dichroism from linear birefringence spectral intensities. Appl Spectrosc 2000;54:1634–1645.
14. Zhu W-L, Puah CM, Tan X-J, Jiang H-L, Chen K-X, Ji R-Y. A density functional theory calculation of the geometry and vibrational spectrum of natural product, ginkgolide B. J Mol Struct (Theochem) 2000;528:193–198.
15. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA. Gaussian 03, Revision D. 01. Wallingford, CT: Gaussian; 2004.
16. MacroModel, version 9.6. New York, NY: Schrödinger, LLC; 2008.
17. Dupont L, Dideberg O, Germain G, Braquet P. Structure of ginkgolide B (BN 52021) monohydrate, a highly specific PAF/acether receptor antagonist isolated from Ginko biboba L. Acta Crystallogr 1986;C42:1759–1762.
18. Sbit M, Dupont L, Dideberg O, Braquet P. Structure of ginkgolide A (BN52020) monohydrate and ginkgolide C (BN52022).Ethanol. 1·5 hydrate, isolated from Ginko biloba L. Acta Cryst 1987;C43:2377–2381.
19. Bohr HG, Jalkanen KJ, Elstner M, Frimand K, Suhai S. A comparative study of MP2, B3LYP, RHF and SCC-DFTB force fields in predicting the vibrational spectra of N-acetyl-L-alanine-N′-methyl amide: VA and VCD spectra. Chem Phys 1999;246:13–36.
20. Olsen L, Antony J, Ryde U, Adolph H-W, Hemmingsen L. Lactam hydrolysis catalysed by mononuclear metallo-beta-lactamases—a density functional study. J Phys Chem B 2003;107:2366–2375.
21. Yuan C, Pan J, Xu J. Study on the relation between ginkgolide B and water by FTIR. Spectrosc Spect Anal 2007;27:270–274.
22. Okabe K, Yamada K, Yamamura S, Takeda S. Ginkgolides. J Chem Soc C 1967;2201–2206.
23. Foresman JB, Frisch A. Exploring chemistry with electronic structure methods. Pittsburgh: Gaussian; 1996. 303 p.