NMR structural studies of oligosaccharides and other natural products

Kjærulf, Louise; Gotfredsen, Charlotte Held

Publication date:
2014

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):
NMR structural studies of oligosaccharides and other natural products

PhD Thesis

Louise Kjærulff

Department of Chemistry
Technical University of Denmark
June 2014
NMR structural studies of oligosaccharides and other natural products

Louise Kjærulff
Preface

This thesis is the result of three years of research as a PhD student at the Department of Chemistry, Technical University of Denmark, submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Chemistry. The work was carried out under the supervision of Assoc. Prof. Charlotte Held Gotfredsen with greatly acknowledged financing from the Danish Council for Strategic Research (DSF).

First and foremost, I would like to thank Lotte for giving me the opportunity to work with NMR spectroscopy in her group, and for all of her guidance and support for the last three years. I have definitely learnt a lot, and getting involved in the setup of the 400 MHz and going to ENC 2013 was awesome. Right from the start we knew we would be facing many new challenges and that we would need help from others as well. Therefore I owe a special thanks to Ole W. Sørensen and Andy J. Benie (Novo Nordisk A/S) for working with us on the pulse sequences and the programming. They are both priceless sources of knowledge and know-how that I could not have done without.

I want to thank the entire NMR group at DTU Chemistry but especially: Casper, for great help with calculations, alignment media, discussions, and numerous other things; and Anne, for always being very caring and helpful, and for very good food.

Through the years I have always felt like I had a second group at Systems Biology. Therefore I am very grateful to: Thomas Ostenfeld Larsen who gave me bacteria and fungi to work with; Maria Månsson who has been a great mentor, letting me continue to work with the peptides from my master’s thesis during my PhD; Lene, Kristian, Marie Louise, Tanja, Sara and all the others from building 221, who have been very welcoming. I enjoyed all of our discussions and your good input. And thank you, Kir, Lisette, and Hanne for 200 plates of Aspergillus and running lots of MS samples.

I would also like to express my gratitude to Anita Nielsen, Hanne Ingmer, and Lone Gram for our fruitful collaborations on the Galathea project - it was a real joy to work with you; likewise thanks for all the good collaborations in the HMO project.
The Danish Center for NMR Spectroscopy of Biological Macromolecules and the Carlsberg Laboratory is gratefully acknowledged for NMR time on the 800 MHz.

Thanks to the office mates in room 154 – I think it would have been three long years without you – and a general thanks to everyone from building 201, who have made my last three years at DTU unforgettable. Besides that, I am indebted to Mathilde for proofreading my thesis.

Last but not least, I must thank my friends and family for supporting me and just being there even though I sometimes have prioritized science over you. Thank you for understanding.

Kgs. Lyngby, June 2nd, 2014

Louise Kjærulff
Abstract

NMR spectroscopy is an important tool in chemical analysis and for structural studies in various research areas. The subject of this thesis is liquid state NMR spectroscopy for structural analysis of small molecules, both regarding method development and structure elucidation by NMR spectroscopy.

HMBC+ is a new pseudo-3D NMR experiment for measurement of long-range homonuclear $J_{HH}$ coupling constants in small molecules. Based on two existing experiments, multiplicity edited HMBC and HAT HMBC by Benie, Nyberg, and Sørensen, the new HMBC+ experiment takes advantage of a π phase difference with respect to $^{n+1}J_{HH}$ between these two experiments, observed in the $^nJ_{CH}$ HMBC cross peak. Through a double editing procedure this enables straightforward determination of both sign and magnitude of $^{n+1}J_{HH}$, including for very small coupling constants. Excellent results were obtained for the natural product strychnine, and 28 $^{n+1}J_{HH}$ coupling constants were determined, of which 10 had not previously been measured due to their small sizes. By comparing to calculated coupling constants, an RMSD of 0.28 Hz was obtained, all coupling constant signs fit with the calculations, and all errors were below 1 Hz, also when comparing the absolute values to experimental coupling constants measured by Carter et al. using 1D $^1H$ NMR on deuterated strychnine analogues. Providing an excellent tool for measurement of long-range $J_{HH}$ with the extra convenience of labeling the coupling constant information to the $^{13}C$ chemical shift through the $^nJ_{CH}$ correlation, this experiment has exciting applications for configurational assignment of e.g. carbohydrates and for residual dipolar couplings.

Identification of known molecules and discovery of novel molecules are other important applications of NMR spectroscopy. Bacteria and fungi produce secondary metabolites for signaling and competing against other organisms, and these molecules are important in drug discovery due to their inherent biological activities. From a marine Photobacterium (P. halotolerans) we isolated the solonamides and the ngercheumicins, two families of cyclic depsipeptides capable of attenuating virulence in S. aureus by quorum sensing inhibition through the accessory gene regulator agr. This is likely due to their structural similarities with the auto-inducing peptides of S. aureus, controlling quorum sensing in bacterial populations. A filamentous fungus, A. fijiensis, was also investigated for production of novel secondary metabolites, and a new pyranonigrin (E) was isolated and structure elucidated by
NMR spectroscopy along with JBIR-74 and decumenone A, two known metabolites previously isolated from *Aspergillus* and *Penicillium* species.

Oligosaccharides found in human milk are important for infant nutrition, and a collaborative effort of university and industry partners was aimed at establishing methods for production of human milk oligosaccharides. Two different bioenzymatic methods for production of 3’-sialyllactose were investigated, and a screening of trans-fucosidases enabled the NMR spectroscopic identification of three pNP-fucosylfucopyranosides as major reaction products.
Resumé

NMR spektroskopi er et af de vigtigste værktøjer til analyse og strukturoklaring af kemiske stoffer og bruges inden for mange forskellige forskningsområder. Denne afhandling omhandler NMR-spektroskopisk strukturanalyse af små molekyler i vandig eller organisk opløsning, både i forhold til metodeudvikling og strukturoklaring vha. NMR spektroskopi. HMBC+ er et nyudviklet pseudo-3D NMR eksperiment til måling af $n+1J_{HH}$ homonukleare koblingskonstanter over 3-4 bindinger i små molekyler, observeret gennem $nJ_{CH}$ korrelationer, hvor $n$ typisk er 2-3 bindinger. Baseret på to eksisterende NMR eksperimenter, editeret HMBC og HAT HMBC af Benie, Nyberg og Sørensen, benytter HMBC+ eksperimentet sig af en $\pi$ faseforskell ifht. $n+1J_{HH}$ mellem editeret HMBC og HAT HMBC. En dobbelt editeringsprocedure muliggør dermed direkte aflæsning af både fortegn og størrelse af de homonukleare koblingskonstanter, så både små og store koblingskonstanter kan måles med høj nøjagtighed. Dette vises for naturproduktet stryknin, hvor 28 $n+1J_{HH}$ koblingskonstanter bestemmes med fortegn, deraf 10 som ikke før er målt. Sammenlignet med beregnede koblingskonstanter fås middelværdien af kvadratafvigelsessummen (RMSD) til 0,28 Hz, alle fortegn stemmer overens med beregningerne, og alle målte koblingskonstanter ligger indenfor 1 Hz af både beregnede og eksperimentelle værdier. De målte koblingskonstanter kombineres desuden med $nJ_{CH}$ HMBC korrelationer så koblingskonstanten forbindes med et tilhørende $^{13}$C kemisk skift, som gør strukturoklaring og tilordning af relativ konfiguration lettere i fx kulhydrater, og HMBC+ eksperimentet vil kunne bruges til måling af residuale dipolare koblinger, hvorved man derved kan opnå yderligere strukturel information.

Identifikation af kendte kemiske stoffer og opdagelse af nye molekyler er også vigtige discipliner indenfor NMR spektroskopi. Bakterier og svampe danner sekundær metabolitter til kommunikation med og konkurrence imod andre organismer, og disse molekyler har biologiske aktiviteter som er vigtige i udviklingen af nye lægemidler. To familier af cykliske depsipeptider blev isoleret fra en marin fotobakterie, *P. halotolerans*: solonamiderne og ngercheumicinerne. Disse peptider kan hæmme virulens-mekanismer i multi-resistente *S. aureus* gennem *agr*, som i stafylokokker styrer den fælles kommunikation (quorum sensing) i bakteriekolonier når tilstrækkeligt høje populationer opnås. Fundet af denne aktivitet i en fotobakterie skyldes højst sandsynligt depsipeptidernes strukturelle ligheder med de naturlige vi
signalstoffer (AIP) for agr. Skimmelsvampen A. fijiensis blev ligeledes undersøgt i håb om at opdage andre nye sekundære metabolitter, og en ny pyranonigrin (E) blev isoleret og strukturoklaret ved brug af NMR spektroskopi, sammen med JBIR-74 og decumbenone A, to kendte stoffer tidligere isoleret fra stammer af Apergillus og Penicillium.

Oligosakkarider i modernælk er vigtige for spædbørns sundhed og velfærd, og et større projekt med universitets- og industripartnere havde til mål at finde nye metoder til at producere disse vigtige sukkerstoffer. To bioenzymatiske metoder til produktion af 3’-sialyllaktose blev udviklet, og en undersøgelse af trans-fucosidase enzymer ledte til NMR-spektroskopisk identifikation af tre pNP-fucosylfucopyranosider, som viste sig at være de primære oligosakkarider dannet ved reaktionen.
Abbreviations

1D  One dimensional
2D  Two dimensional
3D  Three dimensional
AA  Amino acid
Ac  Acetyl
ACCA Automatic coupling constants analyzer
ADEQUATE Adequate double quantum transfer experiment
agr Accessory gene regulator
AIP Autoinducing peptide
Ala Alanine
AP Anti-phase
AU Automation
ax Axial
BBO Broadband observe
bs broad singlet
HETSERF Heteronuclear selective refocusing
CA Community associated (or acquired)
CD Circular dichroism
CGMP Caseinoglycomacropeptide
CLIP Clean in-phase
CMP Cytidine-5’-monophosphate
COSY Correlation spectroscopy
CYA Czapek yeast autolysate
d Doublet
DAD Diode array detector
dd Doublet of doublets
ddd Doublet of doublets of doublets
DFT Density functional theory
DMSO Dimethylsulfoxide
DNP Dynamic nuclear polarization
dq Doublet of quartets
DQ Double quantum
DQF Double quantum filtered
dt Doublet of triplets
DTU Technical University of Denmark
E.COSY Exclusive correlation spectroscopy
eq Equatorial
ESI Electrospray ionization
E-SPE Explorative solid-phase extraction
EtOAc Ethyl acetate
FC Fermi contact
FDA Food and drug administration
FDAA | 1-fluoro-2,4-dinitrophenyl-5-L-alanine amide (Marfey’s reagent)
---|---
FID | Free induction decay
FL | Fucosyllactose
FOS | Fructooligosaccharides
Fuc | Fucose
g | gradient-selected (e.g. gHSQC)
Gal | Galactose
GalNAc | N-acetylgalactosamine
Glc | Glucose
GlcNAc | N-acetylglucoasamine
GOS | Galactooligosaccharides
H2BC | Heteronuclear 2-bond correlation
HAT | Homonuclear J-attenuated
HECADE | Heteronuclear couplings from aSSCI-domain experiments with E.COSY-type cross peaks
HETLOC | Heteronuclear long-range coupling
hla | α-hemolysin (encoding gene)
HMBC | Heteronuclear multiple bond correlation
HMO | Human milk oligosaccharide
HMQC | Heteronuclear multiple quantum correlation
HPAEC | High performance anion exchange chromatography
HPLC | High performance liquid chromatography
HR | High resolution
HSQC | Heteronuclear single quantum coherence
HSQMB | Heteronuclear single quantum multiple-bond connectivity
HTS | High-throughput screening
Ile | Isoleucine
INADEQUATE | Incredible natural abundance double quantum transition experiment
INEPT | Insensitive nuclei enhanced by polarization transfer
IP | In-phase
IPAP | In-phase/anti-phase
JM | J-modulated
LC | Liquid chromatography
Leu | Leucine
LNFP-V | Lacto-N-fucopentaose V
LNNFP-V | Lacto-N-neofucopentaose V
LNNT | Lacto-N-neotetraose
LNT | Lacto-N-tetraose
m | Multiplet
MALDI | Matrix-assisted laser desorption ionization
MAX | Mixed-mode anion exchange
MEA | Malt extract agar
MeOH | Methanol
MRI | Magnetic resonance imaging
MRSA | Methicillin-resistant *S. aureus*
MS | Mass spectrometry
MTPA | α-Methoxy-α-trifluoromethylphenylacetic acid
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>NeuAc</td>
<td>N-acetylneuraminic acid (IUPAC: 5-acetamido-3,5-dideoxy-D-glycero-α-D-galacto-non-2-ulopyranosonic acid)</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>NOE</td>
<td>Nuclear Overhauser effect</td>
</tr>
<tr>
<td>NOESY</td>
<td>Nuclear Overhauser effect spectroscopy</td>
</tr>
<tr>
<td>NP</td>
<td>Natural product</td>
</tr>
<tr>
<td>NUS</td>
<td>Non-uniform sampling</td>
</tr>
<tr>
<td>PAD</td>
<td>Pulsed amperometric detector</td>
</tr>
<tr>
<td>P.E.</td>
<td>Primitive exclusion</td>
</tr>
<tr>
<td>Phe</td>
<td>Phenylalanine</td>
</tr>
<tr>
<td>PHP</td>
<td>Parahydrogen-induced polarization</td>
</tr>
<tr>
<td>pNP</td>
<td>p-nitrophenyl or p-nitrophenol</td>
</tr>
<tr>
<td>ppm</td>
<td>Parts per million</td>
</tr>
<tr>
<td>PTFE</td>
<td>Polytetrafluoroethylene</td>
</tr>
<tr>
<td>QS</td>
<td>Quorum sensing</td>
</tr>
<tr>
<td>QSI</td>
<td>Quorum sensing inhibitor</td>
</tr>
<tr>
<td>qTOF</td>
<td>Quadrupole time of flight</td>
</tr>
<tr>
<td>RDC</td>
<td>Residual dipolar coupling</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root-mean-square deviation</td>
</tr>
<tr>
<td>ROESY</td>
<td>Rotating frame nuclear Overhauser effect spectroscopy</td>
</tr>
<tr>
<td>RP</td>
<td>Reverse phase</td>
</tr>
<tr>
<td>S/N</td>
<td>Signal-to-noise ratio</td>
</tr>
<tr>
<td>s</td>
<td>Singlet</td>
</tr>
<tr>
<td>SAX</td>
<td>Strong anion exchange</td>
</tr>
<tr>
<td>SCX</td>
<td>Strong cation exchange</td>
</tr>
<tr>
<td>SDU</td>
<td>University of Southern Denmark</td>
</tr>
<tr>
<td>Ser</td>
<td>Serine</td>
</tr>
<tr>
<td>SJS</td>
<td>Selective J-scaled</td>
</tr>
<tr>
<td>SL</td>
<td>Sialyllactose</td>
</tr>
<tr>
<td>spa</td>
<td>Staphylococcal Protein A</td>
</tr>
<tr>
<td>SPE</td>
<td>Solid-phase extraction</td>
</tr>
<tr>
<td>spp.</td>
<td>Species</td>
</tr>
<tr>
<td>t</td>
<td>Triplet</td>
</tr>
<tr>
<td>Thr</td>
<td>Threonine</td>
</tr>
<tr>
<td>TLC</td>
<td>Thin layer chromatography</td>
</tr>
<tr>
<td>TOCSY</td>
<td>Total correlation spectroscopy</td>
</tr>
<tr>
<td>UHPLC</td>
<td>Ultra high pressure liquid chromatography</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>VCD</td>
<td>Vibrational circular dichroism</td>
</tr>
<tr>
<td>WP</td>
<td>Work package</td>
</tr>
<tr>
<td>XLOC</td>
<td>X nucleus for long-range couplings</td>
</tr>
<tr>
<td>YES</td>
<td>Yeast extract sucrose</td>
</tr>
</tbody>
</table>
Publications


The publications I-V are included in this thesis and can be found in the Appendix, with supplementary information when relevant.
## Contents

1 Introduction ......................................................................................................................2
  1.1 70 years of nuclear magnetic resonance ..............................................................2

2 New NMR experiments ....................................................................................................5
  2.1 Structure and function - stereochemistry and absolute configuration .................5
  2.2 Residual dipolar couplings ..................................................................................6
  2.3 Measurement of one-bond heteronuclear coupling constants ..........................10
  2.4 Measurement of long-range coupling constants .................................................11
  2.5 HMBC+, a new NMR experiment for measurement of $^{n+1}J_{HH}$ ................13
      2.5.1 The editing procedure ..............................................................................17
      2.5.2 Determination of coupling constants ....................................................22
      2.5.3 Application in oligosaccharide analysis ...................................................28
  2.6 Experiments for measurement of $^2J_{CH}$ ......................................................29
  2.7 Conclusions and perspectives .............................................................................36

3 Microbial natural products ............................................................................................39
  3.1 Drug discovery today .........................................................................................39
  3.2 Sources of natural products ..............................................................................40
  3.3 Analysis of natural products .............................................................................41
      3.3.1 The supply problem and dereplication ....................................................41
      3.3.2 NMR spectroscopy of small amounts ......................................................42
      3.3.3 Determination of absolute stereochemistry .............................................42
  3.4 Natural products from a marine *Photobacterium* .........................................46
  3.5 Results and discussion .......................................................................................46
      3.5.1 Small biologically active peptides from a marine *Photobacterium* ......46
      3.5.2 Quorum sensing and autoinducing peptides ............................................50
      3.5.3 Fungal metabolites from *A. fijiensis* .......................................................53
  3.6 Conclusions and perspectives .............................................................................56

3.7 Experimental .............................................................................................................58
      3.7.1 Fungal extracts .......................................................................................58
      3.7.2 Fractionation and purification .................................................................58
      3.7.3 Bioactivity testing ....................................................................................59
3.7.4  LC-MS and NMR analyses ................................................................. 59

4  Human milk oligosaccharides ........................................................................ 61
   4.1  Introduction ............................................................................................ 61
   4.2  The HMO project .................................................................................... 62
   4.3  Biological role of HMOs ....................................................................... 63
   4.4  Sources of fucose and sialic acid ............................................................. 64
      4.4.1  Perspectives on the production of HMOs ....................................... 65
   4.5  Characterization of HMOs ..................................................................... 66
      4.5.1  HMO analysis by NMR .......................................................... 67
      4.5.2  NMR structural assignment of HMO molecules ................................ 68
      4.5.3  Purity assessment ......................................................................... 70
      4.5.4  CGMP as sialic acid donor .......................................................... 72
      4.5.5  Sialyl-oligosaccharides from trans-sialidases .................................. 74
      4.5.6  Trans-fucosylations ....................................................................... 77
      4.5.7  Enzymes in reaction monitoring .................................................... 79
   4.6  Conclusions and perspectives ................................................................. 83
   4.7  Experimentals .......................................................................................... 85
      4.7.1  Reaction monitoring with CGMP .................................................... 85
      4.7.2  Reaction monitoring with 3’-SL ....................................................... 85

5  Overall conclusions and perspectives ......................................................... 86

6  References ..................................................................................................... 88

Appendix ........................................................................................................... A-1
   Paper I .......................................................................................................... A-39
   Paper II ......................................................................................................... A-67
   Paper III ....................................................................................................... A-103
   Paper IV ....................................................................................................... A-119
   Paper V ......................................................................................................... A-127
1 Introduction
1 Introduction

The subjects within this thesis range from design of new experiments for measurement of coupling constants, over discovery of bacterial and fungal natural products, to NMR structural studies of human milk oligosaccharides. However, the link connecting all these projects is nuclear magnetic resonance (NMR) spectroscopy. This chapter gives a brief introduction to NMR spectroscopy and some of the discoveries within the field.

1.1 70 years of nuclear magnetic resonance

Much has changed since Purcell and Bloch earned the Nobel Prize in 1952 for their early discoveries and developments within nuclear magnetic resonance (NMR) spectroscopy. The respective research groups at Massachusetts Institute of Technology and Stanford University observed the first NMR spectra of paraffin and water in 1945, and today NMR is one of the most important and informative analytical methods of all. NMR is used in many diverse areas of research and science; in chemistry, biology, medicine, physics, material science, and on the interfaces (Figure 1.1). In these areas NMR has become an important tool to probe structure and interactions, including dynamics. Experiments are performed in the liquid, solid, and gas phases as well as in biological systems. Magnetic resonance imaging (MRI) is a whole world of its own, and in essence one can examine almost any kind of material using a powerful magnet and radio waves.

The first NMR spectrum of water by Bloch et al. had a $^1\text{H}$ larmor frequency of 7.76 MHz with a signal-to-noise ratio (S/N) of about 6. The first spectrometers used a continuous wave to sweep the magnetic field, but with the introduction of Fourier transform spectroscopy in 1966, NMR spectroscopy took a great leap forward. With this improvement, the NMR spectrum was calculated mathematically by Fourier transforming the ringing caused by a single pulse. A few years later Jeener introduced 2D NMR spectroscopy which was later published in the form of 2D correlation spectroscopy (COSY) by Aue, Bartholdi and Ernst in 1976.

Today, spectrometers operate at up to 1 GHz with ultra-shielded, sub-cooled, superconducting magnets and specialized probes for optimum resolution and sensitivity, running advanced experiments on sub-mg samples. During the years a major objective in
NMR has been the race for the highest field magnets, as higher field means better S/N ratio and higher resolution, e.g. doubling field strength almost triples S/N and doubles sensitivity.\textsuperscript{9} Nowadays there is a slight shift in the paradigm as we cannot continue increasing the field strength much further, and the spiraling cost of cryogens cause spectrometer vendors to look in other directions, \textit{vide infra}. New systems are increasingly built for minimizing cryogen refills and pleasing the general costumer rather than pushing frontiers in field strength, however current advances in probe design increase spectrometer sensitivity, and e.g. cryoprobes enhance S/N fourfold.\textsuperscript{10}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Some of the different research areas and types of samples analysed by NMR.}
\end{figure}

One of the later great achievements in NMR is the use of dynamic nuclear polarization (DNP), which is a technique where electron spin-polarization is transferred to nuclear polarization in a sample of interest resulting in an extreme gain in sensitivity.\textsuperscript{11} DNP is
gaining a lot of interest both in liquid and solid state NMR and particularly for imaging purposes and for investigating metabolic pathways in vivo.\textsuperscript{12–14} Unfortunately the polarization decays rapidly, and it is therefore not applicable to time consuming experiments with the current technology. A related and very promising new technique is zero-field NMR – the measurement of NMR spectra without magnets.\textsuperscript{15} The method uses parahydrogen-induced polarization (PHIP) and makes a fingerprint of a molecule’s $J$ couplings with no chemical shift information. A recent development to increase resolution or decrease experiment time is non-uniform sampling (NUS), which is applied to multidimensional NMR experiments with great advantage.\textsuperscript{16,17} Here, the idea is to sample the time domain data non-uniformly in $nD$ experiments with an experiment time decreasing by $3^{\frac{n}{1}}$, which can alternatively be traded off for increased resolution.\textsuperscript{18}

In NMR spectroscopy today, there are two major groups of research: One group is concerned with small molecule NMR in the liquid state, where natural isotopic abundance allows the acquisition and interpretation of various $nD$ experiments for structure elucidation and investigating molecular interactions. The other group focuses primarily on protein NMR in the solid state, usually requiring isotopic labelling and using a variety of other $nD$ experiments, most often specifically tailored for proteins. Without doubt NMR has had a tremendous impact on many areas of science, and for a technology only about 70 years old we can expect to see many more exciting applications and improvements in the future.

The work described in this thesis is concerned with structure elucidation of small molecules through the identification of known and unknown molecules in solution, both in pure form and in mixtures, but also through the investigation of molecular structure and conformation by the use of NMR coupling constants.

The thesis is divided into three major chapters. Following this introduction, chapter 2 describes new NMR experiments for measurement of long-range coupling constants in small molecules, chapter 3 describes the isolation of new natural products from bacteria and fungi with a discussion of their bioactivities (including Paper I and II), and chapter 4 covers NMR spectroscopic studies of human milk oligosaccharides (including Paper III, IV, and V).
2 New NMR experiments

This chapter is related to the development of new NMR experiments for measurement of coupling constants with the aim of creating new tools for structural studies by NMR spectroscopy. An introduction to methods for studying absolute and relative configuration is given, along with a brief description of the background for the experimental work and related applications. Finally, the new NMR experiments for measurement of coupling constants are described and discussed, with emphasis on the applications rather than the underlying theory. A publication related to the work described in this chapter is in preparation, but not included. For technical details, assignments, program codes, and more, see the Appendix.

2.1 Structure and function - stereochemistry and absolute configuration

Nature is chiral. Like the right and left hands are different from each other and cannot be superimposed, nature uses right- and left-handedness in many aspects. The world is three dimensional and so are our bodies and the molecules which make it function. This chirality is reflected in the configuration and conformation of molecules. E.g. proteins and carbohydrates require specific configurations to be recognized in the body, and especially for proteins conformation is crucial. If a protein is folded incorrectly the biological function disappears. In the structure elucidation of new molecules it is also very important to find the absolute structure – to know the configuration of all stereocenters in the molecule.

In 1951 Bijvoet et al. solved the absolute configuration of (+)-tartrate as (2R,3R) by the use of X-ray crystallography. This was the first time chirality was linked to molecular structure, and X-ray crystallography remains superior for elucidating absolute configuration in crystalline materials. However, the method requires a well-defined crystal lattice and it can be very difficult, sometimes impossible, to obtain a crystal of sufficient quality. Absolute configuration can be measured by other methods, but they all require a basis for comparison, usually from previous measurements and/or other methods of analysis. These methods include optical rotation and circular dichroism (CD). Chiral molecules interact with polarized light, and measurement of a solution’s optical activity is a measure of its ability to rotate plane polarized light of a specific wavelength. In CD, the light (or more correctly electromagnetic radiation) is circularly polarized, and the difference in absorption of right
and left polarized light as a function of wavelength yields a CD spectrum. However these methods require prior knowledge from analogues or similar scaffolds to yield proof of a certain configuration. Novel compounds with several stereocenters are not easily analysed by these methods and meso-compounds, although containing several stereocenters, are optically inactive. Vibrational CD (VCD) uses infrared radiation and enables determination of absolute configuration by comparison to calculated spectra, as enantiomers will have exact opposite VCD spectra.\textsuperscript{21,22}

Chemical methods with chiral derivatizing agents have been widely used to determine absolute configuration for certain types of molecules. For example secondary alcohols and amines can be derivatized to form Mosher’s esters which are subsequently analysed by NMR spectroscopy,\textsuperscript{23} and the absolute configuration in peptides and alkaloids can be determined using Marfey’s method.\textsuperscript{24} In natural products chemistry, total synthesis is also used to prove absolute structure, although it is a slow and costly method in most instances.

NMR spectroscopy also offers more direct ways to examine stereochemistry. Chiral Solvating agents create a diastereomeric environment in enantiomers and thus results in changes in the chemical shifts.\textsuperscript{25} But like the optical methods there is no clear correlation between a spectrum and the absolute configuration of a molecule, and thus the method has little use without something to compare to. More widespread ways to use NMR spectroscopy for obtaining stereochemical and conformational relationships is the use of scalar ($J$) coupling constants and nuclear Overhauser effect (NOE) spectroscopy. The direct relationship between dihedral angles and $J$ coupling constants\textsuperscript{26,27} together with the distance information from NOEs\textsuperscript{28} have long been important tools to obtain structural information from NMR spectroscopy, however NMR methods usually only yield relative configuration rather than absolute structures. During the last 20 years a third method has gained increasing interest, particularly in combination with the $J$ couplings and NOEs: Residual dipolar couplings (RDC).

### 2.2 Residual dipolar couplings

In 1963 Saupe and Englert were the first to observe dipolar couplings when they acquired “high resolution” $^1$H NMR spectra of a small molecule in liquid crystals.\textsuperscript{29} One year later Saupe had established the theoretical foundations\textsuperscript{30} but it should take many years before the NMR community realized the full potential of these discoveries. Dipolar couplings have long been known as a nuisance from solid state NMR, but it took 30 years before Tolman and Prestegard realized that RDCs could be used together with long-range NOEs for the study of
field-oriented proteins in the liquid state. Two years later Tjandra and Bax had shown how RDCs could be controlled and used in structure calculations of proteins in aqueous solution and since then RDCs have gained increasing interest for structural studies of proteins and small molecules in solution.

Dipolar couplings are observed in the liquid state upon introduction of anisotropy by an alignment medium, as opposed to the conventional liquid state NMR spectra acquired in isotropic solution. Dipolar couplings as known from solid state NMR are in the order of 20 kHz, however observing coupling constants of this magnitude is not desirable and the dipolar couplings are conveniently scaled with the degree of alignment, explaining the term “residual”. Partial alignment in the order of 1‰ is said to be optimal, and with rapid averaging between aligned and unaligned molecules the RDCs in a weakly aligned sample are usually in the order of ±0-20 Hz. The dipolar coupling $D$ is observed together with the scalar coupling as the total coupling constant $T$, which is the sum of $J$ and $D$.

$$T = J + D$$  (Eq. 2.1)

There is however some researchers who instead define $T$ as $J + 2D$ due to conventions from solid state NMR. Most research groups as well as the software PALES for structure refinement uses $J + D$. If $D$ is viewed in a manner similar to $J$, it makes most sense to use the definition in Eq. 2.1 with $D$ as the contribution to the line splitting, and as long as it is used consistently, no errors will result from either.

Residual dipolar couplings result from dipole-dipole interactions through space and are thus related to NOEs via the nuclear Overhauser effect. The mathematical expression for $D$ between two spins $I$ and $S$ is

$$D_{IS} = \frac{\gamma_I \gamma_S \mu_0 h}{(2\pi)^3 r_{IS}^3} \left(\frac{3\cos^2 \theta - 1}{2}\right)$$  (Eq. 2.2)

Where $\gamma$ is the gyromagnetic ratio for the individual nuclei, $\mu_0$ is the vacuum permeability, $h$ is Planck’s constant, $r_{IS}$ is the internuclear distance, and $\theta$ is the angle between the interatomic vector and the magnetic field, $B_0$ (Figure 2.2). The brackets around the right hand part of the equation denote the averaging of observed molecules as they tumble in solution. As seen from Eq. 2.2, the RDC between two spins varies solely with internuclear distance and orientation.
Whereas the NOE depends on internuclear distance by a factor of $r^6$, the RDC is not only longer ranging, by $r^3$, but yields valuable orientational information. The angular relationship makes it possible to relate configurational relationships and determine conformations for mainly rigid structures. Whereas X-ray diffraction is limited to the crystalline state, RDCs opens up the possibility to study configuration and conformation in the liquid state and take on a more dynamic approach with e.g. protein-drug interactions. About a decade after the paper by Tolman and Prestegard researchers began to see the potential of RDCs in relation to small molecules. A search in SciFinder shows that the number of publications per year on residual dipolar couplings (both in relation to proteins and small molecules) have been more or less constant for the last decade with a slight decline after it peaked in 2004 (Figure 2.3). Today there are more than 2,500 publications on the subject of RDCs.

The first alignment media were for biomolecules and therefore these media were mostly compatible with aqueous samples. Today’s alignment media for small molecules are conveniently compatible with organic solvents such as dimethylsulfoxide (DMSO) or chloroform, and the predominant alignment media for small molecules are strain-induced polymers and liquid crystals.

RDCs generally yield only relative stereochemical information, which is a considerable disadvantage when it comes to small molecules. It is debated whether the use of chiral alignment media makes it possible to determine absolute configuration, and a few papers describe initial attempts. It will however be difficult to find one chiral alignment medium suitable for a wide range of samples, and understanding the specific interactions involved will be even more difficult. Therefore enantiomeric discrimination in aligned samples is not yet attainable as a general method, but the possibility to elucidate stereochemical

---

**Figure 2.2.** Two-dimensional vector representation of the angle $\theta$ between the $B_0$ field and the internuclear vector between spins I and S.
relationships in rigid molecules over larger distances than NOEs nevertheless makes dipolar couplings very interesting, and in rigid molecules with one known stereocenter the relative structural relationships can be determined more reliably with RDCs than by using $J$ coupling and NOE analysis alone.\textsuperscript{55,56}

Figure 2.3. Stick plots of SciFinder hits on “Residual dipolar coupling” as entered and the concept (top) and a cumulative plot of the hits on the concept (bottom). This search covers both proteins and small molecules. “The concept” includes the entered term, synonymous terms and similar terms when searching within the records.

RDCs are used in calculations to either confirm a given structure or calculate one that fits restraints from the experimental data. Restrained molecular dynamics are used to simulate a structure consistent with the data available, e.g. $J$ couplings, NOEs and RDCs. A simple
example of how RDCs can be used is for the discrimination of axial and equatorial bonds in six-membered rings (incl. carbohydrates) by $^{1}D_{CH}$ couplings, as bonds pointing in the same direction will display one-bond RDCs of the same magnitude. Analysis of more complicated molecules requires the construction of an alignment tensor; a symmetric 3x3 matrix with five independent parameters, which basically describes the molecule’s orientation in relation to the magnetic field. Specific software for use of RDCs in calculations exists, e.g. PALES and REDCAT primarily designed for macromolecular applications, and MSpin-RDC for small molecules. One of the limitations of RDCs for small molecule structure calculations is the need for minimum five RDCs in order to create the local alignment tensor, and additional RDCs are used to validate it. Each flexible element in a structure requires a minimum of five non-parallel RDCs and therefore a calculation quickly becomes complicated for molecules with high degrees of torsional freedom. In these cases the structural ensemble cannot be embodied by just one structure, and as there is only one set of NMR data for each molecule, the output is always an average solution structure.

2.3 Measurement of one-bond heteronuclear coupling constants

The most commonly used RDCs are one-bond HSQC-type $^{1}D_{CH}$ couplings. Unlike for long-range RDCs, there is only one $^{1}D_{CH}$ for each proton, and in some cases it is not possible to obtain five independent RDCs to create the alignment tensor, making it necessary to calculate the missing information by extrapolation. The one-bond heteronuclear RDC ($^{1}D_{CH}$) is a popular choice for several reasons. The data interpretation and setup of the experiment is straightforward for anyone acquainted with the HSQC experiment and a typical $^{1}J_{CH}$ coupling constant is about 145 Hz and always positive in sign. This is convenient as the RDC creates an additional splitting of the $J$ coupling upon alignment, and with a relatively small degree of alignment ($D << J$) it is straightforward to estimate both magnitude and sign of $^{1}D_{CH}$. A range of HSQC-type experiments have been designed over the years for the purpose of measuring $^{1}J_{CH}$ and $^{1}D_{CH}$. The simplest version is the $t_2$-coupled HSQC, where $^{1}J_{CH}/^{1}T_{CH}$ couplings are determined in the direct dimension, but it suffers from distorted peak shapes and interferences from long-range homonuclear couplings, which can make accurate measurement difficult. The $t_1$-coupled HSQC has interferences from long-range heteronuclear couplings and it requires many $t_1$-increments and thus longer acquisition time to afford accurate measurement of the couplings. Several other HSQC/HMQC-type experiments have been developed for measuring one-bond $J/D$ coupling constants.
2.4 Measurement of long-range coupling constants

It is highly desirable to supplement the $^1D_{CH}$ with homo- and heteronuclear long-range RDCs to extend the number of available dipolar couplings and thereby improve the structural determination. The long-range $J$ coupling constants are however smaller in magnitude compared to $^1J_{CH}$ and they may be negative, although many of the experiments for measurement of long-range couplings are not sign selective, which complicates the structural analysis. These small $J$ coupling constants are difficult to measure accurately, and opposed to the large $^1J_{CH}$, an error of a few Hz can have great impact on the long-range coupling and its interpretation. Any experiment for measurement of $J$ couplings can be used to measure RDCs, and therefore the term “coupling constant” refers to both $J$ and residual dipolar couplings. If $J$ is however not determined sign sensitively, the determination of $D$ becomes very difficult due to the many possible combinations to give $T$.

Basically there are currently two ways to obtain long-range $^0J_{CH}$ couplings: Either by $^{13}\text{C}$-edited TOCSY correlation-type experiments or by long-range $^1\text{H}-^{13}\text{C}$ correlations in HMBC-type experiments. The key to measuring long-range coupling constants (both $J$ and RDCs) is to minimize spectral overlap and resonance cancellation, which is attempted in different ways. Most often the coupling constant is measured as a line-splitting between the $\alpha$- and $\beta$-spin states which can be in-phase (IP), anti-phase (AP) or using both to edit the multiplet components into individual subspectra (IPAP) (Figure 2.4).

![Figure 2.4](image_url)

**Figure 2.4. Principle of IPAP-based experiments, where addition and subtraction of IP and AP data yields separate $\alpha$ (top) and $\beta$ (bottom) spin state components. The two lines in the stick plots are separated by $J$ or $T$. Modified from 71.**

For analysis of small coupling constants an IP cross peak often yields an underestimated coupling constant while the AP cross peak has a tendency to make the coupling constant seem larger due to cancellation of resonances in the cross peak center (Figure 2.5). These methods are therefore best suited in combination with fitting procedures which complicates data analysis. The IPAP separates the individual cross peak components and overlay of the spectra affords easy and accurate data interpretation, and is more suitable for measurement of small ($< 2$-$3$ Hz) coupling constants (Figure 2.4). Another general method is $J$-resolved
spectroscopy, where the desired (in this case long-range) coupling constants are developed independent of the heteronuclear chemical shift, yielding a splitting of $J$ in the indirect dimension (F1) (Figure 2.5). This has the inherent drawback of long experiment time as it requires many $t_1$ increments for sufficient resolution. Several experiments for long-range coupling constants build on the exclusive correlation spectroscopy (E.COSY) principle, where passive $J_{HH}$ coupling constants are suppressed to instead displace the active coupling pattern in the respective dimensions in the size of the passive coupling constants.\textsuperscript{72–75} This allows accurate measurement of small coupling constants, as long as at least one of the passive coupling constants is large enough so the cross peak fine structure is resolved. All experiments for sign-selective measurement of $J_{HH}$ have so far been based on this principle, and the XLOC (X-nucleus for long-range couplings) experiment\textsuperscript{76–79} for measurement of $^1J_{XH}$ and $^{n+1}J_{HH}$ is very robust and yields valuable information.\textsuperscript{38,80}

![Figure 2.5. Simple cross peak patterns for IP, AP, IPAP, $J$-resolved and E.COSY type experiments. Blue is positive resonance intensity and grey is negative. Modified from \textsuperscript{71}.](image)

The E.COSY-type multiplet pattern enables the measurement of two coupling constants from one cross peak as well as obtaining the relative sign of the coupling constants from the direction of the displacement vector of the cross peaks, and for the experiments where $^1J_{CH}$ or $^1J_{NH}$ are involved, the long-range coupling constants are thus measured sign-selectively as the signs of these coupling constants are always positive or negative, respectively.\textsuperscript{38} These experiments include the P.E.HSQC (primitive exclusion HSQC)\textsuperscript{66} yielding $^1J_{CH}$ and $^2J_{HH}$, as well as XLOC for measurement of $^1J_{XH}$ and $^{n+1}J_{HH}$ (mentioned above) and HETLOC (heteronuclear long-range couplings)\textsuperscript{81} for $^1J_{CH}$ and $^nJ_{CH}$. The HETLOC experiment was the first experiment for sign-sensitive measurement of long-range heteronuclear coupling constants, however due to a TOCSY (total correlation spectroscopy) coherence transfer, no correlations to or across quaternary carbons or heteroatoms are observed and resonances may disappear when $J_{HH}$ vanishes. A modification of the HETLOC experiment is called HECADE.\textsuperscript{82} This reduces resonance overlap in F1, but the slightly longer pulse sequence
also results in reduced sensitivity. In 1993 Zhu et al. developed a $J$-quantitative experiment to extract coupling constants from the cross peak intensities. The literature reveals a wide range of HMBC-type experiments which may allow the determination of coupling constants to and across quaternary carbon atoms. However they all seem to have shortcomings of different kinds: Many of them are not sign-selective, some have long sequences and consequently low sensitivity and some have very complex crosspeaks or antiphase peakshapes that require peak-fitting procedures for extraction of the coupling constants. In any case, the use of experiments for measurement of long-range RDCs is limited and there is a demand for better experiments developed for this purpose, especially experiments yielding correlations to quaternary carbons. Table 2.1 provides a rough overview of some of the many 2D NMR experiments for measurement of $^1J_{CH}$ and other long-range coupling constants in small molecules.

Table 2.1. Overview of some of the most important NMR experiments for measurement of long-range homo- and heteronuclear $J$ coupling constants (and RDCs). $C_q$: quaternary carbon atoms. Inspired by.

<table>
<thead>
<tr>
<th>Coupling constants*</th>
<th>Experiment name/type</th>
<th>Principle for coupling constant determination, with comments</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^1J_{CH}$, $^nJ_{HH}$</td>
<td>XLOC</td>
<td>E.COSY</td>
<td>76–78</td>
</tr>
<tr>
<td>$^1J_{CH}$, $^nJ_{CH}$</td>
<td>HETLOC</td>
<td>E.COSY</td>
<td>81,88</td>
</tr>
<tr>
<td>$^nJ_{HH}$</td>
<td>HECADE</td>
<td>E.COSY. Can measure $^nJ_{CH}$ to $C_q$</td>
<td>82,89</td>
</tr>
<tr>
<td>$^1J_{CH}$, $^nJ_{CH}$, $^{n+1}J_{HH}$</td>
<td>C-HETSERF</td>
<td>E.COSY. Complicated cross peaks</td>
<td>53,90</td>
</tr>
<tr>
<td>$^2J_{CH}$, $^3J_{CH}$</td>
<td>SJS-HSQC</td>
<td>E.COSY</td>
<td>91</td>
</tr>
<tr>
<td>$^1J_{CH}$, $^nJ_{CH}$, $^{n+1}J_{HH}$</td>
<td>P.E.HSQMBC</td>
<td>E.COSY</td>
<td>92</td>
</tr>
<tr>
<td>$^nJ_{CC}$</td>
<td>J-M-ADEQUATE</td>
<td>IP, read in F1</td>
<td>93</td>
</tr>
</tbody>
</table>

*For experiments where determination of more than one coupling is possible, F1 is mentioned before F2.

| $^nJ_{CH}$ | Phase sensitive HMBC | IP (fitting needed) | 94–98 |
| $^nJ_{CH}$ | HSQC-TOCSY | IP or IPAP. No $^nJ_{CH}$ to $C_q$ | 99–104 |
| $^nJ_{CH}$ | J-HMBC | J-resolved | 78,86,105–110 |
| $^nJ_{CH}$ | HSQMBC | IP, AP (fitting needed), IPAP, J-resolved or E.COSY | 92,111–122 |
| $^nJ_{CH}$ | HSQMBC-COSY/TOCSY | IPAP. Sign-sensitive. Can measure $^nJ_{CH}$ to $C_q$ | 123,124 |

2.5 HMBC+, a new NMR experiment for measurement of $^{n+1}J_{HH}$
A new NMR experiment was developed for measurement of long-range homonuclear coupling constants. It is based on two existing HMBC-type experiments from the Carlsberg.
Laboratory (Valby, Denmark), the multiplicity-edited HMBC and the homonuclear J attenuated (HAT) HMBC by Benie, Nyberg and Sørensen. The pulse sequences for the two experiments are shown in Figure 2.6 and Figure 2.7, respectively.

a) Improved edited HMBC standard

b) Improved edited HMBC -

c) Improved edited HMBC +

Figure 2.6. Multiplicity edited HMBC pulse sequence with a third-order low-pass J filter. Filled and open bars refer to π/2 and π pulses, respectively. The dashed open boxes represent 13C decoupling. τ = (2 \( J_{CH} \))⁻¹, δ corresponds to a gradient delay, ε = 2 \( t_{\Delta/\min} \) + t(π), \( \epsilon' = \epsilon + \pi(\pi) \), \( \tau_1 = \frac{1}{2} \left( J_{\min} + 0.07 \left( J_{\max} - J_{\min} \right) \right) \), \( \tau_2 = \left( J_{\max} + J_{\min} \right) \), \( \tau_3 = \frac{1}{2} \left( J_{\max} - 0.07 \left( J_{\max} - J_{\min} \right) \right) \). φ₁ = \( \{ x, -x, -x, x \} \), φ₂ = \( \{ x, x, 4(\mathbf{x}), x, x \} \), φ₃ = \( \{ 4(x), 4(y), 4(-x), 4(-y) \} \). Δ = \( (2 \hat{J}_{CH})\) is the delay for evolution of long-range couplings.
One of the differences between the two experiments is a $\pi$ phase difference in the multiplet structure with respect to $n+1J_{HH}$. The idea of the new HMBC+ experiment is to utilize this difference by adding and subtracting the multiplicity-edited subspectra from the HMBC and the HAT HMBC and thus separate the multiplet structure into different subspectra to afford accurate measurement of $J$ and $D$. This principle is illustrated in Figure 2.8, which shows the multiplet structures for the multiplicity-edited HMBC and the HAT HMBC (top left corner). Upon addition and subtraction of the data, the multiplet is split in two, so that (in this case) the addition spectrum shows the left-hand $J_{CH}$ splitting and the subtracted spectrum shows the right-hand side. This is processed as magnitude data and when the add and subtract spectra are overlaid (right-hand side of Figure 2.8) both homo- and heteronuclear long-range couplings can in theory be extracted. As a result of the editing, the $J_{HH}$ couplings can be determined very accurately, but for the heteronuclear coupling, in practice only large $^{a}J_{CH}$ are obtainable due to an IP coupling pattern (Figure 2.9). The individual multiplet structures in the original spectra would often become fused when the cross peak fine structures are IP and
in close proximity, and these experiments were not suitable for measurement of small coupling constants. But because of the π phase difference with respect to $n+1J_{HH}$, this coupling constant is antiphase in the HAT HMBC, compared to the edited HMBC, as drawn in Figure 2.8 for a three-spin system of one proton observing a CH group.

![Diagram of coupling constants in HMBC, HAT HMBC, and HMBC+](image)

**Figure 2.8.** Schematic representation of structure and phases of a three-spin CH…H multiplet in HMBC, HAT HMBC and HMBC+. Addition and subtraction (left side) of the HMBC and HAT subspectra yields complimentary phases that when combined (right side) will afford accurate measurement of scalar and dipolar couplings in HMBC+.

The figure assumes that $|J_{CH}| > |J_{HH}|$, however these long-range couplings are comparable in size and therefore the opposite situation also occurs. In cases where $J_{HH}$ is larger than $J_{CH}$ the splitting from $J_{CH}$ will often not be observed as the two lines become one for small coupling constants (they are IP in HMBC+), but the $J_{HH}$ is still measured accurately by the IPAP principle. Stick plot representations of theoretical multiplicity patterns for other spin systems in the case of $|J_{CH}| > |J_{HH}|$ are shown in Figure 2.9. Other alternatives can be analysed by drawing tree diagrams of the coupling patterns. The figure is read one row at a time from left to right, showing first the edited HMBC and HAT HMBC theoretical cross peak pattern
(stick plots, in blue), then the add (+) and subtract (-) subspectra created by linear combination of the first two spectra. Here it can be seen that $nJ_{CH}$ gives rise to a splitting in the add/subtract subspectra. The last column represents the overlaid add/subtract spectra in magnitude mode, where it is possible to extract $n+1J_{HH}$ as the distance between individual lines in the two subspectra. Especially for diastereotopic protons, the coupling constant determination can be difficult when the two $n+1J_{HH}$ are similar in size, as the two middle lines may not be separated. In this situation, the coupling constants cannot directly be determined accurately (fourth row, Figure 2.9).

![Figure 2.9. Stick plot examples of multiplet structures in HMBC (HMBCED), HAT HMBC and HMBC+ when $|J_{CH}| > |J_{HH}|$. Green represents the add subspectrum, while black represents the subtract subspectrum. By calculating the absolute value spectra (magnitude) and overlaying the add and subtract spectra, long-range homonuclear couplings can be extracted. Direct extraction of $n+1J_{HH}$ (and $D$) for diastereotopic methylene protons (second last row) may be difficult in HMBC+ when the $J_{HH}$ coupling constants are of similar size.

### 2.5.1 The editing procedure

The add and subtract subspectra are created by linear combinations of the datasets from the HMBC and the HAT HMBC. This procedure is in principle similar to the multiplicity editing procedure used in the original experiments where updown- and updown+ (the sign indicates if the $^{13}C \pi$ pulse is before or after the $^1H \pi$ pulse) spectra are combined to produce subspectra according to even/uneven number of protons. To minimize spectral overlap...
the HMBC+ is created from the multiplicity-edited spectra, thus creating four new subspectra from the four initial spectra (Figure 2.10).

For this kind of editing to be successful the pulse sequences must undergo a few changes. In order to obtain comparable resonance intensity and correct multiplet structures, four conditions must be met:
1. The $^{13}$C chemical shift must be refocused for $t_{1,\text{min}}$ between the two $^{13}$C $\pi/2$ pulses
2. Any sample heating must be similar for all sequences
3. The sequences must have the same evolution time for $^1$H chemical shifts as well as $J_{\text{HH}}$ and $J_{\text{CH}}$
4. The sequences must have the same total length and must also have the same timing between the initial $^1$H $\pi/2$ pulse and acquisition of the FID.

Fulfilling these requirements can only be done by taking them all into account at the same time, as the solution to one condition is strongly interlinked with the rest. The first condition (1) is already met in the original experiments due to the $\varepsilon$ delay in the sequences, which ensures that the $^{13}$C $\pi$ pulse for $t_{1,\text{min}}$ is placed exactly between the two $^{13}$C $\pi/2$ pulses. Regarding the timing between the $^1$H $\pi/2$ pulse and the end of the sequence (4), there is a difference of $\tau$, which is what generates the $\pi$ phase difference between the experiments. Decoupling periods can cause sample heating (2), and therefore the decoupling of length $\tau + \varepsilon'/2$ in the beginning of the HMBC should be added to the HAT sequence as well. HAT HMBC is $\tau$ shorter than the multiplicity edited HMBC (4), and this must therefore be added to HAT HMBC some time after the first $^1$H pulse (4). Prior to this pulse the HMBC +/-
sequences have the $\tau + \varepsilon'/2$ long decoupling period, and therefore this delay must be added to the HAT HMBC as well, this time in front of the $^1H \pi/2$ pulse (3 and 4). Ensuring the exact same pulse sequence lengths also includes counting and adjusting for all the small correctional delays which ensure that the hardware has time to switch between e.g. gradients and pulses. The resulting pulse sequences for the HMBC+ experiment are shown in Figure 2.11. The HMBC sequences are unchanged, whereas the HAT HMBC +/- have undergone some changes compared to the original sequences.

The decoupling period of $\tau + \varepsilon'/2$ is inserted around the $^1H \pi/2$ pulse. After the first pulse a decoupling period of $\tau/2$ is used, as we need to add $\tau/2$ in both ends of the HAT experiment to lengthen it by $\tau$, and $^{13}C$ decoupling in the first $\tau/2$ period ensures that evolution of $J_{CH}$ is not affected. Another $\tau/2$ period is added in the end of the HAT sequences, ensuring that $J_{HH}$ is refocused and the total time for all the sequences are exactly the same. The HMBC experiment contains three pulse sequences, the standard, minus (-) and plus (+). This was an implementation by Benie and Sørensen to enhance the editing procedure from the first multiplicity-edited HMBC experiment. Processing of these complex datasets is usually done with a program containing a list of user defined processing commands. In Topspin (Bruker Biospin) they are called AU programs and may contain standard commands and C statements, functions and variables. In the improved multiplicity edited HMBC by Benie and Sørensen the updown+/- sequences (Figure 2.6 b and c) were acquired in an interleaved fashion with the standard HMBC experiment (Figure 2.6 a). The editing procedure was achieved by linear combination corresponding to $2*\text{standard} \pm [(\text{updown-}) + (\text{updown+})]$. In the case of five sequences and four (directly) resulting spectra in HMBC+, the solution was to ensure that the data is recorded in an interleaved fashion where the HMBC updown+/- are cycled half as much as the HMBC standard sequence and the HAT +/- sequences. In practice this was done by acquiring the experiment as a pseudo-3D experiment with four planes from the four pulse sequences recorded in the experiment, where plane #1 is HMBC std., plane #2 is interleaved HMBC+ and HMBC-, plane #3 is HAT-, and plane #4 is HAT+. This distribution ensures that the five sequences are distributed into four slices, two for HMBC and two for HAT HMBC.
a) Improved edited HMBC standard

b) Improved edited HMBC -

c) Improved edited HMBC +
2 New NMR experiments

Figure 2.11. HMBC+ pulse sequence containing the modified sequences from the multiplicity-edited HMBC and the HAT HMBC (includes previous page). The experiment uses a third-order low-pass $J$ filter. Filled and open bars refer to $\pi/2$ and $\pi$ pulses, respectively. The dashed open boxes represent $^{13}$C decoupling. $\tau = (2^{1/2})J_{CH}, \delta$ corresponds to a gradient delay, $\epsilon = 2^{1/2}(2 \cdot J_{CH} + t(\pi/2)), \epsilon' = t(2^{1/2})(J_{CH} + t(\pi/2)), \tau_1 = (J_{CH} + t(\pi/2)), \tau_2 = (J_{CH} + t(\pi/2)), \tau_3 = (J_{CH} + t(\pi/2)), \phi_1 = \{x, -x, -x, x\}, \phi_2 = \{x, x, 4(-x), x, x\}, \phi_3 = \{4(x), 4(y), 4(-x), 4(-y), 4\}$. $\Delta = (2^{1/2})J_{CH}$ is the delay for evolution of long-range couplings. See the corresponding pulse program in the Appendix.

Instead of running all five sequences interleaved in 2D and then adjusting for the resonance intensity by dividing the HMBC+/ by 2, this both saves experiment time and minimizes artefacts. A small correctional factor of 1.05 (also used in the edited HMBC) is however needed as the resonance intensity of the standard HMBC is slightly higher than for the HMBC+/ due to the value of $|\cos(\phi)|$ being slightly smaller for the +/- experiments, because the angle $\phi$ is close to $\pi$, but slightly smaller or larger for HMBC+/.-126 The pseudo-3D scheme furthermore enables easy processing of each of the four slices. To create HMBC+ the AU program is however necessary. One of the main challenges for the AU program was caused by the fact that the HAT HMBC subspectrum #2 required a $\pi/2$ phase correction in F1 to give absorption lineshapes. This meant that just adding and subtracting the raw HMBC and HAT HMBC data would be done with a phase mismatch and therefore not give the
correct result. Therefore a $\pi/2$ phase correction in $t_1$ of HAT prior to Fourier transformation was required to add and subtract the data properly. This was done by taking advantage of the fact that both real and imaginary data are recorded, and while the real data is in absorption the imaginary data is in dispersion, due to a $\pi/2$ phase difference caused by the two detectors being placed orthogonally in the probe. This is what conventionally allows phase corrections on the frequency domain data, and by swapping the real and imaginary data as well as changing the sign of the new imaginary data as described by Palmer III et al.,$^{130}$ the desired phase shift was achieved in the time domain data of the mentioned subspectrum, allowing the editing procedure to run without phase mismatch (for pulse program and AU program, see the Appendix).

### 2.5.2 Determination of coupling constants

Then linear combination of the data as presented in Figure 2.10 yields the HMBC+ subspectra which are processed by the AU program to give the subspectra in magnitude mode in the F2-dimension. overlays of the add and subtract spectra enables determination of $^{n+1}J_{HH}$ from the displacement of the two spectra as it was illustrated in Figure 2.9. Because the coupling constants are determined using this editing procedure, there is no lower boundary for the size of the $J_{HH}$ coupling constant to be determined. In practice the coupling constant is measured by taking slices through the spectral rows (F2) in a multiplet and displacing the subtract spectrum onto the add subspectrum. In this order the sign of the coupling is measured correctly.

The small organic molecule vinyl acetate was used to test the experiment (Figure 2.12), as it is a simple molecule with the large $^3J_{HH}$ couplings of the vinyl group appearing in both the even (C/CH$_2$) and uneven (CH/CH$_3$) subspectra, and in addition the $^2J_{CH}$ couplings are known.$^{131,132}$ Four HMBC+ cross peaks of vinyl acetate are shown in Figure 2.12 along with the respective 1D slices. The blue lines correspond to the add subspectrum and the red lines are the subtract subspectrum (cf. Figure 2.10). The subtract spectra have been shifted slightly upfield in F1 for purely visual reasons, due to the complicated coupling pattern of the H3-C4 cross peak. The NMR assignment of vinyl acetate is shown in Table 2.2, along with the coupling constants measured from the HMBC+ crosspeaks, which are in excellent agreement with the coupling constants from the $^1$H 1D spectrum (Figure 2.12).
Figure 2.12. Structure of vinyl acetate with atom numbering in grey (top) and cross peaks with F1-slices from HMBC+ on vinyl acetate (DMSO-d$_6$, 400 MHz, 16 scans, 4096 data points, 128 increments, zero filled to 1024 in F1). Each subspectrum was scaled separately and the subtract subspectrum (red) was shifted upfield to make it easier to see the individual components.

Table 2.2. NMR assignment table for vinyl acetate, including coupling constants measured with HMBC+, in a coupled $^{13}$C spectrum (400 MHz, 0.185 Hz FID resolution) and by Crecey et al. in 1971.\textsuperscript{131}

<table>
<thead>
<tr>
<th>#</th>
<th>$^1$H chemical shift [ppm], $J_{HH}$</th>
<th>$^1$C-chemical shift [ppm]</th>
<th>$^3$J$_{HH}$ [Hz] HMBC+</th>
<th>$^3$J$_{CH}$ [Hz] HMBC+</th>
<th>$^3$J$_{CH}$ [Hz] coupl. $^{13}$C</th>
<th>Ref.\textsuperscript{131}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.11, 3, s</td>
<td>20.4</td>
<td>-</td>
<td>6.8</td>
<td>7.0</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>167.8</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>7.19, 1, dd, 14.0, 6.3</td>
<td>141.2</td>
<td>+14.04, +6.28</td>
<td>9.4</td>
<td>9.6</td>
<td>+9.65</td>
</tr>
<tr>
<td>4a</td>
<td>4.63, 1, dd, 6.3, 1.5</td>
<td>97.9</td>
<td>+6.25</td>
<td>6.8</td>
<td>7.7</td>
<td>-7.9</td>
</tr>
<tr>
<td>4b</td>
<td>4.88, 1, dd, 14.0, 1.5</td>
<td>97.9</td>
<td>+14.06</td>
<td>7.4</td>
<td>7.7</td>
<td>+7.6</td>
</tr>
</tbody>
</table>

The measurement of $^{n+1}$J$_{HH}$ couplings for spin systems with one active $J_{HH}$ is straightforward, as the coupling constant is determined from the splitting in F2 between the components of the add and subtract subspectra (mid section of Figure 2.12). With no active $J_{HH}$ couplings as seen for quaternary carbons, there is no HMBC+ editing and the entire cross peak ends up in the add subspectrum as a HMBC correlation. For the H3-C4 cross peak in vinyl acetate, two active $^3$J$_{HH}$ to H4a and H4b are involved, and the coupling pattern becomes slightly more complicated. For spin systems with large coupling constants where the individual peaks are well resolved this is not a problem, and the couplings can be analysed, e.g. by drawing tree diagrams (Figure 2.13).\textsuperscript{133} The couplings in the H3-C4 multiplet are most accurately determined by displacing the rightmost red line (#8) onto the
rightmost blue line (#7) to give the small coupling constant, and the other is found by continuing the displacement so that the left blue line (#2) intercepts the third red line from left (#6). If the two middle lines #4-5 were separated in the 2D spectrum, the lines #5 and #8 could have been used to give the same result. In addition to the homonuclear coupling constants, the size of four $^{2}J_{CH}$ were determined, however since the $^{8}J_{CH}$ can only be measured in-phase, this is not suitable for small heteronuclear couplings. Interference from passive $J_{HH}$ couplings should also be taken into consideration, e.g. a HMBC correlation between H3 and C2 shows a dd shape (data not shown) originating from the $^{3}J_{HH}$ couplings to H4a-b while the $^{3}J_{CH}$ is not resolved. The $^{2}J_{CH}$ couplings measured in vinyl acetate by HMBC+ are slightly smaller than those measured from $^{1}H$-coupled $^{13}C$ spectra.$^{131}$ Even these large couplings in vinyl acetate are estimated too small, which shows why the IP method generally does not yield accurate results for small coupling constants.

Figure 2.13. Tree diagrams for the HMBC+ cross peaks from Figure 2.12 of vinyl acetate (with approximate values of $J$ in Hz). The editing happens by the action of the active long-range $J_{HH}$ coupling constant, as highlighted by the coloring of the bottom lines. Because the H3-C4 cross peak contains two active homonuclear couplings, the coupling pattern becomes symmetric. The geminal coupling H4a-H4b is a passive coupling in the HMBC experiment, making it IP like the $^{3}J_{CH}$ couplings and it is therefore not measured.

Of the coupling constants shown in the tree diagrams, only the geminal coupling between H4a and H4b could not be determined. Because this is a HMBC-type experiment it cannot be used to measure geminal coupling constants, but is more suitable for long-range $J_{HH}$ coupling constants. To test the HMBC+ experiment on a complex molecule with coupling constants in all ranges, including negative values of $J_{HH}$, the plant alkaloid strychnine was used. Strychnine has undergone extensive investigation by NMR spectroscopy and a number of
coupling constants have been measured and calculated.\textsuperscript{134–136} Figure 2.14 shows three examples of HMBC+ cross peaks with two positive $^3J_{\text{HH}}$ and one small and negative $^4J_{\text{HH}}$ (see the full spectra and NMR assignment in the Appendix).

Using HMBC+, 28 $^{n+1}J_{\text{HH}}$ coupling constants in strychnine were determined sign-selectively and compared to couplings from the literature\textsuperscript{134,136} as well as coupling constants from DFT calculations. Starting from the crystal structure of strychnine\textsuperscript{137} (PDB entry SY9) the structure was energy minimized using B3LYP/6-31G(d) in chloroform. $J$ coupling constants were calculated through gas-phase single point calculations in accordance with recommendations by Bally and Rablen.\textsuperscript{138} This was done in Gaussian\textsuperscript{139} with the B3LYP functional and a 6-31G(d,p)+1s basis set (Table 2.3, calculations by Casper Hoeck, DTU Chemistry). Two other basis sets were examined for the single point calculation (6-311+G and cc-pVDZ), but 6-31G showed the best correlation with experimental data (data not
shown). The u+1s extension is short for “uncontracting and augmenting by extra compact 1s functions”. Interestingly, the best correlation between experiment and calculations was obtained when using Fermi contact (FC) terms only, giving a root mean square deviation (RMSD) of 0.40 Hz, while the full coupling constant term gave an RMSD of 0.47 Hz (data not shown). Completely in line with recommendations by Bally and Rablen, a linear scaling factor of 0.9117 gave the best fit with an RMSD of 0.28 Hz (Table 2.3). The FC term is the dominating contribution to the coupling constant, taking into account the magnetic interactions of the nucleus with electrons at the site of the nucleus. The other terms are the diamagnetic and paramagnetic spin-orbit and spin-dipole terms, which partly cancel each other out and thereby have smaller contributions to the $^1\text{H}^1\text{H}$ scalar couplings. Calculation of $J$ coupling constants were performed in gas-phase, as adding implicit chloroform had no significance for the result and RMSD in accordance with Bally and Rablen. The correction factor is a linear fitting that scales the values by a fixed factor, found by linear regression analysis on a large training set of small molecules. This is a cheap and easy way to obtain high accuracy results with minimal cost in computational power. The scaling factor is specific for a basis set and evens out systematic errors in calculations. Cobas et al. reported $J_{1\text{H}1\text{H}}$ coupling constants using several methods to extract couplings from 1D data while developing an automatic coupling constants analyzer (ACCA) for MestReC. Using these extracted couplings as a supplement to the extensive work by Carter et al. from 1974, some of the HMBC+ coupling constants could be compared to experimental reference values. Ten of the measured coupling constants had not previously been determined experimentally, and could therefore only be compared to calculations. Carter et al. used incorporation of deuterium to facilitate the measurement of $J_{1\text{H}1\text{H}}$ coupling constants in complicated 1D $^1\text{H}$ multiplets. The coupling constants measured by HMBC+ are in excellent agreement with both references values and calculations, and all sizes of coupling constants correlate very well with predicted values (Figure 2.15). H17ab is a second order multiplet system in the $^1\text{H}$ NMR spectrum, however a coupling between H17ab and H16 was measured. Since the chemical shift of H17a and H17b could not be discriminated and only one correlation was observed, it was assumed to be an average coupling constant, keeping in mind that special care must be taken when second order effects are at play. However it turned out that the measured coupling constant fit exactly to the combined values of $J_{1\text{H}1\text{H}1\text{H}}$ and $J_{1\text{H}1\text{H}1\text{H}}$ and therefore it seemed correct. Further work will clarify if this is a correct assumption. Another coupling constant between H20b and H22 was measured to 1.20 Hz by Carter et al., but HMBC+, ACCA, and calculations all agreed on a
coupling constant of approximately -0.6 Hz (although for ACCA, the sign was of course not determined), and therefore this seems to have been a mistake in the original reference.\textsuperscript{136}

Table 2.3. Experimental and calculated $^{\text{a}}J_{\text{HH}}$ coupling constants of strychnine, where experimental HMBC+ coupling constants were measured and compared to reference coupling constants.\textsuperscript{134,136} Calculations were performed with FC only using B3LYP/6-31G(d,p)u+1s and then linearly scaled by a factor of 0.9117. *H7\textsuperscript{ab} was seen as a broad singlet and the measured coupling was assumed a average. ** Carter \textit{et al.}\textsuperscript{136} determined this coupling to 1.20 Hz, however this seems like a mistake and therefore the ACCA coupling was used for comparison. For more calculated coupling constants, see an expanded table in the Appendix.

<table>
<thead>
<tr>
<th>#H1</th>
<th>#H2</th>
<th>Experimental HMBC+</th>
<th>Experimental $^{1}\text{H}$ NMR\textsuperscript{136, b: 134}</th>
<th>Calculated FC only, scaled</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>1.17</td>
<td>1.18</td>
<td>1.13</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>0.50</td>
<td>0.45</td>
<td>0.55</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>1.09</td>
<td>1.12</td>
<td>0.97</td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>-0.19</td>
<td></td>
<td>-0.10</td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>10.49</td>
<td>10.47</td>
<td>9.96</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>-0.70</td>
<td></td>
<td>-0.27</td>
</tr>
<tr>
<td>11a</td>
<td>12</td>
<td>8.34</td>
<td>8.44</td>
<td>8.43</td>
</tr>
<tr>
<td>11a</td>
<td>13</td>
<td>-0.42</td>
<td></td>
<td>-0.31</td>
</tr>
<tr>
<td>11b</td>
<td>12</td>
<td>3.24</td>
<td>3.28</td>
<td>3.75</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>3.20</td>
<td>3.12</td>
<td>3.58</td>
</tr>
<tr>
<td>12</td>
<td>15b</td>
<td>0.40</td>
<td></td>
<td>0.02</td>
</tr>
<tr>
<td>12</td>
<td>23a</td>
<td>-0.08</td>
<td></td>
<td>-0.02</td>
</tr>
<tr>
<td>13</td>
<td>14</td>
<td>2.97</td>
<td>3.10</td>
<td>3.27</td>
</tr>
<tr>
<td>13</td>
<td>15a</td>
<td>0.34</td>
<td>0.49-0.59\textsuperscript{b}</td>
<td>0.38</td>
</tr>
<tr>
<td>13</td>
<td>15b</td>
<td>-0.30</td>
<td></td>
<td>-0.28</td>
</tr>
<tr>
<td>14</td>
<td>15a</td>
<td>5.42</td>
<td>4.58</td>
<td>4.60</td>
</tr>
<tr>
<td>14</td>
<td>15b</td>
<td>2.03</td>
<td>1.98</td>
<td>1.93</td>
</tr>
<tr>
<td>14</td>
<td>16</td>
<td>0.76</td>
<td></td>
<td>0.54</td>
</tr>
<tr>
<td>14</td>
<td>20b</td>
<td>-0.58</td>
<td></td>
<td>-0.50</td>
</tr>
<tr>
<td>14</td>
<td>22</td>
<td>-2.89</td>
<td>2.82\textsuperscript{b}</td>
<td>-3.03</td>
</tr>
<tr>
<td>15a</td>
<td>16</td>
<td>3.97</td>
<td>4.11</td>
<td>3.89</td>
</tr>
<tr>
<td>15b</td>
<td>16</td>
<td>2.29</td>
<td>1.82</td>
<td>2.02</td>
</tr>
<tr>
<td>16</td>
<td>17a</td>
<td></td>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td>16</td>
<td>17b</td>
<td></td>
<td></td>
<td>-0.18</td>
</tr>
<tr>
<td>16</td>
<td>17ab\textsuperscript{*}</td>
<td>-0.14</td>
<td></td>
<td>-0.14</td>
</tr>
<tr>
<td>16</td>
<td>20b</td>
<td>-0.35</td>
<td></td>
<td>-0.33</td>
</tr>
<tr>
<td>20a</td>
<td>22</td>
<td>-1.37</td>
<td>1.47\textsuperscript{b}</td>
<td>-1.82</td>
</tr>
<tr>
<td>20b</td>
<td>22</td>
<td>-0.63</td>
<td>0.63\textsuperscript{b, c}</td>
<td>-0.58</td>
</tr>
<tr>
<td>22</td>
<td>23a</td>
<td>6.08</td>
<td>5.74</td>
<td>6.14</td>
</tr>
<tr>
<td>22</td>
<td>23b</td>
<td>7.00</td>
<td>6.88</td>
<td>6.86</td>
</tr>
</tbody>
</table>

Only the full term calculation (see the Appendix) gave rise to coupling constant errors of more than 1 Hz, and the signs of the coupling constants are also in perfect agreement with the calculations.
Figure 2.15. Plot of coupling constants in strychnine measured by HMBC+ against FC only calculated couplings with linear scaling factor 0.9117 using B3LYP/6-31G(d,p)//u+1s (▲, black line is linear regression with $R^2 = 0.992$) and against the experimental reference coupling constants in Table 2.3 (●, compared as absolute values, therefore only showing positive coupling constants). The grey lines define an area of perfect fit ± 1 Hz between HMBC+ and experimental/calculated coupling constants, which all data points fall within.

2.5.3 Application in oligosaccharide analysis

HMBC+ has many prospective applications for coupling constant analysis and use thereof. In the last section it was shown how the experiment can be used to measure very accurate coupling constants sign-selectively in vinyl acetate and in the plant alkaloid strychnine, showing the potential of HMBC+ for investigation of relative configuration in small molecules. In oligosaccharide analysis, the experiment has two convenient applications: The $n+1J_{HH}$ splitting in the HMBC cross peak supports the assignment of the stereochemical relationship between the protons, due to the fact that the vicinal coupling $3J_{HH}$ in pyranose rings is largest for axial-axial protons and smallest when both protons are equatorial.\textsuperscript{144,145} For complex oligosaccharides it can be difficult to elucidate neighboring CH pairs when resonance overlap makes the COSY spectrum practically useless. In HMBC+ the $nJ_{CH}$ correlations are combined with the $n+1J_{HH}$ coupling constants to help assign the correct stereochemical relationship and thus the monomer identity. This is in a way analogous to the advantage of the HSQC-TOCSY experiment over TOCSY, as the spectral dispersion of $^{13}C$ is highly advantageous and especially $3J_{HH}$ is vital for determination of monomer identity. The combination of $^{1}H$-$^{13}C$ long-range $nJ_{CH}$ correlations with the measurement of $n+1J_{HH}$ is
therefore believed to facilitate easier oligosaccharide analysis. Furthermore, the experiment can be used to measure $^4J_{HH}$ across glycosidic bonds, which may aid in conformational studies. These coupling constants are very small in size and often negative, and therefore have so far not been reported for structural studies of carbohydrates, although Otter and Bundle reported long-range interglycosidic homonuclear correlations over 4 and 5 bonds via a gCOSY experiment.\textsuperscript{146} Currently, most often a combination of distance calculations with NOESY/ROESY-type experiments and other coupling constants are used, e.g. $J_{CH}$ or $J_{CC}$, also in combination with alignment for measurement of RDCs.\textsuperscript{147–149}

A high degree of resonance overlap is often seen in oligosaccharides and therefore it is necessary for at least one of the $^3J_{CH}$ correlations to be non-overlapped and unique to measure the correct coupling constant, and here it is advantageous that the $^4J_{HH}$ coupling constant can be found by observing either of the two proton resonances. E.g. the trans-glycosidic $^4J_{HH}$ in LNT between GlcNAc H3 and Gal2 H1 can be determined (-0.3 Hz) from H3 to H1, however in the opposite direction there is an ambiguity between the trans-glycosidic $^4J_{HH}$ and an intra-residual $^4J_{HH}$ in Gal1 (Figure 2.16).

Figure 2.16. Partial assignment of LNT, where resonance overlaps for two protons at 4.42 ppm and two carbon atoms at 82.8-82.9 ppm make it difficult to differentiate the $^4J_{HH}$ (red chemical shifts). The coupling constant of -0.3 Hz across the glycosidic bond between GlcNAc and Gal2 can however be measured from H3 to H1 (green chemical shifts) using the new HMBC+ experiment. See the CH/CH$_3$ subspectrum in the Appendix.

Further studies will show the usefulness of HMBC+ in its various applications; however experiments so far demonstrate a very robust experiment with good sensitivity for very accurate determination of long-range $J_{HH}$ coupling constants in small molecules. Initial results have shown that HMBC+ can be used to measure RDCs in aligned samples, but these results will be treated elsewhere.

### 2.6 Experiments for measurement of $^2J_{CH}$

Long-range heteronuclear coupling constants are also highly desired in structural studies, and especially $^2J_{CH}$ couplings to quaternary carbon atoms yields important structural information.
In collaboration with Ole W. Sørensen we wanted to pursue the sign-selective measurement of $2J_{CH}$, using double-quantum (DQ) coherence ADEQUATE-type experiments to gain access to this coupling constant. This experiment could find use both for isotropic and aligned samples, and measurement of the sign of $2J_{CH}$ was important, as this type of coupling constant is often negative, but it is not always the case.$^{132}$

Several $^1$H-detected $^{13}$C-$^{13}$C DQ experiments could form the basis of the new experiment.$^{75}$ Weigelt and Otting’s $^1$H-detected INEPT-INADEQUATE experiment uses proton magnetization both for excitation and detection,$^{150}$ giving it an advantage over the original INADEQUATE experiment which suffered from very low sensitivity since only $^{13}$C-$^{13}$C pairs were detected.$^{151}$ In natural abundance this meant that only one in 8,000 molecules gave rise to measurable magnetization.$^{152}$

With the $^1$H-detected INEPT-INADEQUATE (Figure 2.17), later known as 1,1-ADEQUATE due to the use of $^1J_{CH}$ and $^1J_{CC}$, the S/N ratio was improved about 13-fold compared to the 2D INADEQUATE experiment.

![Figure 2.17. Original 1,1-ADEQUATE pulse sequence by Weigelt and Otting.$^{150}$ Filled and open bars refer to $\pi/2$ and $\pi$ pulses, respectively. The dashed open box represents $^{13}$C decoupling. $\tau_1 = (4^1J_{CH})^{-1}$ and $\tau_2 = (4^1J_{CC})^{-1}$. $\Delta$ corresponds to a gradient delay and $\phi = \{x, y, -x, -y\}$. A slightly better phase cycle can be employed with $\phi = \{x, y, -x, -y, y, -x, -y, x, -y, x, y, -x, y, x, y, -x\}$ and cycling the $^{13}$C $\pi$ pulse immediately after $t_1$ with $\phi = \{4(x), 4(y), 4(-x), 4(-y)\}$. The gradient ratios are 1:1:1:0.004.](image) While the INADEQUATE experiment enabled direct connectivities between neighbouring $^{13}$C nuclei, the 1,1-ADEQUATE yielded cross peaks between a proton and a $^{13}$C DQ coherence corresponding to $\delta_{C1} + \delta_{C2}$, where C1 is the directly bonded $^{13}$C atom and C2 is any neighbouring $^{13}$C (Figure 2.18). This allowed determination of the chemical shift of a neighbor $^{13}$C atom ($\delta_{C2}$). The experiment starts from $^1$H magnetization which through an INEPT sequence (the first two $\tau_1$ delays) is transferred to $^{13}$C (Figure 2.17).
New NMR experiments

Figure 2.18. The J couplings involved in the INADEQUATE, 1,1-ADEQUATE and the new experiment for measurement of \(^2J_{CH}\). The new experiment is also based on the couplings shown for 1,1-ADEQUATE, but the \(^2J_{CH}\) is the aim of the experiment. The red rectangular area symbolizes the DQ coherence which is manifested in a DQ frequency axis in the \(^{13}\)C dimension, F1, with cross peaks between \(\delta_H\) and \(\delta_{C1} + \delta_{C2}\).

This is followed by generation of DQ coherence which evolves during \(t_1\) and in the end magnetization is transferred back to \(^1\)H for detection.

The general idea for the new experiment was to remove the \(^1\)H π pulse in the middle of \(t_1\) (Figure 2.17) to prevent the refocusing of \(J_{CH}\) and \(J_{HH}\). At the same time an improvement to the phase cycle was made, with \(\phi_2\) running a 32-step cycle in steps of 45° or \(\pi/4\) (Figure 2.19).

The new phase cycle makes the experiment less sensitive to adjustment of correct pulse lengths and thereby reduces spectral artefacts, however with a tradeoff of a lower S/N ratio.

Because the experiment uses DQ coherence during the evolution period, a coupling between a proton and two adjacent \(^{13}\)C atoms was expected, with \(J = J_{CH} + \gamma J_{CH}\) appearing along F1.

In case of a CH-CH system, the cross peak should appear as a doublet of doublets with \(J_A = ^1J_{CH1} + ^1J_{CH2}\) and \(J_B = ^1J_{CH2} + ^1J_{CH1}\). Because \(^1J_{CH}\) is easily measured and always positive in sign, the magnitude and sign of \(^2J_{CH}\) could be determined.

Figure 2.19. New pulse sequence based on INEPT-INADEQUATE. Filled and open bars refer to \(\pi/2\) and \(\pi\) pulses, respectively. The dashed open box represents \(^{13}\)C decoupling. \(\tau_1 = (4 ^1J_{CH})^{-1}\) and \(\tau_2 = (4 ^1J_{CC})^{-1}\). \(\Delta\) corresponds to a gradient delay and \(\phi_1 = \{x, y, -x, -y, y, -x, -x, -y, x, y, -y, x, y, -y, x, y, -y, x, y, -y, x, y, -y, x, y, -y, x, y, -y, x, y, -y, x, y, -y, x, y, -y, x, y, -y, x, y, -y, x, y, -y\}\) and \(\phi_2 = \{0, 4, 2, 6, 3, 7, 1, 5, 6, 2, 4, 0, 5, 1, 7, 3, 2, 6, 0, 4, 1, 5, 3, 7, 4, 0, 6, 2, 7, 3, 5, 1\}\) where the numbers correspond to \(\pi/4^*n\). The gradient ratios are 1:1:1.004. See the pulse program (ADEQ11ETGPLR_LKJ3) in the Appendix.
Compared to the INEPT-INADEQUATE experiment, the S/N ratio is expected to be lower, due to the fact that the resonance is dispersed by multiplicity rather than a single peak, in addition to some loss in resonance intensity resulting from the 45° phase cycle. However, the splittings observed were much larger than $^1J_{CH} + ^2J_{CH}$, possibly due to multiplication by the number of protons on each carbon atom. A CH-CH$_3$ spin system in N-methyl ephedrine resulted in a dd-like pattern (observing the CH proton) with splittings of 370 and 124 Hz, and with respective $^1J_{CH}$ of 133.2 and 125.8 Hz determined from a $^1$H-coupled $^{13}$C spectrum, this pointed towards a connection with the number of protons attached to the neighboring $^{13}$C atom. Unfortunately it was difficult to find suitable molecules with known $^2J_{CH}$ to test the exact correlation and the extra splitting created unexpected problems for higher multiplicities, possibly due to interference from unwanted coherence transfer pathways. Splittings in the expected size range were however seen for three spin systems with no $J_{HH}$ coupling, as e.g. H12 in brazilin, where couplings to three different quaternary carbons showed couplings in the correct size range of $J = ^1J_{CH} + ^2J_{CH}$ (Figure 2.20 and Figure 2.21). For DQ CH-CH pairs (C1-C2) and for C-CH$_3$ pairs (C6-C7), the splitting was about doubled, which suggested that not only protons on the neighbor $^{13}$C atom were responsible for the extra splitting. Passive couplings to other neighboring $^{13}$CH pairs were however not at play, as e.g. H1 has one small DQ-coupling (to C1a) and one large DQ-coupling (to C2).

Figure 2.20. Structure of brazilin with numbering.$^{153,154}$ The NMR assignment can be found in the Appendix.
New NMR experiments

It seemed that the center line in the CH-CH multiplet cancelled out, possibly resulting in $J = J_{CH1} + J_{CH2} + J_{CH1}$. This was not an optimal situation, as this type of splitting could not be used to determine either of the $J_{CH}$ coupling constants. To try to overcome this problem, a change to the pulse sequence was made, from INEPT to DEPT magnetization transfer (Figure 2.22 and Figure 2.23).  

$$\phi_1 = \{x, y, -x, -y, y, -x, x, -y, x, y, -y, x, y, -y, x, y, -y, x, y, -y\}$$  
$$\phi_2 = \{y, -x, -y, x, -x, -y, x, y, -y, x, y, -y, x, y, -y, x, y, -y, x, y, -y\}$$  
$$\phi_3 = \{0, 4, 2, 6, 3, 7, 1, 5, 6, 2, 4, 0, 5, 1, 7, 3, 2, 6, 0, 4, 1, 5, 3, 7, 4, 0, 6, 2, 7, 3, 5, 1\}$$

Here, $\theta = 60^\circ$ was used for optimal resonance intensity of CH with some intensity for CH2. See the pulse program in the Appendix (DEPT2LR_LKJ).
The DEPT2LR experiment, as it was tentatively named, now showed the center lines in e.g. the H1-C2 cross peak, although it created a separation between the α- and β-spin states of the observed proton as if the high energy proton could only couple to the low energy carbon states and *vice versa* (Figure 2.24). Some of these cross peak patterns somewhat resembles that of the P.E.HSQMBC.\textsuperscript{92} This type of cross peak however showed different splittings in the upper and lower part. Experimental and calculated coupling constants were compared for menthol and brazilin, but this proved problematic. The $^2J_{CH}$ couplings are usually negative and in the order of 0-10 Hz, and calculated and experimental values therefore had a high chance of being of the same size, however this was not the case for all couplings, and whether the convergence for some couplings was coincidence rather than correct determination was so far not possible to decide. Without experimental data to compare to, it was difficult to say whether the poor fit was due to the basis set not being suitable for calculation of $^2J_{CH}$, and therefore gave a poor correlation, or the experimentally determined coupling constants were simply not correct. Some $^2J_{CH}$ coupling constants have been reported in the literature,\textsuperscript{132} but mostly for proton-rich systems, and a suitable model system with known coupling constants was not found. Furthermore, the cross peak patterns were not fully understood, and with a low S/N experiment demanding much NMR time for a thorough investigation, the project seemed difficult to finish with the current tools available.
further attempts of pulse sequence modifications were attempted, but no further conclusions were reached at this point.

Verdier et al. measured $^2J_{\text{CH}}$ coupling constants in menthol, however only three such coupling constants could be compared between the experiments, and H1-C2 was determined by DEPT2LR to +1.7 Hz (vs. 1.8 Hz) and H2eq-C1 was determined to -3.5 Hz (vs. 3.6 Hz), however the last coupling H8-C4 was far off, determined to -4.4 or -1.3 Hz (vs 7.6 Hz). Numbering was adopted from the reference. There are other examples of experiments for measurement of $^2J_{\text{CH}}$, however more data needs to be acquired using both experiments before further comparisons can be made.

The DEPT2LR experiment in its current form has a narrow applicability for the determination of $^2J_{\text{CH}}$ coupling constants to quaternary carbon atoms; perhaps also for CH-CH spin systems. It is inherently insensitive and the fact that coupling constants need to be extracted from the indirect dimension makes the experiment quite demanding regarding NMR time. ADEQUATE-type experiments are therefore most suitable to small molecules in rather high concentration, and therefore it will probably never be an experiment of choice for
the typical natural products sample. Non-uniform sampling (NUS) could be an advantage for these experiments, making them slightly less time demanding. Nevertheless it seems the minimum requirement for putting the experiment to use would be to fully understand the coupling patterns and be able to extract more coupling constants, not only from CH-C_q spin systems.

2.7 Conclusions and perspectives

The new HMBC+ experiment was shown to give excellent results for vinyl acetate and strychnine, especially regarding the sign-selective measurement of \(^{n+1}J_{HH}\), and in some cases \(aJ_{CH}\) coupling constants may be estimated. Interferences from passive \(J_{HH}\) couplings as well as the IP character of the \(J_{CH}\) coupling constants are expected to limit the use of HMBC+ for measurement of heteronuclear coupling constants. HMBC+ however facilitates easy and accurate measurement of long-range \(^{n+1}J_{HH}\) homonuclear coupling constants of any size and sign, based on HMBC and HAT HMBC experiments by Benie, Sørensen, and Nyberg.\(^{125-127}\)

Acquisition of one single 3D experiment and processing using our new AU program creates eight subspectra, of which the first four are edited HMBC and HAT HMBC, for optional use in structure elucidation, and the last four subspectra form the multiplicity edited HMBC+. Both pulse program and AU program can be found in the Appendix. Like e.g. the XLOC experiment\(^{77}\) the determination of \(J_{HH}\) coupling constants is independent of the \(^1H\) linewidth due to the editing procedure developed, facilitating IPAP-type measurement in the direct dimension with high resolution. Compared to existing experiments where measurement of \(J_{HH}\) is possible, this is the first to directly determine both magnitude and sign of small long-range coupling constants. Simple IPAP-type editing eliminates problems with resolution preventing the extraction of coupling constants, and the possibility to find the same coupling constant in two different cross peaks maximizes the number of obtainable coupling constants in samples with spectral overlap. Ten new \(^{n+1}J_{HH}\) coupling constants were determined for strychnine and shown to correlate very well with calculated values. Because molecular configuration and coupling constants are tightly connected, this is believed to improve future structural studies markedly, supplying a high number of \(^{n+1}J_{HH}\) coupling constants combined with HMBC correlations, and with the sign determination this is a very useful experiment.

The measurement of \(J_{HH}\) couplings from CH\(_x\) to CH and CH\(_3\) groups is straightforward, however for couplings to CH\(_2\), overlap in the multiplet lines can be an issue and understanding the multiplet patterns is more important. Often it is however easier to determine the same coupling constant in the other direction from e.g. CH\(_2\) to CH, as opposed
to from CH to CH$_2$. It would also be possible to implement another level of editing of the CH$_2$ subspectrum by changing the $\tau$ delay and thereby obtain a further separation of the multiplet pattern. In the cases shown so far, most of the coupling constants have been determined from the CH/CH$_3$ subspectra, although it was shown for vinyl acetate that it works for CH$_2$ as well. The coupling constant determination just requires a bit more of the user, and for small coupling constants to CH$_2$, it seems a fitting procedure would be necessary if the coupling constant cannot be determined elsewhere in the spectrum. Since HMBC+ is a long-range HMBC-type experiment, there are some limitations to the $J_{\text{HH}}$ that can be determined. Geminal $J_{\text{HH}}$ coupling constants cannot be measured from HMBC correlations. These couplings are however often large and negative, and therefore easily determined from e.g. a standard $^1$H spectrum. The HMBC+ experiment can be used for measurement of RDCs; however having only started to unravel the potential of the experiment, treatment of aligned samples is next in line, and the experimental measurement of RDCs is not elaborated in this thesis. It should however be noted that samples with rapid $T_2$ relaxation will need a high number of scans to obtain a sufficient S/N ratio, as the resonance dephases during the pulse sequence due to spin-spin relaxation. This unfortunately seems to be an issue for polymer gels, and therefore long experiment time and high sample stability is needed to obtain sufficient data quality. In order to supplement the homonuclear coupling constants from HMBC+ with heteronuclear coupling constants, it was attempted to develop an experiment for sign-sensitive measurement of $J_{\text{CH}}$ coupling constants. The ADEQUATE-based experiments described have however not been fully developed, and problems with “exotic” multiplet structures remain, possibly due to interference from unwanted coherence transfer pathways. So far it seems the experiment can be used for measurement of $J_{\text{CH}}$ coupling constants to quaternary carbon atoms, but with few $J_{\text{CH}}$ experimental couplings to compare to, the accuracy of the method has not yet been determined. As an example, $J_{\text{CH}}$ values in the natural product brazilin were measured, yet calculated values differed several Hz from the measured coupling constants. Whether this was caused by problems with the experiment or with the calculation of $J_{\text{CH}}$ coupling constants in brazilin is difficult to say. It could be that the strained structure in brazilin is not comparable to the training set of the calculation method, and thus pointing towards a problem in the parameterization. Two-bond heteronuclear $^1$H-$^1$H coupling constants have furthermore not gained much attention, except in the SJS-HSQC experiment a few years back,$^{91}$ but the experiment was tested on a natural product not commercially available. We did not test the SJS-HSQC experiment on our
spectrometers yet, nevertheless it is important to obtain experimental $^2J_{\text{CH}}$ for comparison. The experiment under development here will need further work as well as a validation to be of general use, and in its current state the S/N ratio of the ADEQUATE-type experiment is low for measurement of RDCs, however at high concentration and with the most sensitive equipment available the potential of the experiment in that direction should also be evaluated. Generally, for both HMBC+ and ADEQUATE-type experiments, going to higher field strength will increase the spectral intensity for both isotropic and aligned samples.

In this chapter, the experimental work was dedicated to developing new experiments for measurement of $J$ couplings and RDCs, showing examples for isotropic samples measuring $J$ from HMBC+. This naturally only provides relative configurational information over short ranges. However, the number of coupling constants which can now be determined accurately and sign-selectively, especially in combination with other experiments for measurement of $^aJ_{\text{CH}}$, naturally leads to a discussion of whether the $J$ couplings alone, for many small molecules, provides adequate stereochemical discrimination to obtain relative configurations. The best option is without doubt to combine $J$ coupling constants, NOEs, and RDCs. Any of these methods can only be used to obtain relative configuration, and generally require a somewhat locked conformation to be able to use the relative information on the molecular scale. RDCs have so far not found widespread application in research groups not specialized in the area, most likely due to the technical barrier of obtaining aligned samples, and it would be interesting to compare the use of $J$ coupling constants alone to the combination of $Js$ and RDCs for an assortment of small molecules, including diastereoisomers. In many cases RDCs will however be necessary, due to the added global angular restraints. Future studies will show the full potential of the new experiment, and by making the program code available to others online, we hope to see it in use by other research groups as well.
3 Microbial natural products

This chapter evolves around the discovery of natural products and begins with an introduction to natural products and some of the tools and challenges in the identification and analysis of new natural products. The experimental work on the isolation and characterization of secondary metabolites from a marine bacterium and a filamentous fungus is described in Section 3.5.1-3.5.2 and 3.5.3, respectively. Finally, the potential of these sources in today’s drug discovery is discussed.

Two journal articles on bacterial metabolites from a marine bacterium have been published (Paper I and II, see Appendix), and the reader is referred to the respective supplementary materials for experimental details. NMR assignment tables for the fungal metabolites described can be found in the Appendix for chapter 3 and experimental details can be found in the end of the chapter.

3.1 Drug discovery today

Nature has been an important source of medicine for thousands of years, and today natural products (NPs) remain the primary inspiration in drug discovery. In traditional medicine, especially plants have been used for extraction of active principles, and since the discovery of the penicillins and the golden era of antibiotics, microorganisms have been an important source of bioactive molecules.\textsuperscript{155}

All living organisms produce secondary metabolites for intra- and interspecies communication and competition, and the molecules, which also help differentiate between species, are by design optimized in a natural selection process over many years of evolution.\textsuperscript{156} It is said that NPs have so-called privileged structures, meaning that one natural scaffold can give rise to various biological activities with just minor changes to the structure.\textsuperscript{157} Consequently there is a great chance of finding biological activity in a natural extract (a lead molecule), and further modifications on the structure can result in highly specific biological activities with good pharmacological properties and low toxicity.\textsuperscript{158} On the downside, both the discovery of new NPs and the synthesis of analogues in search of a hit molecule are slow processes which are not very desirable for pharmaceutical companies today.\textsuperscript{159} During the last two decades, focus has been turned towards high-throughput
screening (HTS) of combinatorial chemistry libraries, made to resemble NPs but with fewer stereocenters and simpler structural scaffolds to facilitate easy synthesis. Despite this method being prevalent for some years, only one new chemical entity (NCE) has reached market from combinatorial chemistry: The antitumor agent sorafenib approved by the U.S. Food and Drug Administration (FDA) in 2005.\textsuperscript{160} The chemical space covered by combinatorial chemistry is somewhat limited compared to the enormous diversity within NPs, and it is likely for a library synthesized by a given range of reactions to fall outside the drug space.\textsuperscript{157} 36\% of NCEs (1981-2010) are purely synthetic, having no inspiration from NPs, the other 64\% are NPs, NP derivatives, or inspired by NPs (Figure 3.25).\textsuperscript{160,161} It seems that drug discovery and development will continue to require inspiration from nature.\textsuperscript{162,163}

![Figure 3.25. Source of all small molecule drugs approved in 1981-2010. ND: NP derived (28\%), NP (6\%), S*: Synthetic NP mimic or with NP pharmacophore (30\%), S: Purely synthetic (36\%). Modified from\textsuperscript{160}.]

### 3.2 Sources of natural products

The isolation and purification of NPs often demand large amounts of the producing organisms to yield minute amounts of the active substances. One example is the anticancer drug paclitaxel (Figure 3.26), which was originally produced by isolation from the Pacific yew tree \textit{Taxus brevifolia}.\textsuperscript{164} One course of treatment (~2 g paclitaxel) required the bark extracts from six full-grown yew trees, and as it takes more than a century for the trees to mature, it posed a great stress on the natural source and an ethical dilemma on society. During the late stage clinical trials and launch of paclitaxel as the drug Taxol (Bristol-Myers Squibb) in the early nineties, it was produced semi-synthetically by modifications of 10-deacetylbaccatin III (Figure 3.26), extracted from leaves and twigs of a faster growing and better yielding yew species. Today paclitaxel is produced in plant cell cultures of \textit{T. brevifolia}, a more sustainable process with no risk of extinction for the natural source.\textsuperscript{165,166} Several synthetic routes to paclitaxel were developed, but none were suitable for large scale production.\textsuperscript{167,168}
This is a good example of some of the difficulties faced in drug development from NPs, especially for complicated structures in low yields and with limited supply of the producing species. Microbial sources such as fungi and bacteria have the benefits of fast growth and reproduction making them easier to sample and grow. The environmental stress of taking small microbial samples for cultivation is also limited, compared to the collection of larger organisms for direct use in the laboratory.

### 3.3 Analysis of natural products

#### 3.3.1 The supply problem and dereplication

A common problem for microbial NPs is that the secondary metabolites are often isolated in very low yields. The production of a target compound may be improved by testing different growth conditions, e.g. growth medium or by co-culturing with another species, which may even induce expression of silent gene clusters to produce other exciting metabolites.\textsuperscript{169–171}

Furthermore it necessitates that the isolated material is used in the most sensible way possible regarding structure elucidation and biological assays.

In the initial work with a specific fungal or bacterial species, a small culture (agar plate or liquid culture) is grown to assess the potential of finding novel compounds. This investigation is known as dereplication - the efforts to not replicate previous findings.\textsuperscript{172} It is highly undesirable to spend time in the laboratory working on the isolation and structure elucidation of a molecule that later turns out to be known, and therefore often not particularly interesting. The search of biologically active molecules is often coupled to one or more bioassays in bio-guided dereplication,\textsuperscript{173} but the order can be reversed to assessment of the bioactivity after purification and structure elucidation has taken place. In the case of highly active molecules, the first method can sometimes be difficult if the analysis methods are less sensitive than the bioassay, whereas the second method may not result in coupling a
bioactivity to a new molecule due to the often small amounts isolated and essentially unlimited number of bioassays available for testing against different diseases.

3.3.2 NMR spectroscopy of small amounts

NMR spectroscopy is a rather insensitive analysis method compared to e.g. LC-MS and most initial biological screens, but because NMR is nondestructive this is mainly an issue in case of limited sample amounts. Over the years, various technological improvements have increased the sensitivity of NMR spectrometers. One obvious improvement is the increase in field strength, but especially the development of cryoprobe technology has had a large impact on sensitivity, compared to conventional room temperature probes. A state-of-the-art helium-cooled cryoprobe improves the signal-to-noise (S/N) ratio up to four times, whereas a nitrogen-cooled cryoprobe improves the S/N ratio about two times, though the exact improvement is sample dependent. Alternatively the gain in sensitivity for high-field instruments and cryoprobes may enable the use of \(^{13}\)C-\(^{13}\)C correlated experiments such as INADEQUATE, which are inherently insensitive. For NPs and other samples where only limited amounts are available, it is advantageous to decrease the sample volume in the NMR tube. This can be achieved with e.g. 3 mm probes, 1.7 mm probes, or nanoprobes, where the sub-mg sample is spun at the magic angle of 54.7°. Solvent susceptibility-matched Shigemi tubes are also a good choice when only limited amounts are available, although some extra effort may be needed to ensure optimal sample positioning and shimming.

Ideally, as much of the sample as possible should reside within the observe volume of the probe to increase the resonance intensity as much as possible for samples of limited amounts. Hyphenation is another option which optimizes information output and the time spent in the laboratory to obtain an NMR spectrum. Hyphenation in NMR is the coupling of NMR spectroscopy with UV, HPLC, MS and other detectors and can be done with a flow-probe or with SPE cartridges trapping the constituents and enabling longer NMR experiments, often with a high degree of automation. Coupling to biological assays makes hyphenation in NMR a valuable tool in fast screening of NP extracts.

3.3.3 Determination of absolute stereochemistry

NMR spectroscopy is an excellent tool in determination of 2D structural relationships; however in the third dimension it is not optimal. Enantiomers cannot be differentiated by NMR spectroscopy, and assigning absolute configuration to a molecule is an important part of structure determination. Some of the other analytical methods used in this context were described in Chapter 2, including X-ray crystallography and two chemical derivatization
techniques which were used to structure elucidate some of the new molecules: Marfey’s method and Mosher’s esters, *vide infra*.

Marfey’s method was first reported in 1984 for determination of enantiomeric purity.\(^{24}\) For NPs containing amino acids, it is often used to determine absolute stereochemistry.\(^{186-188}\) In short, it involves acid hydrolysis of the peptide into amino acids which are then derivatized using a chiral auxiliary called Marfey’s reagent (1-fluoro-2,4-dinitrophenyl-5-L-alanine amide, FDAA) via a nucleophilic aromatic substitution to yield diastereoisomers, which can be separated by RP-HPLC and analyzed by UV or MS detection (Figure 3.27). By parallel analysis of FDAA derivatives of pure L- and D- amino acids, the amino acid composition of the NP can be elucidated. About 100 μg of peptide is typically used, but the procedure can easily be scaled down for the method to become even more sensitive and thereby attractive to natural products chemists with very little material available. The postreaction mixture is diluted about five times to a volume of 1 ml and only a few μl of this solution is usually needed for analysis by e.g. UHPLC. By using HPLC vials with inserts, a volume of about 100 μl is sufficient, and consequently only 10 μg of peptide would be required.

![Figure 3.27. Reaction of Marfey’s reagent with L- and D-amino acids (AA) to yield two diastereoisomers.](image)

Mosher’s ester analysis is another derivatization method which can be used to determine the absolute stereochemistry of secondary alcohols (or amines, forming amides).\(^{23,189}\) In this case \(S\)- and \(R\)-enantiomers of e.g. Mosher’s acid chloride (\(\alpha\)-methoxy-\(\alpha\)-trifluoromethylphenylacetyl chloride, MTPA-Cl) are used to create two diastereoisomers where the phenyl group induces anisotropic effects on the protons near the stereocenter, respectively shielding and deshielding the substituents.\(^{190}\) Acquiring \(^1\)H NMR spectra of both esters and calculating the chemical shift changes between the \(S\)- and \(R\)-MTPA analogues, the stereocenter can be assigned (Figure 3.28). Due to the presence of a trifluoromethyl group in MTPA, \(^19\)F NMR provides another option to analyze the stereochemistry, in this case only if one substituent of the secondary alcohol (or amine) is notably more bulky than the other.\(^{191}\)
This is because the conformation of the Mosher’s ester has a significant impact on the shielding effects imposed by the carbonyl in MTPA on the trifluoromethyl group (Figure 3.28), and unless one substituent is markedly more bulky than the other this conformation may not be predominant. When the large substituent (L₁) eclipses the large phenyl group it causes a conformational change due to a steric clash. The structure is shifted slightly further towards a staggered conformation, and the CF₃ group is rotated away from the anisotropic deshielding effect caused by the carbonyl, in turn moving the ¹⁹F resonance upfield (Figure 3.28). A problem with using the ¹⁹F NMR method is that the bulkiness of the substituents does not always differ markedly, and in such cases the method should be used with great care. When using MTPA esters it is therefore highly advantageous to use both ¹H and ¹⁹F NMR in the analysis. Another issue is the possibility of misassignments caused by stereochemical confusion: Due to a change in the priority of the substituents of the stereogenic center in MTPA when the chloride is substituted for an ester, the S-MTPA-Cl yields the R-MTPA Mosher’s ester.¹⁹² The absolute configuration of the stereocenter changes without any inversion or direct bonds being broken, and care should be taken to assign the stereochemistry correctly. In the literature there are however several examples failing to consistently stick to the ΔδSR convention, and in stead using ΔδRS, without doubt adding to the confusion.¹⁹³–¹⁹⁵ Even though other similar auxiliaries have been suggested, MTPA is the one most commonly used.
Figure 3.28. Assignment key for Mosher’s ester analysis using MTPA. Top: S- and R-MTPA esters and corresponding Newman projections, showing the steric clash with the benzene ring. Bottom: The chemical shift difference between the S- and R-MTPA esters \( \Delta \delta^{SR} \) determine the stereochemistry of the secondary alcohol using \(^1\)H, \(^19\)F NMR, or both.\(^{196}\)

*According to Cahn-Ingold-Prelog rules
3.4 Natural products from a marine *Photobacterium*

Covering more than 70% of the earth’s surface, the ocean is a vastly underexplored resource.\(^1\) In the hunt for bioactive molecules, marine microorganisms have potential for discovery of new molecules, considering the average cell densities of \(10^5\)–\(10^6\) ml\(^{-1}\) in seawater.\(^2\) Unfortunately, less than 1% of the marine bacteria are culturable by current methods, but if the range of culturable bacteria can be broadened, the scope of potential bioactive molecules from this resource would be immensely expanded. Marine bacteria are often symbionts of e.g. molluscs or sponges,\(^199,200\) and many bioactive molecules isolated from these sources were eventually found to be produced by an endosymbiotic microorganism. Microorganisms in the aquatic environment are highly exposed to threats and competition from the surroundings and the bioactive potential of their secondary metabolites may thus be greater than for terrestrial species.\(^201\)

Marine bacteria are essentially bacteria found in the marine environment. Differing only at the species level from the terrestrial species, the majority of bacterial orders and families also contain marine species. E.g. *Photobacterium halotolerans* (genus, species) belongs to the family of *Vibrionaceae* in the order Vibrionales. *Vibrionaceae* are widespread in the aquatic environment, but not limited to marine habitats.\(^202\) Marine bacteria must tolerate (and some require) high salinity, known as halotolerance, but bacteria generally adapt well to their environment, which may be the main reason for the relatively vague definition of marine bacteria.\(^203\)

3.5 Results and discussion

3.5.1 Small biologically active peptides from a marine *Photobacterium*

In the work described here, we have focused on isolation and structure elucidation of new bioactive metabolites from a marine *Photobacterium halotolerans*, collected from a mussel surface near the Solomon Islands during the Galathea 3 expedition.\(^198,204\) The bacterium was chosen due to a two-pronged inhibitory effect of bacterial growth and quorum sensing (QS).\(^205\) The known antibiotic holomycin was responsible for the antibacterial activity,\(^204\) and solonamide A and B were responsible for the QS inhibition in *S. aureus* (Paper I). Various other small peptide-like compounds were isolated from the same species, counting diketopiperazines and cyclotetrapeptides (unpublished results), as well as a series of larger depsipeptides belonging to an existing family called ngercheumicins. The four new ngercheumicins (F, G, H, and I) were tested in the same *S. aureus* QS assay as the
solonamides due to the structural resemblance with a 16-membered macrocyclic peptide ring and a hydrophobic alkyl chain. Here it was shown that the ngercheumicins, like the solonamides, inhibited QS in *S. aureus*, although higher concentrations were needed to obtain the inhibitory effect ([Paper II](#)) (Figure 3.29).

![Figure 3.29. Structures of the ngercheumicins (top) and the solonamides (bottom) with assigned stereochemistry and amino acid constituents shown.](image)

In the initial screening of a bacterial extract, the metabolite profile is usually investigated using HPLC-UV-MS. Crude fractionation on small scale enables early dereplication, and working with the *Photobacterium* we used $^1$H NMR for dereplication in addition to the MS-based approach. This was done on the EtOAc extract of a one liter liquid batch after fractionation on a diol column yielding ten fractions. One fraction revealed solonamide B as the major compound, which enabled comparison of the NMR data with known compounds found in the UV-MS dereplication, without the need for further purification. Solonamide B
exhibited no distinct UV spectrum and the exact mass was found to be consistent with the small peptide sansalvamide.\textsuperscript{206} The NMR dereplication could however firmly rule out the hits found in the NP database Antibase 2010.\textsuperscript{207} Especially for small peptides with no distinct UV chromophore, early NMR data can help establish the basic structural features in a fraction. Specific NMR resonances can often be used discriminatively to prevent spending valuable time on purification of known compounds or discarding an unknown because the molecular formula matches that of a known molecule. Horizontal gene transfer within the bacterial world is relatively common, and therefore new, exciting molecules like the solonamides could easily be discarded if the choice to identify a compound depends only on the exact mass and UV chromophore being unique.\textsuperscript{208}

The ngercheumicins were isolated from the cell-associated fraction of a 20 liter culture from a desire to characterize the chemical fingerprint of \textit{Photobacterium halotolerans}, which seemed to produce a wide variety of non-ribosomal peptides. Dereplication of this fraction revealed a series of analogs with relative mass-jumps of 28 and 2 Da, consistent with C\textsubscript{2}H\textsubscript{4} and a double bond equivalent, respectively. Due to the extensive resonance overlap in the fatty acid chain, there was some ambiguity in especially the assignment of the double bond position of ngercheumicin H (Table 3.4).

Table 3.4. Experimental \textsuperscript{1}H and \textsuperscript{13}C NMR data and assignments of the fatty acid chains in the ngercheumicins (complete NMR data in Paper II). Chemical shift values are in ppm.

<table>
<thead>
<tr>
<th>Atom #</th>
<th>Ngercheumicin F</th>
<th>Ngercheumicin G</th>
<th>Ngercheumicin H</th>
<th>Ngercheumicin I</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[\delta^1H]</td>
<td>[\delta^{13}C]</td>
<td>[\delta^1H]</td>
<td>[\delta^{13}C]</td>
</tr>
<tr>
<td>Fatty acid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36 - CO</td>
<td>-2.24 43.3</td>
<td>-2.24 43.4</td>
<td>-2.25 22.2</td>
<td>-2.25 22.2</td>
</tr>
<tr>
<td>37</td>
<td>3.78 67.4</td>
<td>3.77 67.5</td>
<td>3.78 67.5</td>
<td>3.78 67.5</td>
</tr>
<tr>
<td>38 - OH</td>
<td>4.62 -</td>
<td>4.60 -</td>
<td>4.60 -</td>
<td>4.59 -</td>
</tr>
<tr>
<td>39</td>
<td>1.34 36.2</td>
<td>1.31 36.6</td>
<td>1.32 36.6</td>
<td>1.32 36.6</td>
</tr>
<tr>
<td>40</td>
<td>1.42 25.1</td>
<td>1.34 24.8</td>
<td>1.36 24.8</td>
<td>1.35 24.8</td>
</tr>
<tr>
<td>41</td>
<td>1.96 26.6</td>
<td>1.23 ~29</td>
<td>~1.2 ~29</td>
<td>~1.2 ~29</td>
</tr>
<tr>
<td>42</td>
<td>5.31 129.6</td>
<td>~1.2 ~29</td>
<td>1.28 ~29</td>
<td>~1.2 ~29</td>
</tr>
<tr>
<td>43</td>
<td>5.31 129.6</td>
<td>~1.2 ~29</td>
<td>1.97 26.6</td>
<td>~1.2 ~29</td>
</tr>
<tr>
<td>44</td>
<td>1.96 26.5</td>
<td>~1.2 ~29</td>
<td>5.31 129.6</td>
<td>~1.2 ~29</td>
</tr>
<tr>
<td>45</td>
<td>1.28 29.0</td>
<td>~1.2 ~29</td>
<td>5.31 129.6</td>
<td>~1.2 ~29</td>
</tr>
<tr>
<td>46</td>
<td>1.24 28.3</td>
<td>1.21 29.1</td>
<td>1.97 26.6</td>
<td>~1.2 ~29</td>
</tr>
<tr>
<td>47</td>
<td>1.22 31.1</td>
<td>1.22 31.3</td>
<td>1.28 ~29</td>
<td>~1.2 ~29</td>
</tr>
<tr>
<td>48</td>
<td>1.25 22.1</td>
<td>1.25 22.1</td>
<td>1.25 28.2</td>
<td>~1.2 ~29</td>
</tr>
<tr>
<td>49</td>
<td>0.84 13.9</td>
<td>0.84 13.9</td>
<td>1.22 31.1</td>
<td>1.22 31.1</td>
</tr>
<tr>
<td>50</td>
<td>1.25 22.0</td>
<td>1.25 22.1</td>
<td>1.25 22.1</td>
<td>1.25 22.1</td>
</tr>
<tr>
<td>51</td>
<td>0.84 13.9</td>
<td>0.84 13.9</td>
<td>0.84 13.9</td>
<td>0.84 13.9</td>
</tr>
</tbody>
</table>

In the hope of supporting the structural assignments it was attempted to calculate the \textsuperscript{13}C chemical shifts of the fatty acid chains (Table 3.4) by density functional theory (DFT). Minimizations in MacroModel (Maestro, OPLS-2005, H\textsubscript{2}O, standard settings) were followed
by DFT calculations with Jaguar to predict NMR shielding constants in DMSO using the B3LYP functional and the 6-31G** basis set. The structures were simplified to N-methyl amides of the fatty acids to make the calculations faster. With this structural modification it was believed that the chemical shifts in the fatty acid chain would resemble those in the natural product. By calculating chemical shifts for the different structures with variations in cis and trans configurations and moving the double bond up to two carbons in each direction compared to the original assignment, we hoped there would be differences in the calculated chemical shift values that would enable the confirmation or correction of the assignment. The differences in chemical shifts were however very small, and the only unbiased way to compare the values was to sort them by size. Unfortunately this method was not sufficiently discriminative to provide a more reliable assignment (Figure 3.30 and Figure 3.31). Due to time restrictions and realizing that these small differences would be difficult to model in sufficient accuracy, no further studies were conducted.

It should however be noted that the assignments were supported by an n-7 fatty acid pattern in the double bond position within the three unsaturated ngercheumicins (n-7 is fatty acid nomenclature for an unsaturation seven carbon atoms from the chain end). The absolute stereochemistry of the ngercheumicins was not fully determined, in part due to the small amounts isolated, and in part due to the presence of three leucine residues of both L- and D-configuration, which hampered the analysis by Marfey’s method.
It was attempted to assess the stereochemistry of the amino acids in the ring by use of NOEs and backbone $J$ coupling constants, primarily between the amide protons and $H_a$. Comparing these experimental data to energy-minimized structures it was possible to distinguish between L-Thr and L-allo-Thr, as L-allo-Thr showed the best correlation with experimental data – and it was known from the Marfey’s analysis that the Thr residues had the same stereochemistry. There were small differences in fits between the structures but with four unknown stereocenters further conclusions could not be made. The NOESY data was only acquired at 800 ms, which is good for optimizing correlations but difficult to rely on for quantitation of NOEs to calculate distance information due to the susceptibility to spin diffusion. Unfortunately no further spectra could be acquired, as it had been judged that it would not be possible to complete the stereochemical assignment with the amounts available, and therefore the rest was used to screen for biological activity.

3.5.2 Quorum sensing and autoinducing peptides

Both the solonamides and the ngercheumicins were found to inhibit QS in *S. aureus*, including the community associated methicillin-resistant *S. aureus* (CA-MRSA) strain USA300 (Paper I and II). The inhibition occurs via the accessory gene regulator (*agr*) QS system which controls the onset of virulence in *S. aureus*. Dr. Anita Nielsen and Prof. Hanne Ingmer (Department of Veterinary Disease Biology, Faculty of Health and Medical Sciences, University of Copenhagen) showed that these cyclic peptides affect the expression of *agr* by down-regulation of *rnalIII* and *hla*, encoding extracellular toxins, and up-regulation of *spa*,
encoding surface proteins needed in the growth phase before the onset of QS and virulence. Without the presence of a QS inhibitor, the onset of virulence is controlled by autoinducing peptides (AIPs), which induce \textit{agr} expression at sufficiently high cell densities to initiate production of e.g. extracellular toxins. Different \textit{S. aureus} strains express different AIP variants, with QS cross inhibition between the types producing other AIPs. Structurally, AIPs are 16-membered peptide thiolactones, ring-closed between a cysteine residue and the peptide C-terminus (Figure 3.32). The solonamides and the ngercheumicins are all depsipeptides with the same lactone ring size and they all have an exocyclic chain, although the AIPs have peptide chains whereas the solonamides and ngercheumicins have (mainly) fatty acid chains. The amino acids in the C-terminus ends are generally hydrophobic for all the peptides and there are no positively charged amino acids in any of the peptides (Figure 3.32).

Figure 3.32. Amino acid sequences and ring closures of \textit{S. aureus} AIP I-III, solonamide B and ngercheumicin G. Amino acids with hydrophobic side chains are colored green, polar uncharged side chains are red, and polar charged sidechains are bright red. AIP structures were reproduced from 213.
Altogether these structural similarities contribute to explaining the observed bioactivities, and it is likely that the solonamides and the ngercheumicins are competitive inhibitors of the agr QS system in S. aureus. Of the new QS inhibitors identified, Solonamide B showed the highest activity, and Nielsen et al.\textsuperscript{214} recently showed that solonamide B binds to the agr signal receptor histidine kinase AgrC,\textsuperscript{210} confirming the hypothesis of solonamide B as a competitive inhibitor. Furthermore, it was shown that solonamide B causes a remarkable reduction in CA-MRSA USA300 toxicity towards human neutrophils (white blood cells in response to inflammation) and rabbit erythrocytes (red blood cells).\textsuperscript{214}

Solonamide B was produced in remarkably large amounts by the Photobacterium, as the most abundant secondary metabolite of this bacterium. From a 20 liter liquid culture, 201 mg of solonamide B was isolated, compared to 17 mg solonamide A and 0.5-1.1 mg of the ngercheumicins. This suggests that solonamide B has an important function for this marine bacterium or that there may be another biological role of this depsipeptide for the Photobacterium itself, perhaps an intraspecies communication system as seen within the AIPs of the staphylococci; however without further studies this is purely speculations. It is ventured that these depsipeptides are not made for QS inhibition in S. aureus, and therefore some sort of communication or regulation in a different species (or the Photobacterium) seems a more likely evolutionary reason for their production.

For further information see Paper I and II on the solonamides and the ngercheumicins, respectively.
3.5.3 **Fungal metabolites from *A. fijiensis***

Filamentous fungi are historically some of the most prolific producers of various drugs, such as antibiotics and anticancer agents. In the search for novel secondary metabolites we investigated two species of the black aspergilli, namely *A. carbonarius* and *A. fijiensis*, both found in the fungal collection at DTU Systems Biology. *A. carbonarius* is a known producer of ochratoxin A, a foodborne carcinogenic mycotoxin known to be up-concentrated in the food chain.\(^{215,216}\) The species *A. fijiensis* was only recently described, and had therefore not been extensively investigated for novel metabolites before.\(^{217}\) Dereplication of the plug extracts (see Experimental section 3.7) resulted in the decision to make a large-scale extraction of *A. fijiensis* since this seemed to produce more possibly unknown molecules. Dereplication primarily based on chromatographic retention time, masses and explorative solid-phase extraction (E-SPE)\(^{173}\) revealed the known hits in Antibase for *A. fijiensis* to be asperparaline A, preparaherquamide, marcfortine A and C, oxaline, calbistrins and secalonic acids (Figure 3.33). These were all previously isolated from either *Aspergillus* or *Penicillium* spp.\(^{207}\) which makes them likely to be produced by this fungus. Both the asperparalines, the secalonic acids, and the marcfortines have several analogues which could be present in the extract, but only a representative selection is shown.

After dereplication the three major compounds believed to be unknown were targeted for isolation and identification. It turned out that one of these ([M+H]\(^+\) m/z 417.2, 9.0 min) had recently been identified by a colleague from *A. aculeatus* (Lene M. Petersen, DTU Systems Biology, paper submitted for publication) and another was supposedly up-concentrated plastic remnants from the stomacher bags ([M+H]\(^+\) m/z 338.3, 11.8 min). This turned out to be an increasing problem, and identifying the source and reason for these changes is something that often takes long time to unravel. The work with the extract however continued, as there was still one unknown secondary metabolite ([M+H]\(^+\) m/z 266.1, 4.5 min). Even though a hit in Antibase suggested the molecule Nigerloxin as a possible candidate, the UV data were different. The fractionation and purification resulted in identification of four secondary metabolites from *A. fijiensis* (see Experimental section 3.7).
Figure 3.33. Top: Structures from dereplication of the *A. fijiensis* extract. Bottom: Base peak chromatogram (m/z 50-1000 ESI+) of the *A. fijiensis* raw extract in methanol. With annotations of [M+H]+ exact mass (observed or calculated from other adducts) and possible known compounds. See section 3.7.1 for experimental procedures.

The unknown molecule was identified as a new pyranonigrin and tentatively named pyranonigrin E (Figure 3.34, see NMR assignment in the Appendix). This family constituted pyranonigrin A, B, C, D, and S, all known from *A. niger.*218,219 Pyranonigrin A, the closest
analogue, was known as a radical scavenger found in molds on rice and for a moderate insecticidal activity.\textsuperscript{218,220} The new pyranonigrin E was $O$-methylated, which had not been seen before, and contained a longer alkenyl chain than the existing pyranonigrins. The specific rotation of pyranonigrin E was measured and found comparable to that of pyranonigrin A,\textsuperscript{219} and therefore the absolute configuration of the hydroxyl group is likely to be $R$ for pyranonigrin E as well (see details in section 3.7.2).

![Pyranonigrin A and Pyranonigrin E](image)

Figure 3.34. Structures of pyranonigrin A (closest resembling) and the new pyranonigrin E.

The extract also contained the two known diketopiperazines JBIR-74 and JBIR-75 (Figure 3.35) which were purified from the same fraction as pyranonigrin E,\textsuperscript{221} however only 0.2 mg of JBIR-75 was isolated, and therefore it was not subjected to NMR analysis. The NMR assignment for JBIR-74 is in the Appendix. The MS data for the purified compound believed to be JBIR-75 were in agreement with a loss of a CH$_2$ moiety (14 Da) and the UV chromatogram was equivalent to that of JBIR-74 with UV$_{\text{max}}$ at 300 and 226 nm for both. JBIR-75 and pyranonigrin E were subjected to antifungal and anticancer assays (Peter B. Knudsen, DTU Systems Biology), however no activity was found.

![JBIR-75 and JBIR-74](image)

Figure 3.35. Structures of JBIR-75 and JBIR-74. The known compounds contain D-Val and D-allo-Ile, and the optical rotation of JBIR-74 shows positive values similar to those found in the literature.\textsuperscript{222} Therefore the metabolites most likely have the same configurations.

Pyranonigrin E was present in another fraction that also contained a metabolite from the extract eluting at 4.2 min (Figure 3.33) with a base peak at m/z 263.1644. This corresponded to water loss from a molecule with [M+H]$^+$ of 281.1748 as seen from the presence of [M+H-2H$_2$O]$^+$, [M+Na]$^+$, [2M+Na]$^+$. This low molecular mass gave 30 Antibase hits with the same molecular formula and a nondiscriminative UV spectrum ($\lambda_{\text{max}}$ 230-240 nm) made it difficult to dereplicate. Since it was possible to purify more of the new pyranonigrin as well as
identify this unknown molecule, it was purified by RP-HPLC (see Experimental section 3.7.2) and identified by 1D and 2D NMR spectroscopy as decumbenone A (NMR assignment in the Appendix), a melanin biosynthesis inhibitor previously isolated from a *Penicillium decumbens*.222

In collaboration with bachelor student Zeliha Basaran and Assoc. Prof. Thomas O. Larsen the antifungal activity in the *A. fijiensis* extract against *Candida albicans* was investigated. Initially it was found that plug extracts showed high antifungal activity, and testing of a series of fractions led to the identification of the calbistrins and secalonic acids as responsible for the activity. This was inferred from testing E-SPE fractions,173 where all the active fractions contained either calbistrins or secalonic acids. The calbistrins had known antifungal activity against *C. albicans*,223 and the secalonic acids are mycotoxins found in *Penicillium* and *Aspergillus* species.224,225 Testing secalonic acid F (purified from *A. homomorphus* by Casper Hoeck, DTU Chemistry) confirmed the antifungal activity against *C. albicans*.

In summary, the chemistry of *A. fijiensis* was investigated for novel metabolites and one new molecule from *A. fijiensis* was identified and named pyranonigrin E. Two known secondary metabolites from *Aspergillus* spp., JBIR-74 and -75 were identified along with decumbenone A, known from *P. decumbens*.

### 3.6 Conclusions and perspectives

Despite the high technological state of our society, disease continues to be a major challenge in the 21st century. Cancer is a major cause of death in the developed world, and also bacterial infections are today not only a threat to the frail individuals of our society, but community-associated infections of healthy adults by multidrug resistant pathogens are becoming more and more common. While antibiotic resistance is increasing, the number of new drugs to combat these pests decreases. Many actions can be taken to try to stop the development, and probably the optimal strategy is multifaceted. While the use of antibiotics against non-threatening infections both in humans and animals must be restricted, it seems naive to not keep developing new molecular “weapons” against multidrug resistant pathogens. The choice of methods for drug development may be a matter of belief or opinion, however in retrospect nature has been the major supply of drug leads, and neglecting to utilize this superior resource while we try to establish other methods for drug discovery could turn out to be fatal. It is not hard to imagine that nature already has the
solutions to our problems; we “just” have to discover them, and the microbial world seems like the perfect place to search.

Filamentous fungi are being and have been extensively investigated by natural product chemists, and the same secondary metabolites are often found in phylogenetically distant species due to horizontal gene transfer. This highlights the importance of early-stage dereplication and careful strain selection, using continuously updated metabolite databases in the process. Metabolites from *A. fijiensis* were isolated in relatively small amounts and initial attempts of linking a biological activity to the structures were unsuccessful. Our findings from a marine bacterium, on the other hand, indicate that by looking in untraditional niches of the world, there is interesting chemical biology to be found. Despite the isolated metabolites being biosynthetically simple compared to the complex polyketides found in other sources, the biological activity and high yields (at least for the solonamides) proved the search worthwhile. Solonamide B was shown to strongly inhibit the *agr* QS system of *S. aureus*, including the hypervirulent strain USA300. This type of anti-infectious therapy is a hot topic, and while the *in vitro* activity is in the micromolar range and this is not a molecule ready for clinical trials, solonamide B shows very promising results by reducing the killing of human neutrophils in the presence of *S. aureus* USA300.

By optimizing the selection of strains, the dereplication process, and the biological screening of secondary metabolites, the hit rate increases. Whether future natural products based drug discovery will be in a purely academic setting is probably too early to tell. There is however a controversy between the requirement to publish as an academic researcher and the need to monopolize drugs in pharmaceutical companies. In the current state, no private company can be expected to develop a drug spun out of a research publication. Even a highly efficacious bioactive molecule, proven able to solve one of the world’s major problems, is unlikely to become a drug if a pharmaceutical company cannot get return of investments. It seems however that drug discovery and development of new antibiotics must receive imminent attention to stop the current trend within multi-drug resistance from spiraling out of control.
3.7 Experimental

3.7.1 Fungal extracts

*A. carbonarius* (IBT 21089) and *A. fijiensis* (IBT 4580, 13989, 14441, and 14442) from the IBT culture collection at the Department of Systems Biology were inoculated on agar plates with CYA (Czapek Yeast Autolysate), OAT (Oatmeal agar), YES (Yeast Extract Sucrose), and MEA (Malt Extract Agar) and grown in the dark at 25 °C for 7 days. Nine plugs across the colonies from the CYA and YES plates were extracted with 1 ml ethyl acetate containing 1 % (v/v) formic acid and ultrasonicated for one hour. In addition 20 plugs from a CYA plate were deposited at 0 °C for the preparation of a large extract later. The plug extracts from CYA and YES were transferred to HPLC vials and evaporated under N$_2$ at 33 °C until dryness, then redissolved in acetonitrile (1 ml) and ultrasonicated for 10 min following filtration (0.45 μm PTFE filter) to remove solid particles. 50 μl was transferred to another HPLC vial and diluted with 100 μl acetonitrile for LC-DAD-HRMS and dereplication.

For a large-scale extract of *A. fijiensis* (IBT 14442) 200 plates of CYA were inoculated and incubated in the dark at 25 °C for 7 days. The agar plates were extracted after 1 min in a Colworth 400 stomacher with ethyl acetate containing 1 % (v/v) formic acid, and then left to extract for 1 hour following a second round of biomass extraction overnight. Filtration on a Büchner funnel and concentrated to dryness on a rotary evaporator yielded about 4.4 g of pooled extracts (sample for HRMS) ready for desugaring and degreasing. The extract was dissolved in methanol/MilliQ water (9:1, 250 ml) and extracted with heptane (250 ml) and then 200 ml of MilliQ water was added to the aqueous phase and extracted with 200 ml dichloromethane. From each of the three fractions a sample was taken for HRMS (filtrated). After freeze drying the heptane extract yielded 0.2 g fats, the methanol/water phase yielded 2.4 g of polar substances, and the dichloromethane phase yielded 1.0 g containing secondary metabolites.

3.7.2 Fractionation and purification

Both the methanol-water and the dichloromethane extracts of *A. fijiensis* were subjected to E-SPE following the procedure by Månsson et al. using SAX, MAX, SCX and diol SPE cartridges. The E-SPE fractions were subjected to LC-MS and bioguided dereplication via the *C. albicans* assay described below.
The dichloromethane extract was fractionated on an Isolera One automated flash chromatography system (Biotage, Uppsala, Sweden) using a 25 g SNAP Isolute diol column (Biotage, Uppsala, Sweden) with solvents ranging from heptane to dichloromethane to ethyl acetate to methanol yielding 18 fractions. Fraction #12 (148 mg) was further purified by reversed phase flash chromatography using a 10 g SNAP column with ZepraZT C\textsubscript{18} material at a flow of acetonitrile-MilliQ water at 10 ml min\textsuperscript{-1} from 15 to 75 % acetonitrile over 22 min and from 75 to 100 % over 4 min yielding 11 fractions. Fraction #3 (19 mg) from this run was subjected to final purification on a Waters 600 chromatograph (Milford, MA, USA) with a Waters 600 DAD with a Luna II C\textsubscript{18} column (5 μm, 10 x 250 mm, Phenomenex, Torrance, CA, USA) running at 5 ml min\textsuperscript{-1} acetonitrile-MilliQ water from 20 to 60 % acetonitrile over 20 min, yielding 11.0 mg of JBIR-74 (6.5 min, 33 % acetonitrile, white solid, HRMS m/z 248.1172, M\textsuperscript{+} calculated for C\textsubscript{12}H\textsubscript{16}N\textsubscript{4}O\textsubscript{2} m/z 248.1273, [α]\textsubscript{D} in methanol at 20°C, 11 mg ml\textsuperscript{-1}: +122.2°) and 2.7 mg of pyranonigrin E (12 min, 44 % acetonitrile, white solid, HRMS m/z 265.0953, M\textsuperscript{+} calculated for C\textsubscript{13}H\textsubscript{15}NO\textsubscript{5} m/z 265.0950, [α]\textsubscript{D} in methanol at 20°C, 2.7 mg ml\textsuperscript{-1}: +29.6°).

Fraction #11 from the diol fractionation (62 mg) was purified on the same Waters chromatograph with a Gemini C\textsubscript{6}-Phenyl column (5 μm, 10 x 250 mm, Phenomenex, Torrance, CA, USA) from 15 to 45 % acetonitrile in 17 min and from 45 to 100 % in 5 min yielding 0.8 mg of decumbenone A (13 min, yellow oil, HRMS m/z 280.1712, M\textsuperscript{+} calculated for C\textsubscript{16}H\textsubscript{24}O\textsubscript{4} m/z 280.1675) and an additional 0.9 mg of pyranonigrin E (15 min).

### 3.7.3 Bioactivity testing

Major diol fractions (#6-13) were subjected to antifungal activity testing using \textit{C. albicans} (IBT #654 and IBT #656) (with Zeliha Basaran and Assoc. Prof. Thomas O. Larsen) using 20 μl at 20 mg ml\textsuperscript{-1} in an agar well diffusion assay with inoculations of 24 and 48 hours, using DMSO as negative control and the antifungal drug griseofulvin as positive control. IBT 654 however showed very weak susceptibility towards griseofulvin, whereas IBT #656 displayed clearing zones of up to 11 mm when subjected to griseofulvin. \textit{A. fijiensis} fractions showed clearing zones up to 31 mm in diameter, whereas pure secalonic acid D (MFP147 from local metabolite collection, not pure) and F (supplied by Casper Hoeck) gave clearing zones of 22 and 11 mm, respectively.

### 3.7.4 LC-MS and NMR analyses

LC-HRMS analyses were performed on a MaXis quadrupole time of flight (qTOF) electrospray ionization (ESI) mass spectrometer (Bruker Daltonics, Bremen, Germany) with
an Ultimate 3000 UHPLC system and a DAD detector (Dionex, Sunnyvale, CA, USA) and separation by a Kinetex C\textsubscript{18} (2.6 μm, 2.1 x 200 mm, Phenomenex, Torrance, CA, USA) at 40 °C with a 0.4 ml min\textsuperscript{-1} flow of acetonitrile-water (MilliQ) containing 20 μM formic acid from 10 to 100 % acetonitrile in 10 min.

NMR analyses were performed on a Bruker Ascend 400 MHz with a 5 mm BBO Prodigy cryoprobe, a Varian 500 MHz Unity Inova with a 5 mm HCP probe, and a Bruker Avance 800 MHz spectrometer with a TCI cryoprobe.
4 Human milk oligosaccharides

This chapter describes structural studies of oligosaccharides, and begins with a background introduction, followed by a description of the experimental work on NMR structural studies of biologically important oligosaccharides in relation to this thesis. Relevant NMR assignments are in the Appendix for chapter 4, and three publications related to the work described in this chapter can be found in the back of the thesis, Paper III, IV, and V (in preparation).

4.1 Introduction

Breast milk is the perfect food source for an infant in the first period of life. It contains all the things needed for the baby to grow healthy and strong. If for some reason breast milk is not an option, the usual alternative is infant formula, a cow’s milk-based product made to mimic human milk. However data suggests that there are differences in the health and well-being of breastfed vs. bottle-fed infants. Comparing human milk to cow’s milk there is one particular group of molecules found in human milk which commercially produced cow’s milk is virtually devoid of: Human milk oligosaccharides (HMOs).

HMOs are a large group of lactose-based oligosaccharides with backbones of primarily D-glucose (Glc), D-galactose (Gal) and N-acetylglucosamine (GlcNAc), often decorated with L-fucose (Fuc) and Neu5Ac (N-acetylneuraminic acid also known as sialic acid) (Figure 4.36). The simplest structures are lactose decorated with Neu5Ac, usually in the 3’- and 6’-positions on Gal, or with Fuc, usually in the 3- and 2’-positions on Glc and Gal, respectively (Figure 4.37).
4.2 The HMO project

In 2010 a range of industry and university partners began a collaboration aimed at developing enzymatic processes to produce HMOs. The partners (Arla Foods, DuPont, DTU, University of Southern Denmark (SDU), University of Copenhagen (KU-Life), and University of Reading) realized that large-scale production of HMOs resulting from a four year research project was an ambitious aim; however the immediate goal was production of HMOs in laboratory scale and assessing the structural and biological properties. The project started in April 2010 led by Prof. Jørn D. Mikkelsen (DTU Chemical Engineering) and the 35-40 researchers were divided into five work packages (WP) with responsibilities of 1) generating enzymes, 2) bioresource utilization and process development, 3) chemical structure and analysis, 4) functional properties and proof of concept, and 5) project management and intellectual property rights (IPR). One of the end goals was production of 10 HMOs to be analyzed and screened in various assays for biological activity. NMR and MALDI-TOF analyses (by groups at DTU Chemistry and SDU, respectively) should be used to characterize the final ten HMOs and collaborate with WP1 and WP2 in the ongoing analysis of reaction products.

Approximately one year into the project I entered as a PhD student, and the initial work was to prepare for future analyses of potentially small amounts and/or mixtures, by structure elucidating commercially available HMO samples, which were also used by the different WPs until scale-up the various HMOs could be achieved. These analyses were often combined with purity assessment as the biological assays in WP4 are sensitive to contaminants and assessing the purity of a reactant is important to be able to characterize the
process. Carbosynth (Berkshire, UK) and Elicityl (Crolles, France) are major commercial distributors of HMOs for research purposes with mainly chemical synthesis and biomass extraction/fermentation as respective production methods. Different impurities can thus be expected from the products, depending on the production method and purification, which in turn may influence the response of a biological assay.

4.3 Biological role of HMOs

The average lactose content in human milk and cow’s milk is 55-70 g/l and 40-50 g/l respectively, and human milk is estimated to contain 5-15 g/l of HMOs whereas cow’s milk contains only trace amounts of these oligosaccharides, although the composition varies during the different phases of lactation. More than 180 different HMOs have been identified in human milk samples and the biological role of these oligosaccharides is surprisingly manifold, although some of the effects are not fully elucidated:

- HMOs are prebiotic, which defines a non-viable food component not digested by the host organism, but which benefits the gut microbiota, e.g. bifidobacteria and lactobacilli.

- HMOs function as soluble receptor analogues for a wide range of pathogens, preventing their binding to the intestinal epithelium and thereby inhibit infection. These pathogens include influenza virus, HIV-1, V. cholera, E. coli, and many more.

- It is believed that HMOs have beneficial immunomodulatory properties by activating the immune system upon exposure and perhaps even prevent allergies.

- As fucose and sialic acid are abundant constituents of brain gangliosides, HMOs appear to be important factors in infant brain development.

In fact, the brain gangliosides are structurally very similar to gangliosides in the intestinal epithelium where e.g. cholera toxin binds. Gangliosides are ceramide-bound HMO-like molecules found on cell membranes eliciting interaction and communication with the environment. The only other mammal found to have a milk oligosaccharide composition as complex as humans is the elephant. It may not be a coincidence that both humans and elephants are slow-growing species with highly developed neural systems. Other mammals are found to have different building blocks, other ratios of known oligosaccharides or a simpler oligosaccharide composition, but the origins and conclusions to these relationships remain unsettled.
4.4 Sources of fucose and sialic acid

Fuc- and Neu5Ac-containing molecules are very valuable constituents in research, in medicine, and as food supplements due to their biological properties. Unfortunately it is not straightforward to obtain these monomers in pure form, neither from natural sources nor by chemical synthesis. L-fucose is equivalent to 6-deoxy-L-galactose, however L-galactose is neither available as a suitable substrate for large-scale synthesis. Several chemical syntheses converting other monosaccharides into L-fucose are described in the literature, but common disadvantages are low yields and many synthetic steps.\(^{241,250-252}\) Wong \textit{et al.} produced L-Fucose enzymatically in 55\% yield from dihydroxyacetone phosphate and lactaldehyde.\(^{253}\) Pinto \textit{et al.} synthesized a Fuc analogue in 40\% yield, however further deprotection steps would be required to make any use of it.\(^{254}\) A patent application from 2013 by Glycom A/S (Kgs. Lyngby, Denmark) also describes new ways of synthesizing L-Fuc.\(^{255}\)

There are some examples of attempts to synthesize Neu5Ac in the literature,\(^{256-258}\) however many syntheses focus on different analogues\(^{259}\) or use complicated or expensive starting materials.\(^{260,261}\) In general the difficulties in extracting or synthesizing L-fucose and sialic acid are reflected in the cost of the monomers. Carbosynth currently sells 100 g L-Fucose for 608 €, and Neu5Ac for 266 €.\(^{262}\)

Natural sources abundant in L-fucose or Neu5Ac are not easy to find. Plant polysaccharides contain L-fucose and especially seaweeds have a high content of sulfated L-fucose polymers called fucoidan.\(^{241}\) Fucoidan is primarily made up of L-fucose and sulphate groups. It is used as a dietary supplement and has received much attention due to its manifold biological activities.\(^{263,264}\) The best source for enzymatic utilization of Neu5Ac seems to be caseinoglycomacropedite (CGMP). CGMP is a constituent in whey, which is a byproduct in cheese production. It is not to be confused with the secondary messenger cyclic guanosine monophosphate (cGMP) or CMP-Neu5Ac which is the sialic acid donor for sialytransferases in mammals (Figure 4.38). CGMP is a 64 amino acid peptide containing 5-13 \% (w/w) Neu5Ac and dairy companies such as Arla Foods produce CGMP isolates for functional food ingredients.\(^{265,266}\) Due to the bulk production, low cost and relatively high sialic acid content it seems to be a good source of Neu5Ac, which is difficult to obtain from other natural sources. Sialic acid can be cleaved from CGMP by acid hydrolysis or by the use of enzymes (Figure 4.38).
Neuraminidases or sialidases are sialyl hydrolases that cleave the glycosidic linkages of Neu5Ac-glycan structures. Two different types of enzymes can catalyze the transfer of Neu5Ac between residues: Sialyltransferases, which use nucleotide-sugar donors in an SN2-like inversion mechanism to form α-Neu5Ac-glycan bonds, and trans-sialidases, which move Neu5Ac residues between glycan chains (or to water, functioning as sialidases) via a double displacement or retaining mechanism.267

### 4.4.1 Perspectives on the production of HMOs

Enzymes with the ability to transfer Neu5Ac are also used to synthesize and manipulate sialic acid-containing compounds in chemoenzymatic reactions.268 Another option to make sialylated oligosaccharides is the use of cell factories.269 The method requires some genetic engineering to optimize the biosynthesis, but 3'-sialyllactose yields as high as 25 g/l has been obtained from cultures using lactose as acceptor and glycerol as the only carbon and energy source.270 The strain of *E. coli* used in the production of 3'-sialyllactose is also used in the pharmaceutical industry for e.g. large-scale production of human insulin. Restrictions on the production of food supplements (especially for infant formula) are however expected to be harder, and the recombinant genes come from pathogenic bacteria, which produce Neu5Ac as a way to evade the host immune system.271 It is also very doubtful whether parents would allow feeding large amounts of oligosaccharides produced by recombinant *E.
*coli* to their children, especially after the Chinese scandal in 2008 where melamine added to milk caused serious kidney disease in thousands of children.\textsuperscript{272}

A group of Chinese researchers have cloned a herd of cattle to produce human lysozyme, which protects infants from infections and supposedly this “humanized the bovine milk”.\textsuperscript{273} The milk oligosaccharide content is however unchanged, but one could speculate on the possibility of cloning genes necessary for HMO production in a similar future project, and thus further enhance the health benefits of cow’s milk.

Chemical synthesis of HMOs is a complicated process, and reaching high molecular weights in solution is a challenge with the many protection and deprotection steps required, decreasing the synthetic yield. Another issue is the chemistry of Neu5Ac, with the lack of a 3-hydroxy group and the carboxyl group at the anomeric position making glycoside formation difficult.\textsuperscript{274,275} The HMO backbone tetrasaccharide lacto-\(N\)-tetraose (LNT) was synthesized by Noro *et al.*\textsuperscript{276} in 1979 and since then a wide range of HMOs have been synthesized.\textsuperscript{277–281} Recently, Jennum *et al.* synthesized LNT, lacto-\(N\)-neotetraose (LNnT – see Table 4.5 on page 69) and developed one-pot procedures for these rather complex oligosaccharides. Fucosylated LNT and LNnT were also synthesized, yielding LNFP-I and LNnFP-I.\textsuperscript{282}

A combination of chemical and enzymatic reactions seems to be the best route to sialic acid-containing HMO-like molecules, either by chemical modification of enzymatically synthesized Neu5Ac-containing molecules or by final decoration of synthetic HMO backbone structures with Neu5Ac.\textsuperscript{280,283}

Due to the current inaccessibility of HMOs in large scale, other oligosaccharides are used as food additives and in some infant formulas today.\textsuperscript{234,284} Galactooligosaccharides (GOS) are oligomers of D-galactose with a D-glucose moiety in the reducing end and fructooligosaccharides (FOS) are oligomers of D-fructose with D-glucose in the reducing end.\textsuperscript{285} These oligosaccharides are not found in human milk but they have prebiotic properties mimicking some of the highly desired effects of HMOs.\textsuperscript{234,284}

### 4.5 Characterization of HMOs

Analysis and detection of oligosaccharide mixtures is difficult due to the structural similarities and the high number of possible isomers. HPLC, MS, and NMR spectroscopy are probably the best methods to characterize HMOs, preferably used together either directly coupled or one by one.\textsuperscript{286} HMO molecules are very hydrophilic and therefore the most suitable separation method is anion exchange chromatography.\textsuperscript{287} Capillary electrophoresis
has also been used for separation of either sialylated or neutral HMOs, but as the separation is based on electrophoretic mobility, which is different for charged and uncharged molecules, this method is not applicable to HMOs in general.\textsuperscript{288,289}

Mass spectrometric methods (especially if initial separation of mixtures is achieved) are very important tools to examine HMOs due to the exact mass determination. Matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) techniques are especially suitable for carbohydrates which can be difficult to ionize.\textsuperscript{290,291} The MALDI technique cannot be coupled to chromatography and is therefore most suitable for small amounts and relatively uncomplicated samples, whereas ESI is coupled to a liquid stream and therefore finds its strength in the analysis of mixtures by coupling to HPLC. Tandem MS can be used to obtain monomer sequences in simple oligosaccharide samples, and in combination with specific enzyme digestion it is possible to elucidate branching points.\textsuperscript{235,286} Although a bit cumbersome, this method provides a very sensitive method for analysis of HMOs in minute amounts, down to the low femtomole range.

4.5.1 HMO analysis by NMR

Compared to MS, NMR spectroscopy is a relatively insensitive technique; however NMR yields important structural information. The primary strength of NMR spectroscopy in oligosaccharide analysis is within structural studies in the distinction of isomers and conformations of HMOs in solution. Even in mixtures of compounds it is possible to elucidate monomer structures and branching points. So-called structural reporter groups have been used to identify specific monomers due to NMR resonances outside the crowded region of the \textsuperscript{1}H NMR spectrum (3-4 ppm);\textsuperscript{292,293} e.g. the 6-Me group of Fuc or the unique resonances from the equatorial and axial H3s of Neu5Ac. Anomeric resonances in both \textsuperscript{1}H and \textsuperscript{13}C NMR contain information that can be used in the structure elucidation. The anomeric chemical shifts are different for \(\alpha\) and \(\beta\)-anomers and are therefore often used discriminatively, e.g. the differences in \textsuperscript{1}H and \textsuperscript{13}C chemical shifts between \(\alpha\)- and \(\beta\)-Glc in \(3'\)-sialyllactose is about 0.5 ppm and 4 ppm, respectively. Also the anomeric \(J_{\text{CH}}\) coupling constants are known to differ about 10 Hz between \(\alpha\)- and \(\beta\)-anomers.\textsuperscript{294} For Neu5Ac, which has a quaternary anomeric carbon, assignment of the anomeric configuration is difficult. The \(J_{C2,H3Ax}\) and \(J_{C1,H3Ax}\) coupling constants are believed to offer the best discrimination, however the measurement is not straightforward, especially in cases of overlapping resonances.\textsuperscript{295-297} Other coupling constants (especially \(J_{HH}\)) are used in the assignment of axial and equatorial substituents to differentiate especially the hexopyranoses. The \textsuperscript{13}C
chemical shifts are also fairly monomer-specific, and chain elongation or branching causes systematic changes in the chemical shifts of surrounding atoms, depending on the type and stereochemistry of the glycosylation. This is something that automatic assignment procedures and carbohydrate databases take advantage of.\textsuperscript{298,299} Chemical shifts of reference compounds are thus used to great advantage in assignment of carbohydrates, and therefore it is important to have the same spectral reference. Most unprotected carbohydrate samples are acquired in D\textsubscript{2}O, as it simplifies the spectra by removing any resonances from hydroxyl groups. However, D\textsubscript{2}O is not an ideal solvent for spectral referencing. There is no $^{13}$C resonance and the HDO residual solvent resonance around 4.79 ppm at room temperature is highly temperature dependent.\textsuperscript{300} In addition, the residual solvent resonance is often removed by presaturation of the resonance frequency if the sample contains larger amounts of water. This makes it even harder to use as a spectral reference in $^1$H NMR. There are two ways to solve this problem: Either adding a reference compound to all samples and reference in accordance with literature values or using external referencing. External referencing is done by acquiring a spectrum of a reference compound in the given solvent, and then using the spectrum reference frequency obtained to calibrate spectra run in the same solvent at the same temperature, without addition of reference compound and thus avoiding sample contamination.\textsuperscript{301} This method however does not take salt and pH effects into account, which may change the chemical shift values. A reference compound often used for carbohydrates is 1,4-dioxane adjusted to 3.75 ppm and 67.4 ppm at 25 °C for $^1$H and $^{13}$C, respectively.\textsuperscript{300,302}

\subsection*{4.5.2 NMR structural assignment of HMO molecules}

The experimental work within this project began with the construction of a small NMR library of assigned HMOs. Structure elucidation of commercial samples involved the tri-, tetra- and pentasaccharides shown below (Table 4.5, see NMR assignments with HMBC correlations in the Appendix). Reference $^1$H and $^{13}$C NMR assignments are available for many of the HMOs, however in many cases in incomplete form or based on structural reporter groups.\textsuperscript{245,246,303–306} Various reference spectra of lactose, selected monomers and possible contaminants were acquired and analyzed when necessary.
Table 4.5. Commercial HMOs characterized by NMR spectroscopy. NMR assignments are in the Appendix.

<table>
<thead>
<tr>
<th>Trivial name (abbreviation)</th>
<th>Condensed structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>3'-sialyllactose (3'-SL)</td>
<td>Neu5Acα2-3Galβ1-4Glc</td>
</tr>
<tr>
<td>6'-sialyllactose (6'-SL)</td>
<td>Neu5Acα2-6Galβ1-4Glc</td>
</tr>
<tr>
<td>2'-fucosyllactose (2'-FL)</td>
<td>Fucα1-2Galβ1-4Glc</td>
</tr>
<tr>
<td>3-fucosyllactose (3-FL)</td>
<td>Galβ1-4(Fucα1-3)Glc</td>
</tr>
<tr>
<td>Lacto-N-tetraose (LNT)</td>
<td>Galβ1-3GlcNAcβ1-3Galβ1-4Glc</td>
</tr>
<tr>
<td>Lacto-N-neotetraose (LNnT)</td>
<td>Galβ1-4GlcNAcβ1-3Galβ1-4Glc</td>
</tr>
<tr>
<td>Lacto-N-fucopentaose V (LNFP-V)</td>
<td>Galβ1-3GlcNAcβ1-3Galβ1-4(Fucα1-3)Glc</td>
</tr>
<tr>
<td>Lacto-N-neofucopentaose V (LNnFP-V)</td>
<td>Galβ1-4GlcNAcβ1-3Galβ1-4(Fucα1-3)Glc</td>
</tr>
</tbody>
</table>

There are a few differences between the structure elucidation of oligosaccharides compared to natural products, and a general assignment protocol is therefore described here. Since spectra are usually acquired in D$_2$O and small chemical shift differences are often important for the outcome of analysis, careful spectral referencing must be ensured. In this work 1,4-dioxane was used as external reference ($\delta$H 3.75 ppm, $\delta$C 67.4 ppm, 25°C). The majority of carbohydrate resonances in $^1$H NMR are observed in a narrow spectral region between 3.5 and 4 ppm with severe resonance overlap. The less crowded regions of the $^1$H spectrum contains the structural reporter groups$^{293}$ which are more easily used in analysis and includes the anomeric protons between 4 and 6 ppm and any other more or less monomer-specific resonances from e.g. N-acetyl groups and deoxy-sugars (Figure 4.39).

![Figure 4.39](image-url)
Even though compounds like the fucosyl- and sialyllactoses contain the same respective monomers and thus have identical molecular masses, there are distinguishable features in the less crowded areas of the NMR spectra which aid in the differentiation and detection of these isomers. For $^{13}$C NMR, the majority of the carbohydrate resonances of pyranoses are found between 60 and 80 ppm, and the anomeric resonances are around 90-105 ppm. Consequently, protons in the crowded region of the $^1$H spectrum of about 0.5 ppm have their respective (directly bonded) carbon chemical shifts distributed over a 20 ppm range. This helps to unravel resonance overlap and distinguish $^{13}$CH-pairs by using 2D heteronuclear experiments. For an oligosaccharide, little information can be extracted from the 3.5-4 ppm region of a DQF-COSY spectrum, but with sufficient resolution in the second dimension these protons are more easily identified by use of gHSQC, gHMBC, gHSQC-TOCSY and other heteronuclear experiments (Figure 4.40). Especially the gHSQC-TOCSY is highly useful for oligosaccharide samples. Most of the $^{13}$C resonances in each monosaccharide can be assigned from the HSQC-TOCSY correlations to the anomeric proton. HSQC-type correlations are used to identify protons directly correlated to the $^{13}$C resonances in the spin system identified from the gHSQC-TOCSY. Missing resonances within the monosaccharide building blocks are identified by concurrent use of gHMBC, gH2BC and the homonuclear DQF-COSY and TOCSY, depending on the extent of resonance overlap. Monosaccharide building blocks are connected using HMBC- and NOESY/ROESY-correlations across the glycosidic bonds. In cases where this type of HMBC-correlation is absent, one must rely on through-space couplings from NOESY and ROESY spectra. The chemical shifts should however also present clues to the linkage position by comparison to similar molecules. To illustrate some of the points regarding spectral overlap for a pure HMO, selected 1D and 2D NMR spectra for synthetic LNT are shown in Figure 4.40. The spectral overlap seen in the $^1$H and DQF-COSY spectra is markedly reduced in the heteronuclear experiments.

4.5.3 Purity assessment
Purity assessment by NMR is not straightforward. Inorganic impurities are rarely detected and NMR resonances of macromolecules such as polymers or proteins often disappear due to fast ($T_2$) relaxation compared to small molecules. This leads to extensive broadening of the resonances which in some cases are only seen as an elevated baseline. Chemical shifts easily change with differences in temperature, ionic strength and pH, and the optimal parameters (e.g. pulse width and relaxation delay) for one molecule may cause resonances of other molecules to vanish.
Figure 4.40. NMR spectra of synthetic LNT (800 MHz, DQF-COSY at 500 MHz, D₂O, 20 mg, generously supplied by Dr. Thomas Fenger and Prof. Robert Madsen). From top left: 1D ¹H, 1D ¹³C, 2D DQF-COSY, gHSQC, gHMBC, and gHSQC-TOCSY.

Most methods of analysis however have their drawbacks when it comes to mixture analysis, and NMR has its strengths within organic molecules and structural isomers. Examples of
impurities often observed in HMO samples were various reactants, solvents, and buffers, usually resulting from side reactions or inadequate purification, e.g. glycerol was found in many samples from enzymatic reactions due to its use as stabilizer for frozen proteins. Changes in chemical shifts compared to the pure reference samples were often observed.

For analysis of mixtures where the constituents had a high degree of resonance overlap in 1D as well as 2D spectra it could be difficult to establish whether this was simply pH effects or whether a structural modification had occurred. This highlights the importance of combining several methods of analysis, why we should use a combination of NMR and MS for most purposes. Finding a good source of fucose for enzymatic synthesis proved difficult, and in 2012 a large order of LNFP-V and LNnFP-V (Table 4.5) was placed with one of the HMO-producing companies. The two HMOs were intended for development of the biological assays of WP4, and two samples were submitted to NMR analysis prior to use. The compounds were assigned by 1D and 2D NMR spectroscopy and it was found that the samples contained substantial carbohydrate impurities of which some were the corresponding non-fucosylated tetrasaccharides (Figure 4.41).

Other impurities were not fully structure elucidated, however one abundant impurity with anomeric resonances at 5.12 ppm and 5.00 ppm displayed strong NOE correlations between the two anomeric protons mentioned. This suggested some resemblance to trehalose (two α-D-Glc connected by an α,α-1,1-glycosidic bond), although the asymmetry in the chemical shift means the two monomers were not identical. Nevertheless, it meant that these
samples were not suitable for biological assays as it would be difficult to know if a certain response was caused by the pentasaccharides or any of the impurities.

4.5.4 CGMP as sialic acid donor

CGMP is a relatively small hydrophilic peptide which is released from κ-casein by the enzyme chymosin during cheese production. The κ-casein is cleaved between Phe105 and Met106 yielding this 64 amino acid peptide from the C-terminal end of κ-casein, with phosphorylation and glycosylation sites at various Thr and Ser residues. Five different oligosaccharide structures are found in bovine CGMP, of which two are neutral in charge (no Neu5Ac), two are monosialylated and one is disialylated. As a Neu5Ac-containing dairy byproduct, CGMP was widely used in the HMO project as a sialic acid donor for enzymatic reactions, and as sialic acid is difficult to obtain from natural sources, this seemed to be the optimal donor. In the oligosaccharides bound to the protein backbone, Neu5Ac is known to bind to the 3-position of Gal and the 6-position of GalNAc. The glycosylation of CGMP is non-uniform both regarding the glycosylation sites and the size of the oligosaccharides, and some CGMP contains no Neu5Ac at all. There is evidence suggesting that CGMP forms aggregates in solution and it is unclear how much of the sialic acid in CGMP is actually accessible to enzymes, and thus what the true sialic acid utilization potential of CGMP is. Many enzymes are specific for certain Neu5Ac linkage types and therefore it is important to know the relative amounts of the different Neu5Ac species. We therefore decided to investigate CGMP by NMR spectroscopy and in particular the ratio of 3- and 6-bound Neu5Ac. We used Lacprodan CGMP-20 (Arla Foods Ingredients, VibyJ, Denmark), which is a food grade protein powder product. The CGMP-20 contains 5.7 % (w/w) Neu5Ac as determined by our collaborators (Paper III). As expected, both α2-3- and α2-6-linked Neu5Ac variants were found in CGMP-20, however from visual inspection of the NMR data it appeared that the α2-3-linked Neu5Ac was slightly more abundant in this product (Figure 4.42). Assuming that the NMR parameters for these Neu5Ac-species in CGMP are similar, integration of the 2D NOESY (600 ms mixing) and DQF-COSY cross peaks between H3\textsubscript{eq} and H3\textsubscript{ax} indicated about 15-20 % less Neu5Ac bound to the 6-position than to the 3-position of GalNAc/Gal. Judging by the DQF-COSY cross peak shape, especially the α2-6-linked Neu5Ac seems to be present in more than one structural environment reflected by the diagonal elongation of the H3\textsubscript{eq}-H3\textsubscript{ax} cross peak structure. This is likely due to the α2-3-linked Neu5Ac being bound to the nonreducing end of the oligosaccharide, whereas the α2-6-linked Neu5Ac is found as a branching point closer to the peptide backbone.
The molecular weight of CGMP is about 8 kDa whereas the aglyco-CGMP weighs 6.8 kDa. This means that the average CGMP has glycosylations corresponding to a total of 1.2 kDa, however the extent of glycosylation varies. The five common oligosaccharide fragments on CGMP each contribute between 203 and 948 Da to the molecular weight, however the larger sialylated tri- and tetrasaccharides (657-948 Da) are most abundant. Roughly estimated, this means that about two Neu5Ac units can be found in the average CGMP molecule, and judging by the ratio of α2-3- and α2-6-linked Neu5Ac, an α2-3-trans-sialidase may thus be able to utilize one Neu5Ac per CGMP molecule. This is in reasonable agreement with the 5.7 % (w/w) Neu5Ac in CGMP-20 which for an 8 kDa CGMP corresponds to 1.5 Neu5Ac units per molecule (both 3- and 6-bound). Although this seems like a small amount, it has to be taken into account that there are not many cheap and easily accessible Neu5Ac sources available, and transglycosylation enzymes cannot utilize free Neu5Ac. Therefore binding of Neu5Ac to a donor molecule is advantageous for enzymatic utilization.

### 4.5.5 Sialyl-oligosaccharides from trans-sialidases

At the start of the HMO project the first target compound was 3’-SL, and about a year after joining the project, Dr. Jesper Holck and Dr. Malwina Michalak had developed procedures
for production of 3’-SL of high purity by use of native and mutated trans-sialidases from *Trypanosoma* spp. expressed in *Pichia pastoris* (Paper III and IV). *Trypanosoma* spp. are protozoan parasites or single-celled animal-like protists known to cause tropical infections such as Chaga’s disease. As this is therefore not an ideal production host, the enzymes were expressed in *P. pastoris*, which is also used for expression of recombinant proteins in the biotechnological industry. The 3’-SL was structure elucidated by MS and NMR spectroscopy and two research papers were published on the results, which will therefore not be further elaborated here (Paper III and IV). It was also attempted to sialylate GOS, with enzyme and purification work by Dr. Jesper Holck and Yao Guo. It was speculated that this might improve the prebiotic properties and afford an easy route to a relatively simple mixture, mimicking the biological roles of HMOs. Structure elucidation of the different constituents of GOS was not attempted by 2D NMR due to the high homogeneity. However, treatment of GOS with a trans-sialidase in the presence of CGMP, and following fractionation by ion exchange chromatography to remove unreacted GOS, yielded a fraction assumed to contain sialyl-GOS. By comparing the NMR spectra with spectra of GOS and sialyllactoses this assumption was confirmed. Several resonances were indicative of the Neu5Acα2-3Galβ motif as shown in Figure 4.43. The grey boxes highlight the Gal resonances in close proximity with Neu5Ac as well as the H3 resonances of Neu5Ac. The acetyl (Ac) group of Neu5Ac is also present in the less crowded region of the spectra, however singlet resonances at this position are common in other molecules and therefore not discriminative (e.g. acetic acid and GlcNAc). By examining and comparing 1D $^1$H and 2D DQF-COSY spectra it was clear that not only lactose had reacted to form 3’-SL but other Neu5Ac-containing oligosaccharides were present, as seen from the group of resonances slightly shifted from the 3’-SL resonances (Figure 4.44). Especially the resonances indicated by grey rectangles are clearly related to the sialylation. Due to the limited amount of material (low S/N ratio) and the resonance overlap, it was not possible to investigate the structures further by use of other 2D homonuclear and heteronuclear NMR methods.

The strength of NMR in this case is the unambiguous assignment of all Neu5Ac as being bound to the 3-position of Gal. In case of a α2-6-linked Neu5Ac, the H3 resonances around 1.80 ppm (axial) and 2.75 ppm (equatorial) would be shifted further upfield similar to what is seen in 6’-SL. Other linkage types would also be expected to result in chemical shift changes for the two H3 protons of Neu5Ac.

NMR spectroscopy was used to investigate the linkages formed from sialylation of GOS and lactose, and it was shown that the purified products contained exclusively α2-3 linked
Neu5Ac. To determine the number, relative amounts, and sizes of different sialyl-GOS molecules, LC-MS would be superior.

Figure 4.3. Partly annotated $^1$H NMR spectra (500 MHz, D$_2$O) of sialyl-GOS (Dr. Jesper Holck), 3’-SL, GOS, and lactose (from the top). Grey rectangles indicate distinct resonances for molecules containing Neu5Acα2-Glcβ-Galβ.

Figure 4.4. Overlay of DQF-COSY spectra for sialyl-GOS (red), GOS (blue) and 3’-SL (grey). The relative resonance intensity and contour levels were individually optimized for visual inspection.
4.5.6 Trans-fucosylations

A HMO-related project (with Mateusz Lezyk, DTU Chemical Engineering) involved the search for trans-fucosidases for production of fucose-HMOs using lactose and \( p \)-nitrophenyl \( \alpha \)-L-fucose (pNP-Fuc) as substrates. After screening for enzymes in a soil metagenomic library, a small group of enzymes were selected for characterization and assaying of their trans-fucosidase activity. However, the yields were very low and it was not possible to identify reaction products by high performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) for which no standards were available. We therefore started a collaboration to use NMR spectroscopy to identify products observed in higher yields than the fucosyllactoses (2’-FL and 3-FL) which were the primary targets. It was believed that the major constituents could be difucosyllactoses, however by using NMR spectroscopy it was quickly established that the lactose was not the optimal acceptor molecule for the nine enzymes investigated. In fact the fucosyllactoses (2’-FL and 3-FL) were present in too low amounts to confirm their presence by 1D and 2D NMR spectroscopy. The major reaction taking place was hydrolysis of pNP-Fuc, as seen by the presence of free fucose in the postreaction mixture. In analogy with the trans-sialidases, the trans-fucosidases also act as fucosidases, and therefore hydrolyze pNP-Fuc. \(^{316}\) The major trans-fucosylation products were identified and structure elucidated by 2D NMR spectroscopy, revealing that pNP-Fuc or Fuc acts as the glycosyl acceptor to form the pNP-Fuc-Fuc glycosides \( p \)-nitrophenyl \( \alpha \)-L-fucopyranosyl-(1-2)-\( \alpha \)-L-fucopyranoside as well as the 1-3- and 1-4-linked pNP-Fuc-Fuc glycosides (Figure 4.45).

The products were identified and structure elucidated in postreaction mixtures after removal of the enzyme by ultrafiltration (5 kDa cutoff). The results were confirmed by \(^1\)H NMR spectroscopy.
analysis of purified products in minute amounts (Figure 4.46), which enabled linking the chromatographic peaks to the different pNP-Fuc-Fuc glycosides.

Since no commercial standards were available for these products, it was initially attempted to relate $^1$H NMR integral values to integrals from the PAD response (Figure 4.47). This was done by calibrating the $\beta$-Gal H1 resonance of lactose to an integral value of 1 and then using the known anomeric resonances of the products as well as pNP-Fuc to compare the relative amounts of trans-fucosylation and hydrolysis products. Besides the fucosyllactoses there were other minor resonances not accounted for in Figure 4.47. Comparing amounts between the different enzymes should be done with caution as this requires exactly equal amounts of lactose and solvent in the samples as well as comparable relaxation properties. Although this can be assumed to some extent to be fulfilled, it is still best to use the results to compare enzyme regioselectivity qualitatively. By using NMR spectroscopy it was shown that these enzymes primarily act as fucosidases, hydrolyzing the donor, but analyzing reaction mixtures it was possible to investigate the product specificity of the enzymes and
showing that the primary reaction products were self condensation pNP-Fuc-Fuc glycosides. For experimental details and NMR data see Paper V.

![Figure 4.47. Diagram of $^1$H NMR integral values of pNP-Fuc and reaction products relative to lactose in postreaction mixtures from 9 different trans-fucosidases. WT=wildtype, MUT=mutant.](image)

### 4.5.7 Enzymes in reaction monitoring

Studying enzymatic reactions over time while monitoring the disappearance and build-up of reactants and products can yield information of the kinetics of an enzymatic reaction. In collaboration with Dr. Jesper Holck from DTU Chemical Engineering we wanted to monitor the reaction of a trans-sialidase with CGMP and lactose as glycosyl donor and acceptor, respectively. The reactions were run as described in the experimental section 4.7.

We hoped to see a change in the spectrum over time, however it turned out any reaction product was formed in so small amounts that it was indiscernible due to the large resonances of CGMP. The trans-sialidase resonances were broad and almost not visible in the spectra, but the CGMP, being a relatively small peptide in high concentration, made it impossible to discern the sialyllactose formed in minute amounts. There was no baseline separation at places where resonances of $3'$-SL were expected and even comparison to a spectrum of CGMP on its own did not help to simplify matters. Buffer was not added to the CGMP sample, which would have facilitated an easier comparison between the spectra. However it
would not have changed the conclusion that this low yielding enzymatic reaction in the presence of the large CGMP resonances was not optimal for reaction monitoring by 1D $^1$H NMR.

Later in the project we were asked to examine another trans-sialidase in a similar fashion. In this case our collaborators (Rui Xue and Yao Guo from DTU Chemical Engineering) were interested in determining the identity of some byproducts they could observe by thin layer chromatography (TLC), and if possible study the kinetics of the reaction. Using lactose and pNP-Neu5Ac as substrates they could observe the formation of 3'-SL and after one hour it disappeared while other products formed. They also attempted to use 3'-SL and lactose as substrates (in a 2:1 ratio), and it appeared that the 3'-SL was converted into free Neu5Ac and 6'-SL, judging by TLC and comparing to the pure standards. $^1$H NMR spectra of the different carbohydrates in the sample (Figure 4.48) were acquired, along with spectra of the citric acid/disodiumphosphate buffer and the enzyme solution (data not shown). The enzyme sample showed mainly resonances from the buffer and glycerol. A negative control and a postreaction sample were freeze dried and redissolved in D$_2$O to assess the resonance overlap at start and finish (Figure 4.48).

![Figure 4.48. $^1$H NMR spectra (500 MHz, D$_2$O, 25 °C) of carbohydrates found in the reaction mixture. NC is a negative control without enzyme, and R is the reaction after completion (then freeze dried and redissolved in D$_2$O, although not all H$_2$O was removed). Regions chosen for integration are highlighted in the NC and R by grey boxes. From the left, the six boxes represent 3'-SL Gal H1; Lactose and 6'-SL Gal H1; 3'-SL Gal H3; α-Neu5Ac H$_3$eq both free and in sialyllactoses; free β-Neu5Ac H$_3$eq; 6'-SL Neu5Ac H$_3$ax. Glycerol resonances are labelled with * in the R spectrum.](image-url)
The spectra of the negative control (NC) and the postreaction mixture (R) displayed distinct differences. There were no resonances from the enzyme which had been broadened so much that they completely disappeared. There were easily recognizable resonances from 3′-SL, 6′-SL, and Neu5Ac, although the Neu5Ac chemical shifts appeared to be shifted upfield due to the effect of the buffer. The second order spin system at 2.6 ppm came from citric acid, just downfield from the equatorial H3 resonances, which for 6′-SL was shifted slightly upfield compared to 3′-SL. The reaction mixture contained large resonances from glycerol, which came from the enzyme solution (marked with * in Figure 4.48). The Neu5Ac H3_α resonances appear between 1.6 and 1.9 ppm, and they are different for the three Neu5Ac-containing samples, although the free β-Neu5Ac resonance is found closer to the 3′-SL H3_α in the buffered solution. The free α-Neu5Ac H3 resonances are found at 1.7 and 2.7 ppm, more or less overlapping with those of 6′-SL, however the α-anomer constitutes only about 5% of the total free Neu5Ac in solution. The Gal H1 resonance at 4.5 ppm for 3′-SL is completely separated from those of lactose and 6′-SL which are overlapping, however the water resonance turned out to add to the integral and therefore the Gal H3 resonance at 4.1 ppm yielded a more accurate measurement. The H3_α at 1.7 ppm for the 6′-SL is separated from those of 3′-SL and β-Neu5Ac. For Neu5Ac the H3_eq of the β-anomer at 2.2 ppm was isolated. All this together formed a good basis for reaction monitoring by NMR as there were separated resonances for the 3′-SL, the 6′-SL, and the β-Neu5Ac (Figure 4.48). The reactions were run in H$_2$O to prevent the heavier D$_2$O from having an effect on the reaction kinetics. This was done by inserting a glass capillary containing D$_2$O into the NMR tube, which ensured sufficient deuterium resonance for the lock. We could not stir the solution inside the NMR tube and we did not spin the samples, so therefore the free diffusion in the sample as opposed to magnetic stirring may have had an effect. In the case of a slow reaction, the diffusion would however have limited influence on the kinetics. See experimental details in section 4.7.

The integrals were converted to concentrations by calibrating the integral at 2.7 ppm (Figure 4.48) to the starting concentration of 3′-SL (20 mM). This integral also covered the 6′-SL and the free α-Neu5Ac (but not the β-Neu5Ac), however the hydrolysis reaction was slow and this calibration was therefore expected to be an acceptable approximation in comparison to the uncertainty of integration in general. At $t = 3$ min the concentrations of β-Neu5Ac and 6′-SL were measured to 0.06 mM and 0.54 mM, and therefore the error resulting from this approximation was negligible. The integral of the 3′-SL Gal H3 also confirmed this by starting from 20 mM (Figure 4.49).
Figure 4.49. Graph of concentrations over time and test of the reproducibility for enzyme kinetics experiments measured by $^1$H NMR. Integrals were measured by $^1$H NMR and calibrated to the starting concentration of 3’-SL. Starting conditions were 20 mM 3’-SL, 10 mM lactose, 50 mM phosphate buffer, and 50 µl enzyme (37 °C, H$_2$O, 500 MHz).

Figure 4.50. Enzymatic reactions of a trans-sialidase with 20 mM 3’-SL and 10 mM lactose (2, ▲), no lactose (6, ◦), or 100 mM lactose (7, ●).
The Gal H1 integrals had systematic errors as they were closest to the water resonance and the 6'-SL and lactose resonances were combined; however the slopes of these curves followed those of the corresponding sialyllactoses, and nevertheless the two curves were redundant. The reaction was repeated three times under the same conditions to test the reproducibility of the method. There were small deviations in the concentrations of substrates and products within a 2 mM window for the final concentrations (Figure 4.49). The reaction was repeated without lactose and with a ten times higher lactose concentration, and surprisingly lactose slowed down the reaction, most likely as a competitive inhibitor (Figure 4.50). It was believed that the enzyme could not use 6'-SL as a substrate. To test this hypothesis the reaction was run with 6'-SL and lactose which resulted in hydrolysis of 6'-SL, but to lesser extent than for 3'-SL. After 19 hours, 3 mM of β-Neu5Ac had been formed in this reaction (data not shown). The project came to a halt when priorities changed and lack of enzyme prevented further analyses.

### 4.6 Conclusions and perspectives

The initial goal of production and analysis of 10 HMOs was not achieved and we took part in parallel identification of 3'-sialyllactose products from different trans-sialidases. Various commercial HMOs were analyzed by NMR spectroscopy to create a small HMO NMR database. This was believed to afford easier mixture analysis when necessary; however it was sometimes problematic to compare spectral data for pure substances in D₂O to buffered systems in H₂O with different ionic strengths and pH. Nevertheless we identified various impurities and oligosaccharides, usually in complex mixtures, which resulted in contributions to Paper III, IV and V. Reaction monitoring of enzymatic processes was achieved by ¹H NMR in H₂O, and it was possible to correlate integral values to concentration. Even though the accuracy of the integrations can be questioned, the slope of the curves should not be, and the results with this reaction show good potential for kinetic investigations of trans-sialidases by ¹H NMR. CGMP works well as a sialic acid donor in enzymatic reactions; however kinetics studies of low-yielding reactions in the presence of CGMP proved difficult as the ¹H NMR resonances of CGMP dominated in the high concentrations necessary for adequate Neu5Ac availability. An attempt to investigate the reaction kinetics using smaller Neu5Ac donors such as 3'-SL were initiated, which afforded adequate resonance separation to monitor the changes in reactants and products overnight.
It was revealed that lactose seems to act as a competitive inhibitor for the substrate binding site of the trans-sialidase enzyme, shown by the deceleration of the reaction with higher lactose concentration and acceleration at lower concentration and without lactose present. The reaction is slow for an enzymatic reaction, and altogether this suggests that the enzyme’s interactions with donor and acceptor could be optimized further. Despite the anticipations that 6’-SL was not a substrate for this enzyme, it was shown that 6’-SL was slowly hydrolyzed to form Neu5Ac and lactose.

In a related project, a variety of enzymes were assessed for their trans-fucosidase activities, and by analyzing postreaction mixtures by 1D and 2D NMR spectroscopy it was shown that the enzymes’ major activities were as fucosidases, hydrolyzing the pNP-Fuc substrate. The desired trans-fucosylation reaction would be to use lactose as a glycosyl acceptor and pNP-Fuc as donor to form fucosyllactoses, however the major products besides free Fuc were identified as pNP-α-L-Fuc-(1-\(n\))-α-L-Fuc, where \(n\) is 2, 3, and 4. Fucosyllactoses were not detected by NMR spectroscopy but by HPAEC chromatography in small amounts, and therefore further screening of trans-fucosidases for production of fucosyllactoses is necessary.

A general conclusion drawn from this project is that production of HMOs in large scale for human consumption is still a goal that is difficult to reach. Discovering appropriate methods, and in particular finding the right enzymes, has been more difficult than expected. Enzymes able to sialylate and (to limited extent) fucosylate lactose and higher oligosaccharides have been obtained, even though the yields are very low. The HMO backbone was not synthesized enzymatically, and therefore it seems that production of HMOs by current technologies would be best undertaken by joint efforts of biotechnology and synthetic chemistry. Several HMO backbone structures were synthesized within the HMO project by Dr. Camilla Jennum, Dr. Thomas Fenger and Prof. Robert Madsen (DTU Chemistry) and enzymatic decoration with fucose and sialic acid could potentially form a basis for HMO production, however both synthetic and enzymatic approaches must be further improved before this can become reality.
4.7 Experimentals

4.7.1 Reaction monitoring with CGMP
CGMP (10.5 mM), lactose (30.0 mM), and citric acid/disodiumphosphate buffer (50 mM) were mixed in 500 μl D$_2$O and heated to 30 °C (Holck 2013). The NMR magnet was shimmed, tuned and matched and a $^1$H spectrum was acquired. The enzyme was added (20 μl, 1.1 g/l final concentration), the time was noted and the magnet was again quickly shimmed manually. Then a $^1$H spectrum was acquired approx. every 9 minutes overnight.

4.7.2 Reaction monitoring with 3’-SL
When starting a reaction in the 500 MHz NMR spectrometer (see spectrometer details in the Appendix) the magnet was prepared (shimming, tuning and matching) with the sample inside at 37 °C (before enzyme addition) and a $t_0$ reference spectrum was acquired. Addition of 50 μl enzyme solution changed the concentrations in the samples, and therefore the first usable data points were after addition of enzyme and initial shimming (approx. 3 min). The first few measurements (8 scans each) were acquired manually while shimming between the measurements and then after 10-15 min the spectrometer was set up to acquire a spectrum every 10 min with automatic shimming on the z1 and z2 coils for about 60 seconds between acquisitions. The data was saved as a 2D experiment and processed in Topspin 3.0 with the ‘multicmd’ and ‘multi_integ’ AU programs. The initial manual measurements were included and used as the first data points by ensuring that the time for each measurement was noted along with the exact start and finish time of the automatic acquisition, which could then be divided by the number of 1Ds to give the exact time between each measurement.
5 Overall conclusions and perspectives

This thesis contains the parallel results of three different projects. Chapter 2 describes the development of new NMR experiments for measurement of coupling constants; chapter 3 was on identification of secondary metabolites from microbial sources, with the hope of finding new bioactive molecules; and chapter 4 gave a description of the work undertaken for the common goal of producing human milk oligosaccharides, also with biological activities in mind. The experiments developed in chapter 2 can be used for studies of the molecules described in chapter 3 and 4, and in this way the thesis is held together by the general topic: Structural studies of human milk oligosaccharides and other natural products.

A new NMR experiment, named HMBC+, was developed for measurement of $J_{hh}$ coupling constants in small molecules, and it was demonstrated on vinyl acetate and strychnine – a small organic molecule and a natural product. The strength of this experiment lies in the sign sensitive measurement of long-range homonuclear $J_{hh}$ coupling constants, and the results so far point towards a very accurate determination. Further work will show the full potential of this new experiment, but its robustness and ability to sign sensitively determine very small coupling constants makes it superior to existing methods. It is believed that the HMBC+ experiment has potential applications both in oligosaccharide analysis and for determination of relative configuration in small molecules, e.g. by measurement of RDCs.

Chapter 2 also describes work on NMR experiments for sign sensitive measurement of $J_{ch}$ in small molecules, however this is work in progress and the experiment in its current state has narrow applications for determination of coupling constants to quaternary carbon atoms. For proton-rich spin systems, interference from undesired coherences seems to obscure the multiplet patterns and so far the principles behind it have not been understood. Measurement of suitable $J_{ch}$ coupling constants by other methods and comparison of results from the different methods would be the next step in the ongoing work with these experiments. Changes to the pulse sequences may also help elucidate and eventually solve the current problems with this experiment.

A range of new molecules were discovered and described in this thesis. This includes the solonamides and the ngercheumicins: Two families of cyclic depsipeptides isolated from a marine bacterium with the ability to prevent virulence in the pathogenic $S. aureus$ (Paper I
and II). These peptides share some structural traits with the natural quorum sensing inducing AIPs in *S. aureus*, and the ability to inhibit QS in three of four subclasses of *agr* provides important knowledge of the structure-activity relationships of such QS inhibitors.\textsuperscript{214} Future work on this project could involve synthesis of solonamide analogues in attempts to elucidate the pharmacophore as well as improve the activity, and studies of the interaction between solonamide B and its supposed target, the sensor histidine kinase AgrC, would provide important information on the mechanism of action.

Aspergillus fijiensis, a relatively unknown fungal species, was also investigated in the hope of isolating novel secondary metabolites. Regrettably only one new molecule was isolated, an analogue of the pyranonigrins, therefore named pyranonigrin E. Dereplication is an important tool in this research area, and quick update as well as thorough use of metabolite databases is very important when several people are working with the same fungal genera, as it is quite common to find the same secondary metabolites in different species. It turned out that one of the pursued metabolites had recently been isolated by Lene Maj Petersen from *A. aculeatus*. Another supposedly unknown molecule had been labeled as such in the local metabolite database, when in fact the molecule was a contamination from stomacher bags during the extraction, suddenly appearing as a very large peak in all extracts, possibly due to a new production method at the stomacher bag supplier.

Finally, participation in a larger project with the goal of producing human milk oligosaccharides led to the joint efforts of producing 3'-SL by enzymatic methods (Paper III and IV). In a related collaboration, a range of enzymes were screened for their trans-fucosylation abilities, and three “auto”-fucosylation products were identified, both in reaction mixtures and as purified compounds in small amounts (Paper V). This shows the importance of collaborations across various fields of research, and how joint efforts, using several methods of analysis, leads to a clearer interpretation of the results. A great deal of work is still needed before industrial production of complex HMOs is achievable, and work on both enzymes and synthetic procedures seems necessary to reach this ambitious but noble goal.

The HMBC+ experiment is without doubt the most significant contribution of this thesis to the NMR society. With both the experiment and the subsequent processing running smoothly, it is the hope that less experienced NMR users will also make use of this experiment for easier structural analysis of complex molecules in the future.
6 References


Trigo-Mouriño, P.; de la Fuente, M. C.; Gil, R. R.; Sánchez-Pedregal, V. M.; Navarro-Vázquez, A. Conformational Analysis of the Anti-Obesity Drug Lorcaserin in Water: How to


(87) Personal Communication with Ole W. Sørensen, May 2012.


(102) Sørensen, M. D.; Meissner, A.; Sørensen, O. W. $^{13}$C Natural Abundance $^{1}$H and $^{1}$CT Experiments for Measurement of $^{13}$C Coupling Constants between $^{13}$C and $^{1}$H and Other Protons in a Protein. J. Magn. Reson. 1999, 137, 237–242.


References


Appendix

Spectrometers and acquisition ......................................................... A-2

Chapter 2 .......................................................................................... A-2
HMBC+ pulse program ...................................................................... A-2
HMBC+ AU program .......................................................................... A-2
ADEQ11ETGPLR_LKJ3 pulse program ............................................... A-8
DEPT2LR_LKJ pulse program .............................................................. A-16
Strychnine ......................................................................................... A-18
Extended table of calculated $^\text{H}J_{\text{HH}}$ coupling constants in strychnine ............................................... A-20
HMBC+ spectra for strychnine and LNT ............................................ A-21
Brazilin ............................................................................................... A-21

Chapter 3 .......................................................................................... A-25
Pyranonigrin E .................................................................................. A-26
JBIR-74 .............................................................................................. A-26
Decumbenone A ................................................................................ A-27

Chapter 4 .......................................................................................... A-28
3’-sialyllactose (3’-SL) ....................................................................... A-29
6’-sialyllactose (6’-SL) ....................................................................... A-30
2’-fucosyllactose (2’-FL) ................................................................... A-31
3-fucosyllactose (3-FL) ..................................................................... A-32
Lacto-$\text{N}$-tetraose (LNT) .................................................................. A-33
Lacto-$\text{N}$-neotetraose (LNNT) ............................................................. A-34
Lacto-$\text{N}$-fucopentaose V (LNFP-V) .................................................. A-35
Lacto-$\text{N}$-neofucopentaose V (LNNFP-V) ......................................... A-36
pNP-Fuc-Fuc glycosides ................................................................... A-38

Paper I ............................................................................................... A-39
Paper II .............................................................................................. A-67
Paper III ............................................................................................ A-103
Paper IV ............................................................................................ A-119
Paper V .............................................................................................. A-127
Spectrometers and acquisition

- Varian Mercury 300 MHz with a 5 mm 4 nucleus probe (HCFP)
- Bruker Ascend 400 MHz with a 5 mm H – Broadband Dual Channel z-gradient Prodigy cryoprobe
- Varian Inova 500 MHz with a 5 mm HCP probe
- Bruker Avance 800 MHz with a 5 mm TCI cryoprobe (at the Danish Instrument Centre for NMR Spectroscopy of Biological Macromolecules at the Carlsberg Laboratory, Valby, Denmark).

Unless otherwise noted, experiments were acquired using standard pulse sequences at 25 °C, referenced to solvent residual resonances (see below). Zero filling in F2 was done once and zero filling in F1 was done once or to 1024 points to enhance the digital resolution. Linear prediction was only used for ADEQUATE-type experiments.

DMSO-d$_6$: 2.49 ppm, $\delta$C 39.50 ppm
CDCl$_3$: 7.26 ppm, $\delta$C 77.16 ppm
In D$_2$O and H$_2$O, using external 1,4-dioxane: $\delta$H 3.75 ppm, $\delta$C 67.4 ppm

Chapter 2

HMBC+ pulse program

; HMBCplus.c
; pulse program by
; LOKJA, AWBN, CHG, and OWS (May 2014)
; for measurement of magnitude and sign of long-range J(HH)
; Echo/antiecho gradient selection
; With third order low-pass J-filter to suppress one-bond correlations

#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

; NB: set TD1 to 4 or the program will NOT work

; compacted pulseprogram with minor modifications.
; uses NBL so that we can acquire the HMBC+- to the
; same location with only 50% of the time + and 50%
; -.

;SER file: HMBC, HMBC+-, HAT-, HAT+

; the pulseprogram can be thought of as three pulseprograms with different delays
; which we can define "on the fly" in the pulseprogram
define delay AB_delay1
define delay AB_delay2
define delay AB_delay3
define delay AB_delay4
define delay AB_delay5  
define delay AB_delay6  
define delay AB_delay7  

; this loopcounter specifies which experiment type will be run:  
; 0 (std) , 1 (-) or 2 (+)  
define loopcounter updown  

; this loopcounter is used to control when the decoupler is turned off after the first proton 90  
; 1 (HMBC +/-) or 0 (std and HAT)  
define loopcounter HMBCED  

define delay EPS  
define delay EPSA  
define delay LP1  
define delay LP2  
define delay LP3  

; gH=26.752196, gC=6.72828  
; EA = gH+gC : gH-gC (approx.: 5:3)  
define list<gradient> EA1 = {1.0000 -0.5981}  
define list<gradient> EA2 = {1.0000 0.5981}  
define list<gradient> EA3 = {0.5981 1.0000}  

"d10=3u"  
"d11=30m"  
"d12=20u"  
"p2=p1*2"  
"p4=p3*2"  

"TAU = (1s/(2*cnst9))" ; Tau, probably best to set this according to the modal average  
"TAU1 = (1s/((cnst6 + 0.07*(cnst8 - cnst6))*2))"  
"TAU2 = (1s/(cnst6 + cnst8))"  
"TAU3 = (1s/((cnst8 - 0.07*(cnst8 - cnst6))*2))"  
"DELTA = 50u + p16 + d16" ; gradient duration  
"EPS=p2+d10*2"  
"EPSA=EPS+p4"  

"LP1 = TAU1 - DELTA - 2*d12" ; LPJF  
"LP2 = TAU2 - DELTA - p3" ; LPJF  
"LP3 = TAU3 - DELTA - p3" ; LPJF  

"l0=0"  
"updown=0"  
"HMBCED=0"  
"inf1=1inf2/2"  

aqseq 312

1 ze
10u
2 3m
3 3m
30u
4 d11
d1 BLKGRAD  
; handle the HMBC +/- experiments using nsdone  
if "nsdone %2 == 1"  
{  
"AB_delay1=d12"  
"AB_delay2=TAU+0.5*EPSA"  
"AB_delay3=d12"  
"AB_delay4=d6-2*DELTA"  
"AB_delay5=TAU+EPS-DELTA"  
"AB_delay6=TAU-DELTA"  
"AB_delay7=EPSA"  
}
"updown=1"
"HMBCED=1"
}
else
{
"AB_delay1=d12"
"AB_delay2=TAU+0.5*EPSA"
"AB_delay3=d12"
"AB_delay4=d6-2*DELTA+EPSA"
"AB_delay5=TAU-DELTA"
"AB_delay6=TAU-DELTA+EPS"
"AB_delay7=d16"
"updown=2"
"HMBCED=1"
}

; if l0 != 1 then do one of the other experiments
; (this would be more elegant if "else if" was allowed)

; standard HMBC
if "l0==0"
{
"AB_delay1=TAU+0.5*EPSA"
"AB_delay2=d12"
"AB_delay3=TAU+0.5*EPSA"
"AB_delay4=d6-DELTA+0.5*EPSA"
"AB_delay5=0.5*TAU-DELTA"
"AB_delay6=0.5*TAU-DELTA+EPS"
"AB_delay7=DELTA"
"updown=0"
"HMBCED=0"
}

; HAT -
if "l0==2"
{
"AB_delay1=0.5*TAU+d12"
"AB_delay2=0.5*TAU+0.5*EPSA"
"AB_delay3=0.5*TAU"
"AB_delay4=d6-2*DELTA"
"AB_delay5=0.5*TAU+EPSA-DELTA"
"AB_delay6=0.5*TAU-DELTA"
"AB_delay7=0.5*TAU+EPSA"
"updown=1"
"HMBCED=0"
}

; HAT +
if "l0==3"
{
"AB_delay1=0.5*TAU+d12"
"AB_delay2=0.5*TAU+0.5*EPSA"
"AB_delay3=0.5*TAU"
"AB_delay4=d6-2*DELTA+EPSA"
"AB_delay5=0.5*TAU-DELTA"
"AB_delay6=0.5*TAU+EPS-DELTA"
"AB_delay7=d16+0.5*TAU"
"updown=2"
"HMBCED=0"
}

AB_delay1 pl12:f2
AB_delay2 cpd2:f2
(pi ph10):f1
if "HMBCED==1"
{
AB_delay3 do:f2
}
else
{
AB_delay3
  d12 do:f2
}
d12 p12:f2
LP1
50u UNBLKGRAD
p16:gp1
d16
  (p3 ph4):f2
  50u
p16:gp1*-0.5714
d16
LP2
  (p3 ph5):f2
  50u
p16:gp1*-0.2857
d16
LP3
  (p3 ph6):f2
  50u
p16:gp1*-0.1429
d16
AB_delay4
  if "updown==0"
  {
    (p3 ph2):f2
d10
    (p2 ph10):f1
d10
  }
AB_delay5
  50u
p16:gp5*EA1
d16 igrad EA1
  (p4 ph3):f2
  50u
p16:gp5*EA1
d16 igrad EA1
AB_delay6
  (p3 ph1):f2
AB_delay7
}
if "updown==1"
{
  50u
p16:gp5*EA3*-1
d16
  (p3 ph2):f2
  50u
p16:gp5*0.7990
d16
AB_delay5
  (p4 ph3):f2
d10
  (p2 ph10):f1
d10
A\_delay6

50u
p16:gp5*=0.7990
d16
(p3 ph1):f2
50u
p16:gp5*EA3
d16

A\_delay7

) if "updown==2"
{
50u
p16:gp5*EA2*-1
d16
(p3 ph2):f2
50u
p16:gp5*0.7990
d16

A\_delay5

d10
(p2 ph10):f1
d10
(p4 ph3):f2

A\_delay6

50u
p16:gp5*=0.7990
d16
(p3 ph1):f2
50u
p16:gp5*EA2
; no d16 here as we include it in A\_delay7 to prevent it from becoming zero
A\_delay7
}

goscnp ph31
10u st
10u iu0
lo to 2 times nbl ; do the four experiments (HMBC switches between +/- on alternate scans)
3m ipp1 ipp2 ipp3 ipp31 ru0; do phases and ns scans
lo to 3 times ns
d11 mc #0 to 4
F1QP()
F2EA(igrad EA1 & igrad EA2 & igrad EA3 rpp1 rpp2 rpp3 rpp31, id10)
d11 BLKGRAD
exit

; phase cycle
ph1=0 0 2 2 2 2 0 0
ph2=0 0 0 0 2 2 2 2 2 2 0 0 0 0
ph3=0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3
ph4=0
ph5=0
ph6=0
ph10=0
ph31=0 0 2 2

; plw1 :f1 power level for pulse (default)
; plw2 :f2 power level for pulse (default)
; plw12 :f2 power level for CPD/BB decoupling

A-6
Appendix

;p1 : f1 90 degree high power pulse
;p2 : f1 180 degree high power pulse
;p3 : f2 90 degree high power pulse
;p4 : f2 180 degree high power pulse
;p16 : gradient pulse
;d0 : increment delay (3 usec)
;d1 : relaxation delay: 1-5*T1
;d6 : delay for evolution of long range couplings [30 msec]
;d11 : delay for disk I/O
;d12 : delay for power switching [20 usec]
;d16 : delay for homospoil/gradient recovery [200 usec]
;TAU : 1/(J(min) + J(max))
;TAU1 : 1/2(J(min)+0.07(J(max)-J(min))
;TAU2 : 1/(J(min) + J(max))
;TAU3 : 1/2(J(max)-0.07(J(max)-J(min))
;DELTA : Gradient duration, p16 + d16 + 50u
;cnst6 : 1J(min), ~125 Hz
;cnst8 : 1J(max), ~165 Hz
;cnst9 : 1J optimal for editing ~145 Hz
;MC2 : Echo-Antiecho
;EA1 : Echo-Antiecho gradient program, [+5 -3]
;EA2 : Echo-Antiecho gradient program, [+5 +3]
;EA3 : Echo-Antiecho gradient program, [+3 +5]
;pcpd2 : Pulse length for cpd on f2 [65 usec]
;cpdprg2 : Cpdp-program [garp4]
;gpnam1 : Gradient pulse 1 file name [SMSQ10.100]
;gpnam5 : Gradient pulse 5 file name [SMSQ10.100]
;gpz1 : Gradient strength 1, +7 [+12%]
;gpz5 : Gradient strength 5 [+80%]
;ns: 16*n
ds: 32
;Processing:
; F2 : SSB = 4, WDW = QSINE, PH_mod = pk
; F1 : SSB = 2, WDW = QSINE, PH_mod = no
; Au-pgm : Use HMBCplus AU program
; FT : xfb; xf2m; abs2; abs1;
HMBC+ AU program

; HMBCplus AU program
; for processing of HMBCplus datasets in Topspin 3.0
; by LOKJA, AWBN, CHG, and OWS (May 2014)
; It creates 8 expnos:
; 1-2: Edited HMBC (1: C/CH2, 2: CH/CH3)
; 3-4: HAT HMBC (3: C/CH2, 4: CH/CH3)
; 5-6: HMBCplus C/CH2 (5: add, 6: subtract)
; 7-8: HMBCplus CH/CH3 (7: add, 8: subtract)

#define MAXSIZE 32768
#define RSER2DAB(direction, num, eno) \
    {SETCURDATA; (void)sprintf(Hilfs_string,"rser2d s%d %d %d y n", \
        direction,num,eno); \
        AUERR=CPR_exec(Hilfs_string,WAIT_TERM);}

AUERR = hmbcplus(curdat);
QUIT

int hmbcplus(const char* curdat)
{
    char infile1[PATH_MAX], infile2[PATH_MAX], stattext[PATH_MAX];
    char outfile1[PATH_MAX], outfile2[PATH_MAX];
    char path[PATH_MAX + 64];
    int row1[MAXSIZE], row2[MAXSIZE];
    int rowout1[MAXSIZE], rowout2[MAXSIZE];
    int td1s, td2s, td1, td2, tds, td;
    int i, j;
    int oexpno, nexpno, byteorder, parmode, zexpno;
    FILE *fpin1, *fpin2, *fpout1, *fpout2;

    float scaleed=1.05, scalehat=1.0, scaleplus=1.0;

    /***** get dataset and parameters *****/
    FETCHPAR("PARMODE", &parmode)
    if (parmode != 2)
        STOPMSG("Program is only suitable for pseudo 3D data")

    /***** get input *****/
    oexpno = expno;
    nexpno = expno*100 + 1;
    GETINT("Enter first EXPNO to store new dataset:", nexpno)
    if (nexpno <= 0)
        STOPMSG("invalid expno value")

    GETFLOAT("Enter scaling factor for HMBCed (0.8-1.2)", scaleed);
    if((scaleed < 0.8) || (scaleed > 1.2)) STOPMSG("Scaling factor must be between 0.8 and 1.2");
    /* GETFLOAT("Enter scaling factor for HAT HMBC (0.8-1.2)", scalehat);
    if((scalehat < 0.8) || (scalehat > 1.2)) STOPMSG("Scaling factor must be between 0.8 and 1.2");
    GETFLOAT("Enter scaling factor for Plus spectra (0.8-1.2)", scaleplus);
    if((scaleplus < 0.8) || (scaleplus > 1.2)) STOPMSG("Scaling factor must be between 0.8 and 1.2"); */

    /* create the output datasets as 2D dataset */
    /* contains bug fixes needed for some computers */
    RSER2DAB(23, i, nexpno); /* std HMBC */
/* create datasets for the PLUS experiment */
/* we also use them as temp datasets */
RSER2DAB(23,1,nexpno+4);
RSER2DAB(23,2,nexpno+5);
RSER2DAB(23,1,nexpno+6);
RSER2DAB(23,2,nexpno+7);

/* create the HMBCed spectra */
expno = nexpno+4;
(void)strcpy(infile1, ACQUPATH("ser"));
fpin1 = fopen(infile1, "rb");
if (fpin1 == 0)
{
    (void)sprintf(path, "%s:\n%s", "cannot open file for reading", infile1);
    STOPMSG(path)
}
expno = nexpno+5;
(void)strcpy(infile2, ACQUPATH("ser"));
fpin2 = fopen(infile2, "rb");
if (fpin2 == 0)
{
    (void)sprintf(path, "%s:\n%s", "cannot open file for reading", infile2);
    STOPMSG(path)
}
expno = nexpno;
(void)strcpy(outfile1, ACQUPATH("ser"));
fpin1 = fopen(outfile1, "wb");
if (fpin1 == 0)
{
    (void)sprintf(path, "%s:\n%s", "cannot open file", outfile1);
    STOPMSG(path)
}
expno = nexpno + 1;
/* we write to ser2 as we want to remove E/AE later ! */
(void)strcpy(outfile2, ACQUPATH("ser2"));
fpin2 = fopen(outfile2, "wb");
if (fpin2 == 0)
{
    (void)sprintf(path, "%s:\n%s", "cannot open file", outfile2);
    STOPMSG(path)
}

SETCURDATA;
FETCHPARS("BYTEORD", &byteorder)
FETCHPARS("TD", &tds)
tds = ( (tds + 255) / 256 ) * 256;

/* here we get the data from the 2D not the 3D */
FETCHPAR1("TD", &td1)
FETCHPAR1S("TD", &td1s)

for (i=0; i < tds; i++)
{
    fread(row1,sizeof(int),tds,fpin1);
    local_swap4(row1,sizeof(int)*tds,byteorder);
    fread(row2,sizeof(int),tds,fpin2);
    local_swap4(row2,sizeof(int)*tds,byteorder);

    for (j=0; j < tds; j++)
    {
        int row1u = row1[j];
        int row2s = (int)(scaleed * row2[j]);
        rowout1[j] = row1u + row2s;
        rowout2[j] = row1u - row2s;
    }
}
local_swap4(rowout1,sizeof(int)*tds,byteorder);
fwrite(rowout1,sizeof(int),tds,fpout1);
local_swap4(rowout2,sizeof(int)*tds,byteorder);
fwrite(rowout2,sizeof(int),tds,fpout2);
}
fclose(fpin1);
fclose(fpin2);
fclose(fpin1);
fclose(fpout2);

/* remove E/AE from second HMBCed experiment. Because we need to get rid of 90
degree phase error in second HAT subspectrum */
expno = nexpno+1;
(void)strcpy(infile1, ACQUPATH("ser2"));
fpin1 = fopen(infile1, "rb");
if (fpin1 == 0)
{
(void)sprintf(path, "%s:\n%s", "cannot open file for reading", infile1);
STOPMSG(path);
}
(void)strcpy(outfile1, ACQUPATH("ser"));
fpout2 = fopen(outfile1, "wb");
if (fpout2 == 0)
{
(void)sprintf(path, "%s:\n%s", "cannot open file", outfile1);
STOPMSG(path);
}
for (i=0; i < td1s; i+=2)
{
  fread(row1,sizeof(int),tds,fpin1);
  local_swap4(row1,sizeof(int)*tds,byteorder);
  fread(row2,sizeof(int),tds,fpin1);
  local_swap4(row2,sizeof(int)*tds,byteorder);
  for (j=0; j < tds ; j+=2)
  {
    int row_r1 = row1[j];
    int row_r2 = row2[j];
    int row_i1 = row1[j+1];
    int row_i2 = row2[j+1];
    /* see Palmer III, JMagnReson 93(1991), 151-170 */
    /* the second fid has the signs swapped due to the way */
    /* Bruker's FT works */
    rowout1[j] = row_r1 + row_r2;
    rowout1[j+1] = row_i1 + row_i2;
    rowout2[j+1] = (row_r1 - row_r2);
    rowout2[j] = -(row_i1 - row_i2);
  }
local_swap4(rowout1,sizeof(int)*tds,byteorder);
fwrite(rowout1,sizeof(int),tds,fpout2);
local_swap4(rowout2,sizeof(int)*tds,byteorder);
fwrite(rowout2,sizeof(int),tds,fpout2);
}
fclose(fpin1);
fclose(fpout2);
SETCURDATA;
STOREPARN(1, "FnMODE", 4)
unlink(ACQUPATH("ser2")); /* delete tmp ser file */

/* now create the HAT spectra */
expno=nexpno;
SETCURDATA
RSER2DAB(23,3,nexpno+4);
RSER2DAB(23,4,nexpno+5);
Appendix

expno = nexpno+4;
(void)strcpy(infile1, ACQUPATH("ser"));
fpin1 = fopen(infile1, "rb");
if (fpin1 == 0)
{
    (void)sprintf(path, "\%s:\n\%s", "cannot open file for reading", infile1);
    STOPMSG(path)
}
expno = nexpno+5;
(void)strcpy(infile2, ACQUPATH("ser"));
fpin2 = fopen(infile2, "rb");
if (fpin2 == 0)
{
    (void)sprintf(path, "\%s:\n\%s", "cannot open file for reading", infile2);
    STOPMSG(path)
}
expno = nexpno+2;
(void)strcpy(outfile1, ACQUPATH("ser"));
fpout1 = fopen(outfile1, "wb");
if (fpout1 == 0)
{
    (void)sprintf(path, "\%s:\n\%s", "cannot open file", outfile1);
    STOPMSG(path)
}
expno = nexpno+3;
(void)strcpy(outfile2, ACQUPATH("ser2")); /* as for HMBCed */
fpout2 = fopen(outfile2, "wb");
if (fpout2 == 0)
{
    (void)sprintf(path, "\%s:\n\%s", "cannot open file", outfile2);
    STOPMSG(path)
}
SETCURDATA;
FETCHPARS("BYTORDA", &byteorder)
FETCHPARS("TD", &tds)
tds = ( (tds + 255) / 256 ) * 256;
FETCHPAR1("TD", &td1)
FETCHPAR1S("TD", &td1s)
for (i=0; i < td1s; i++)
{
    fread(row1,sizeof(int),tds,fpin1);
    local_swap4(row1,sizeof(int)*tds,byteorder);
    fread(row2,sizeof(int),tds,fpin2);
    local_swap4(row2,sizeof(int)*tds,byteorder);
}
for (j=0; j < tds ; j++)
{
    int row1u = row1[j];
    int row2s = (int)(scalehat * row2[j]);
    rowout1[j] = row1u + row2s;
    rowout2[j] = row1u - row2s;
}
local_swap4(rowout1,sizeof(int)*tds,byteorder);
fwrite(rowout1,sizeof(int),tds,fpout1);
local_swap4(rowout2,sizeof(int)*tds,byteorder);
fwrite(rowout2,sizeof(int),tds,fpout2);
}
fclose(fpin1);
fclose(fpin2);
fclose(fpout1);
fclose(fpout2);

/* remove E/AE from second HAT HMBC experiment */
exnpo = nexpno+3;
SETCURDATA;
(void)strcpy(infile1, ACQUPATH("ser2"));
fpin1 = fopen(infile1, "rb");
if (fpin1 == 0)
{
    (void)sprintf(path, "\%s:\n\%s", "cannot open file for reading", infile1);
    STOPMSG(path)
}
(void)strcpy(outfile1, ACQUPATH("ser3")); /* as we need to phase correct the data afterwards */
fpout2 = fopen(outfile1, "wb");
if (fpout2 == 0)
{
    (void)sprintf(path, "\%s:\n\%s", "cannot open file", outfile1);
    STOPMSG(path)
}

for (i=0; i < td1s; i+=2)
{
    fread(row1,sizeof(int),tds,fpin1);
    local_swap4(row1,sizeof(int)*tds,byteorder);
    fread(row2,sizeof(int),tds,fpin1);
    local_swap4(row2,sizeof(int)*tds,byteorder);

    for (j=0; j < tds ; j+=2)
    {
        int row_r1 = row1[j];
        int row_r2 = row2[j];
        int row_i1 = row1[j+1];
        int row_i2 = row2[j+1];

        rowout1[j] = row_r1 + row_r2;
        rowout1[j+1] = row_i1 + row_i2;
        rowout2[j+1] = (row_r1 - row_r2);    
        rowout2[j] = -(row_i1 - row_i2);
    }
    local_swap4(rowout1,sizeof(int)*tds,byteorder);
    fwrite(rowout1,sizeof(int),tds,fpout2);
    local_swap4(rowout2,sizeof(int)*tds,byteorder);
    fwrite(rowout2,sizeof(int),tds,fpout2);
}
fclose(fpin1);
fclose(fpout2);
SETCURDATA;
STOREPARN(1, "FnMODE", 4)
unlink(ACQUPATH("ser2")); /* delete tmp ser file */

/* now remove the phase error from this dataset */
expno = nexpno+3;
SETCURDATA;
(void)strcpy(infile1, ACQUPATH("ser3"));
fpin1 = fopen(infile1, "rb");
if (fpin1 == 0)
{
    (void)sprintf(path, "\%s:\n\%s", "cannot open file for reading", infile1);
    STOPMSG(path)
}
(void)strcpy(outfile1, ACQUPATH("ser"));
fpout2 = fopen(outfile1, "wb");
if (fpout2 == 0)
{
    (void)sprintf(path, "\%s:\n\%s", "cannot open file", outfile1);
    STOPMSG(path)
}

for (i=0; i < td1s; i+=2)
/* read in rows pairwise so as to allow for phase correction */
fread(row1,sizeof(int),tds,fpin1);
local_swap4(row1,sizeof(int)*tds,byteorder);
fread(row2,sizeof(int),tds,fpin1);
local_swap4(row2,sizeof(int)*tds,byteorder);
for (j=0; j < tds ; j+=2)
{
    int row_1 = row1[j];
    int row_2 = row2[j];
    int row_3 = row1[j+1];
    int row_4 = row2[j+1];

    /* this gives a 90 degree phase correction in F1 */
    rowout1[j] = row_2;
    rowout1[j+1] = row_4;
    rowout2[j] = -row_1;
    rowout2[j+1] = -row_3;
}
local_swap4(rowout1,sizeof(int)*tds,byteorder);
fwrite(rowout1,sizeof(int),tds,fpout2);
local_swap4(rowout2,sizeof(int)*tds,byteorder);
fwrite(rowout2,sizeof(int),tds,fpout2);
}
fclose(fpin1);
fclose(fpout2);
unlink(ACQUPATH("ser3")); /* delete tmp ser file */

/* now make the HMBC PLUS spectra */
exno = nexno;
(void)strcpy(infile1, ACQUPATH("ser"));
fpin1 = fopen(infile1, "rb");
if (fpin1 == 0)
{
    (void)sprintf(path, "%s:\n%s", "cannot open file for reading", infile1);
    STOPMSG(path)
}
exno = nexno+2;
(void)strcpy(infile2, ACQUPATH("ser"));
fpin2 = fopen(infile2, "rb");
if (fpin2 == 0)
{
    (void)sprintf(path, "%s:\n%s", "cannot open file for reading", infile2);
    STOPMSG(path)
}
exno = nexno+4;
(void)strcpy(outfile1, ACQUPATH("ser"));
fpout1 = fopen(outfile1, "wb");
if (fpout1 == 0)
{
    (void)sprintf(path, "%s:\n%s", "cannot open file", outfile1);
    STOPMSG(path)
}
exno = nexno+5;
(void)strcpy(outfile2, ACQUPATH("ser"));
fpout2 = fopen(outfile2, "wb");
if (fpout2 == 0)
{
    (void)sprintf(path, "%s:\n%s", "cannot open file", outfile2);
    STOPMSG(path)
}
for (i=0; i < td1s; i++)
{
    fread(row1,sizeof(int),tds,fpin1);
local_swap4(row1,sizeof(int)*tds,byteorder);
forread(row2,sizeof(int),tds,fpin2);
local_swap4(row2,sizeof(int)*tds,byteorder);

for (j=0; j < tds ; j++)
{
    int rowlu = row1[j];
    int row2s = (int)(scaleplus * row2[j]);
    rowout1[j] = rowlu + row2s;
    rowout2[j] = rowlu - row2s;
}
local_swap4(rowout1,sizeof(int)*tds,byteorder);
fwrite(rowout1,sizeof(int),tds,fpout1);
local_swap4(rowout2,sizeof(int)*tds,byteorder);
fwrite(rowout2,sizeof(int),tds,fpout2);
}
fclose(fpin1);
fclose(fpin2);
fclose(fpout1);fclose(fpout2);

expno = nexpno+1;
(void)strcpy(infile1, ACQUPATH("ser"));
fpin1 = fopen(infile1, "rb");
if (fpin1 == 0)
{
    (void)sprintf(path, "%s:\n%s", "cannot open file for reading", infile1);
    STOPMSG(path)
}
expno = nexpno+3;
(void)strcpy(infile2, ACQUPATH("ser"));
fpin2 = fopen(infile2, "rb");
if (fpin2 == 0)
{
    (void)sprintf(path, "%s:\n%s", "cannot open file for reading", infile2);
    STOPMSG(path)
}
expno = nexpno+6;
(void)strcpy(outfile1, ACQUPATH("ser"));
fpout1 = fopen(outfile1, "wb");
if (fpout1 == 0)
{
    (void)sprintf(path, "%s:\n%s", "cannot open file", outfile1);
    STOPMSG(path)
}
expno = nexpno+7;
(void)strcpy(outfile2, ACQUPATH("ser"));
fpout2 = fopen(outfile2, "wb");
if (fpout2 == 0)
{
    (void)sprintf(path, "%s:\n%s", "cannot open file", outfile2);
    STOPMSG(path)
}
for (i=0; i < td1; i++)
{
    fread(row1,sizeof(int),tds,fpin1);
    local_swap4(row1,sizeof(int)*tds,byteorder);
    fread(row2,sizeof(int),tds,fpin2);
    local_swap4(row2,sizeof(int)*tds,byteorder);
    for (j=0; j < tds ; j++)
    {
        int rowlu = row1[j];
        int row2s = (int)(scaleplus * row2[j]);
        rowout1[j] = rowlu + row2s;
        rowout2[j] = rowlu - row2s;
    }
    local_swap4(rowout1,sizeof(int)*tds,byteorder);
fwrite(rowout1,sizeof(int),tds,fpout1);
local_swap4(rowout2,sizeof(int)*tds,byteorder);
fwrite(rowout2,sizeof(int),tds,fpout2);
}
fclose(fpin1);
fclose(fpin2);
fclose(fpout1);
fclose(fpout2);

expno=nexpno+6;
SETCURDATA;
STOREPARNS(1, "FnMODE", 4)
STOREPARN(1, "FnMODE", 4)
expno=nexpno+7;
SETCURDATA;
STOREPARNS(1, "FnMODE", 4)
STOREPARN(1, "FnMODE", 4)

expno=oexpno;
SETCURDATA;

for(i=0; i < 4; i++)
{
expno=nexpno+i;
STOREPARN(1,"PH_mod",0);
SETCURDATA;
XF2;
XF1;
XF2M;
}

for(i=4; i < 8; i++)
{
expno=nexpno+i;
STOREPARN(1,"PH_mod",0);
SETCURDATA;
XF2;
XF1;
XF2M;
}
return;
}
ADEQ11ETGPLR_LKJ3 pulse program

;adeq1letgplr_LKJ3
;avance version (May 2014)
;Modified 1.1-ADEQUATE
;for longrange 2JCH couplings in F1 dimension
;phase sensitive using Echo/Antiecho gradient selection
;with decoupling during acquisition
;and new improved phase cycling scheme with 45 degree steps
;
; Program by LOKJA-CHG-OWS
;
; Based on the pulse sequence by
;
;$CLASS=HighRes
;$DIM=2D
;$TYPE=
;$SUBTYPE=
;$COMMENT=

#include <Avance.incl>
#include <Delay.incl>
#include <Grad.incl>

"p2=p1*2"
"d0=3u"
"p4=p3*2"
"d11=30m"
"d4=1s/(cnst2*4)"
"d23=1s/(cnst3*4)"
"in0=inf1/2"

"DELTA=d4-p16-d16-6u"

1 ze
   d11 p112:f2
   2 d1 do:f2
3 (p1 ph2)
   d4 p12:f2
   (center (p2 ph1) (p4 ph1):f2 )
  d4
   (p1 ph1) (p3 ph3):f2
  d23
   (p4 ph3):f2
  d23 UNBLKGRAD
   (p3 ph3):f2
   d0
; 180x pulse removed to obtain multiplicity patterns in F1
   d0
   3u
   p16:gp1*EA
   d16
   (p4 ph4):f2
   3u
   p16:gp2*EA
   d16
   6u ; compensate initial d0 to refocus 13C for t1=0
   (p3 ph1):f2
   d23
   (p4 ph1):f2
   d23
   (center (p1 ph1) (p3 ph1):f2 )
   d4
Appendix

Appendix

```
(cenrer (p2 ph1) (p4 ph1):f2 )
DELTA
3u
pl6:gp3
d16 pl12: f2
4u BLKGRAD
go=2 ph31 cpd2: f2
d1 do: f2 mc #0 to 2 FLEA(igrad EA, id0)
exit

ph1=0
ph2=1
ph3=0 1 2 3 1 2 3 0 2 3 0 1 3 0 1 2
ph4=(8) 0 4 2 6 3 7 1 5 6 2 4 0 5 1 7 3 2 6 0 4 1 5 3 7 4 0 6 2 7 3 5 1
ph31=0 2

;p11 : f1 channel - power level for pulse (default)
p12 : f2 channel - power level for pulse (default)
p112 : f2 channel - power level for CPD/BB decoupling
p1 : f1 channel - 90 degree high power pulse
p2 : f1 channel - 180 degree high power pulse
p3 : f2 channel - 90 degree high power pulse
p4 : f2 channel - 180 degree high power pulse
p16 : homospoil/gradient pulse
;d0 : incremented delay (2D) [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d4 : 1/4J(CH)
;d11 : delay for disk I/O [30 msec]
;d16 : delay for homospoil/gradient recovery
d23: 1/(4J(CC))
cnst2 : J(CH) = 127 .. 160 Hz
cnst3 : J(CC) = 35 .. 55 Hz
inf1: 1/SW(C) = 2 * DW(C)
in0: 1/(2 * SW(C)) = DW(C)
nd0: 2
;DS: 4 * n (optimal is 32, minimum is 4)
td1: number of experiments
FnMODE: echo-antiecho
;cpd2: decoupling according to sequence defined by cpdprg2
;pcpd2: f2 channel - 90 degree pulse for decoupling sequence

;use gradient ratio: gp 1 : gp 2 : gp 3
; +49.8 : -49.8 : +50 for C=13

;for z-only gradients:
gpz1: +49.8%
gpz2: -49.8%
gpz3: +50%
which equals 1:-1:1.004 (Weigelt & Otting)

;use gradient files:
gpnam1: SMSQ10.100
gpnam2: SMSQ10.100
gpnam3: SMSQ10.100
```

A-17
DEPT2LR_LKJ pulse program

;dept2lr_lkj
;avance version
;proton detected 13C-DQ experiment for measurement of 2JCH;coupling constants,
with splitting in F1 as
;1J(C1H1) + 2J(C2H1)
;
based on dept2 proton detected inadequate by OWS,
;
by LKJ-CHG-OWS May 2014
; modified from version by AXME-OWS

#include <Avance.incl>
#include <Delay.incl>
#include <Grad.incl>

"d0=3u"
"d2=1s/(cnst2*2)"   ;1/2J(CH)
"d4=1s/(cnst3*4)"   ;(1/2J(CC))/2
"d11=30m"
"p2=p1*2"
"p4=p3*2"
"p8=p1*(cnst8/90)"   ;Theta1
"p9=p1*(cnst9/90)"   ;Theta2
"l3=(td1/2)"
"d24=(1s/(cnst3*4))-(1s/(cnst2*2))-p9"  ; d4-d2-p9
"d29=d2-50u-p16-d16" ; d2-50u-p16-d16
"d20=6u"
"in0=inf1/2"

1   ze
   d11 p112:f2
2   d1 do:f2    ; decoupler off during d1
3   d11
4   (p1 ph1)
   d2 p12:f2
   (center (p2 ph1) (p3 ph7):f2)
   d2
   (p8 ph3)   ;1H Theta 1 pulse
   d24
   (p4 ph7):f2
   d4    ;tau_c/2
   (p3 ph8):f2
   d0
   d0
50u UNBLKGRAD
   p16:gp1*EA
   d16
   (p4 ph10):f2
50u
   p16:gp2*EA
   d16
   d20    ; compensate initial d0
   ;(refocus 13C completely for t1=0)
   (p3 ph11):f2
   d4
   (p4 ph12):f2
   d24    ;d4-d2-p9
   (p9 ph5)   ;1H Theta 2 pulse
   d2
   (center (p2 ph6) (p3 ph12):f2)
50u
   p16:gp3
   d16 p112:f2
   d29 BLKGRAD ;d29=d2-(50u+p16+d16)
g0=2 ph31 cpd2:f2
Appendix

d1 do:f2 wr #0 if #0 zd ;decoupler off during d1
d11 igrad EA
lo to 3 times 2
d11 id0
lo to 4 times 13
exit

ph1 = 0
ph3 = 1
ph5 = 0
ph6 = 0
ph7 = 0 1 2 3 1 2 3 0 2 3 0 1 3 0 1 2
ph8 = 1 2 3 0 2 3 0 1 3 2 6 0 4 1 5 3 7 4 0 6 2 7 3 5 1
ph11 = 1
ph12 = 0
ph31 = 0 2

;pl1 : f1 channel - power level for pulse (default)
;pl2 : f2 channel - power level for pulse (default)
;pl12: f2 channel - power level for CPD/BB decoupling
;pl1 : f1 channel - 90 degree high power pulse
;pl2 : f1 channel - 180 degree high power pulse
;pl3 : f2 channel - 90 degree high power pulse
;pl4 : f2 channel - 180 degree high power pulse
;pl8: Theta1 1H pulse
;pl9: Theta2 1H pulse
;pl10: homospoil/gradient pulse
;d0 : incremented delay (2D) [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d2: 1/2J(CH) or 2*taul
;d4: 1/4J(CC) or Tau2
;d11: delay for disk I/O [30 msec]
;d16: delay for homospoil/gradient recovery
;d20: compensate initial d0 and p2
;d21: 1/(4J(CC))
;d24: d4-d2-p9
;d29: d2-50u-p16-d16
;cnst2 : J(CH) = 127 .. 160 Hz
;cnst3 : J(CC) = 35 .. 55 Hz
;cnst8: Theta1 (degrees)
;cnst9: Theta2 (degrees)
;inf1: 1/SW(C) = 2 * DW(C)
;in0: 1/(2 * SW(C)) = DW(C)
;nd0: 2
;l3: td1/2
;NS: 4 * n (optimal is 32, minimum is 4)
;DS: 4 * n
;td1: number of experiments
;FnMODE: echo-antiecho
;cpd2: decoupling according to sequence defined by cpdprg2
;pcpd2: f2 channel 90 degree pulse for decoupling sequence

;use gradient ratio: gp 1 : gp 2 : gp 3
; +49.8 : -49.8 : +50.1 for C=13

;for z-only gradients:
gpz1: +49.8%
gpz2: -49.8%
gpz3: +50.1%
which equals 1:-1:1.006

;use gradient files:
;gpnam1: SMSQ10.100
;gpnam2: SMSQ10.100
;gpnam3: SMSQ10.100
**Strychnine**

**Journal Number:** LKJ020067a  
**Amount:** 20 mg  
**Solvent:** CDCl$_3$

<table>
<thead>
<tr>
<th>Atom assignment</th>
<th>$^1$H chemical shift [ppm], multiplicity, $J$ [Hz]</th>
<th>$^{13}$C chemical shift [ppm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.20, 1H, d, 7.4</td>
<td>122.4</td>
</tr>
<tr>
<td>2</td>
<td>7.10, 1H, t, 7.4</td>
<td>124.4</td>
</tr>
<tr>
<td>3</td>
<td>7.26, 1H, m</td>
<td>128.8</td>
</tr>
<tr>
<td>4</td>
<td>8.09, 1H, d, 8.0</td>
<td>116.2</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>142.1</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>131.7</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>51.9</td>
</tr>
<tr>
<td>8</td>
<td>3.88, 1H, d, 10.5</td>
<td>59.9</td>
</tr>
<tr>
<td>9-N</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>169.2</td>
</tr>
<tr>
<td>11$\alpha$</td>
<td>3.13, 1H, dd, 17.5, 8.5</td>
<td>42.4</td>
</tr>
<tr>
<td>11$\beta$</td>
<td>2.67, 1H, dd, 17.4, 3.0</td>
<td>42.4</td>
</tr>
<tr>
<td>12</td>
<td>4.29, 1H, m</td>
<td>77.5</td>
</tr>
<tr>
<td>13</td>
<td>1.29, 1H, dt, 10.4, 2.8</td>
<td>48.0</td>
</tr>
<tr>
<td>14</td>
<td>3.17, 1H, bs</td>
<td>31.4</td>
</tr>
<tr>
<td>15$\alpha$</td>
<td>2.40, 1H, dt, 14.6, 4.1</td>
<td>26.6</td>
</tr>
<tr>
<td>15$\beta$</td>
<td>1.49, 1H, d, 14.4</td>
<td>26.6</td>
</tr>
<tr>
<td>16</td>
<td>4.07, 1H, m</td>
<td>60.1</td>
</tr>
<tr>
<td>17$\alpha$</td>
<td>1.94, 1H, m</td>
<td>42.5</td>
</tr>
<tr>
<td>17$\beta$</td>
<td>1.92, 1H, m</td>
<td>42.5</td>
</tr>
<tr>
<td>18$\alpha$</td>
<td>3.32, 1H, m</td>
<td>50.2</td>
</tr>
<tr>
<td>18$\beta$</td>
<td>2.91, 1H, m</td>
<td>50.2</td>
</tr>
<tr>
<td>19-N</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>20$\alpha$</td>
<td>3.78, 1H, d, 14.6</td>
<td>52.4</td>
</tr>
<tr>
<td>20$\beta$</td>
<td>2.81, 1H, d, 14.7</td>
<td>52.4</td>
</tr>
<tr>
<td>21</td>
<td>-</td>
<td>139.2</td>
</tr>
<tr>
<td>22</td>
<td>5.98, 1H, m</td>
<td>128.7</td>
</tr>
<tr>
<td>23$\alpha$</td>
<td>4.16, 1H, dd, 13.8, 7.0</td>
<td>64.5</td>
</tr>
<tr>
<td>23$\beta$</td>
<td>4.07, 1H, m</td>
<td>64.5</td>
</tr>
</tbody>
</table>
## Extended table of calculated $^{n+1}\textit{J}_{HH}$ coupling constants in strychnine

<table>
<thead>
<tr>
<th>#H1</th>
<th>#H2</th>
<th>Exp.</th>
<th>Ref.</th>
<th>6-31G(d,p)</th>
<th>6-31G(d,p)</th>
<th>6-31G(d,p)</th>
<th>6-311+G(d,p)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HMBC+</td>
<td>FC only</td>
<td>corr. FC only</td>
<td>Full</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>1.17</td>
<td>1.18</td>
<td>1.24</td>
<td>1.13</td>
<td>1.11</td>
<td>1.11</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>0.50</td>
<td>0.45</td>
<td>0.60</td>
<td>0.55</td>
<td>0.60</td>
<td>0.52</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>1.09</td>
<td>1.12</td>
<td>1.06</td>
<td>0.97</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>-0.19</td>
<td></td>
<td>-0.11</td>
<td>-0.10</td>
<td>-0.18</td>
<td>-0.10</td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>10.49</td>
<td>10.47</td>
<td>10.93</td>
<td>9.96</td>
<td>10.91</td>
<td>10.14</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>-0.70</td>
<td></td>
<td>-0.29</td>
<td>-0.27</td>
<td>-0.33</td>
<td>-0.33</td>
</tr>
<tr>
<td>11a</td>
<td>12</td>
<td>8.34</td>
<td>8.44</td>
<td>9.24</td>
<td>8.43</td>
<td>9.48</td>
<td>8.78</td>
</tr>
<tr>
<td>11a</td>
<td>13</td>
<td>-0.42</td>
<td></td>
<td>-0.34</td>
<td>-0.31</td>
<td>-0.38</td>
<td>-0.31</td>
</tr>
<tr>
<td>11b</td>
<td>12</td>
<td>3.24</td>
<td>3.28</td>
<td>4.12</td>
<td>3.75</td>
<td>4.00</td>
<td>3.58</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>3.20</td>
<td>3.12</td>
<td>3.92</td>
<td>3.58</td>
<td>4.21</td>
<td>3.82</td>
</tr>
<tr>
<td>12</td>
<td>15b</td>
<td>0.40</td>
<td></td>
<td>0.02</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>12</td>
<td>23a</td>
<td>-0.08</td>
<td></td>
<td>-0.02</td>
<td>-0.02</td>
<td>0.23</td>
<td>-0.11</td>
</tr>
<tr>
<td>13</td>
<td>14</td>
<td>2.97</td>
<td>3.10</td>
<td>3.59</td>
<td>3.27</td>
<td>3.76</td>
<td>3.44</td>
</tr>
<tr>
<td>13</td>
<td>15a</td>
<td>0.34</td>
<td></td>
<td>0.42</td>
<td>0.38</td>
<td>0.39</td>
<td>0.46</td>
</tr>
<tr>
<td>13</td>
<td>15b</td>
<td>-0.30</td>
<td></td>
<td>-0.31</td>
<td>-0.28</td>
<td>-0.04</td>
<td>-0.25</td>
</tr>
<tr>
<td>14</td>
<td>15a</td>
<td>5.42</td>
<td>4.58</td>
<td>5.05</td>
<td>4.60</td>
<td>5.15</td>
<td>4.77</td>
</tr>
<tr>
<td>14</td>
<td>15b</td>
<td>2.03</td>
<td>1.98</td>
<td>2.11</td>
<td>1.93</td>
<td>2.16</td>
<td>2.05</td>
</tr>
<tr>
<td>14</td>
<td>16</td>
<td>0.76</td>
<td></td>
<td>0.60</td>
<td>0.54</td>
<td>0.59</td>
<td>0.52</td>
</tr>
<tr>
<td>14</td>
<td>20b</td>
<td>-0.58</td>
<td></td>
<td>-0.55</td>
<td>-0.50</td>
<td>-0.64</td>
<td>-0.59</td>
</tr>
<tr>
<td>14</td>
<td>22</td>
<td>-2.89</td>
<td>2.82</td>
<td>-3.32</td>
<td>-3.03</td>
<td>-3.43</td>
<td>-3.00</td>
</tr>
<tr>
<td>15a</td>
<td>16</td>
<td>3.97</td>
<td>4.11</td>
<td>4.26</td>
<td>3.89</td>
<td>4.36</td>
<td>4.11</td>
</tr>
<tr>
<td>15b</td>
<td>16</td>
<td>2.29</td>
<td>1.82</td>
<td>2.21</td>
<td>2.02</td>
<td>2.32</td>
<td>2.20</td>
</tr>
<tr>
<td>16</td>
<td>17a</td>
<td></td>
<td></td>
<td>0.05</td>
<td>0.04</td>
<td>0.16</td>
<td>0.06</td>
</tr>
<tr>
<td>16</td>
<td>17b</td>
<td></td>
<td></td>
<td>-0.20</td>
<td>-0.18</td>
<td>-0.27</td>
<td>-0.15</td>
</tr>
<tr>
<td>16</td>
<td>17ab*</td>
<td>-0.14</td>
<td></td>
<td>-0.15</td>
<td>-0.14</td>
<td>-0.11</td>
<td>-0.10</td>
</tr>
<tr>
<td>16</td>
<td>20b</td>
<td>-0.35</td>
<td></td>
<td>-0.36</td>
<td>-0.33</td>
<td>-0.42</td>
<td>-0.35</td>
</tr>
<tr>
<td>20a</td>
<td>22</td>
<td>-1.37</td>
<td>1.47</td>
<td>-1.99</td>
<td>-1.82</td>
<td>-2.10</td>
<td>-1.83</td>
</tr>
<tr>
<td>20b</td>
<td>22</td>
<td>-0.63</td>
<td>0.63</td>
<td>-0.63</td>
<td>-0.58</td>
<td>-0.42</td>
<td>-0.59</td>
</tr>
<tr>
<td>22</td>
<td>23a</td>
<td>6.08</td>
<td>5.74</td>
<td>6.74</td>
<td>6.14</td>
<td>6.54</td>
<td>5.97</td>
</tr>
<tr>
<td>22</td>
<td>23b</td>
<td>7.00</td>
<td>6.88</td>
<td>7.53</td>
<td>6.86</td>
<td>7.77</td>
<td>7.38</td>
</tr>
</tbody>
</table>

**RMSD =** 0.40 0.28 0.47 0.29

All calculations are at the B3LYP level. Linear correction factor 0.9117. *17ab is assumed to be an average.


**HMBC+ spectra for strychnine and LNT**

(following pages)
HMBC+ CH/CH₃. The subtract subspectrum (red) has been shifted slightly upwards, to make it easier to see the splitting. When the red curve (subtract) is on the left side, $J$ is negative.
HMBC+ C/CH$_2$ (region with couplings to CH$_3$). The subtract subspectrum (red) has been shifted slightly upwards, to make it easier to see the splitting.
HMBC+ CH/CH₃ subspectrum for LNT. The subtract subspectrum (red) has been shifted slightly upwards, to make it easier to see the splitting.

H₃ GlcNAc – H₁ Gal2 = -0.3 Hz
Brazilin

Journal Number: LKJ020060a
Amount: 20 mg
Solvent: DMSO-$d_6$

<table>
<thead>
<tr>
<th>Atom assignment</th>
<th>$^1$H chemical shift [ppm], multiplicity, $J$ [Hz]</th>
<th>$^{13}$C chemical shift [ppm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.14, 1H, d, 8.3</td>
<td>131.0</td>
</tr>
<tr>
<td>1a</td>
<td>-</td>
<td>114.4</td>
</tr>
<tr>
<td>2</td>
<td>6.42, 1H, dd, 8.3, 2.4</td>
<td>108.7</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>156.5</td>
</tr>
<tr>
<td>4</td>
<td>6.21, 1H, d, 2.4</td>
<td>102.8</td>
</tr>
<tr>
<td>4a</td>
<td>-</td>
<td>154.1</td>
</tr>
<tr>
<td>6</td>
<td>3.85, 1H, m</td>
<td>69.6</td>
</tr>
<tr>
<td></td>
<td>3.57, 1H, d, 11.3</td>
<td></td>
</tr>
<tr>
<td>6a</td>
<td>-</td>
<td>76.3</td>
</tr>
<tr>
<td>7</td>
<td>2.88, 1H, d, 15.7</td>
<td>42.0</td>
</tr>
<tr>
<td></td>
<td>2.70, 1H, d, 15.7</td>
<td></td>
</tr>
<tr>
<td>7a</td>
<td>-</td>
<td>129.8</td>
</tr>
<tr>
<td>8</td>
<td>6.54, 1H, s</td>
<td>112.1</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>144.3</td>
</tr>
<tr>
<td>10</td>
<td>-</td>
<td>144.0</td>
</tr>
<tr>
<td>11</td>
<td>6.64, 1H, s</td>
<td>111.7</td>
</tr>
<tr>
<td>11a</td>
<td>-</td>
<td>135.6</td>
</tr>
<tr>
<td>12</td>
<td>3.85, 1H, m</td>
<td>49.5</td>
</tr>
</tbody>
</table>

Broad singlets are observed at 9.24, 8.61, 5.26 and 3.5 ppm.
Chapter 3

Pyranonigrin E

Journal Number: LKJ032404
Amount: 2.7 mg
Solvent: DMSO- $d_6$

![Pyranonigrin E structure](image)

<table>
<thead>
<tr>
<th>Atom assignment</th>
<th>$^1$H chemical shift [ppm], multiplicity, $J$ [Hz]</th>
<th>$^{13}$C chemical shift [ppm]</th>
<th>HMBC correlations</th>
<th>Key NOE correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>154.1</td>
<td>5, 7, 7a, (4a), (4)</td>
<td>7-OH</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>143.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>169.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4a</td>
<td>-</td>
<td>164.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>174.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>8.63, 1H, bs</td>
<td>-</td>
<td>5, 7, 7a, (4a), (4)</td>
<td>7-OH</td>
</tr>
<tr>
<td>7</td>
<td>5.73, 1H, dd, 9: 1.3</td>
<td>74.9</td>
<td>4a, 5, (7a)</td>
<td></td>
</tr>
<tr>
<td>7-OH</td>
<td>6.80, 1H, d, 9.1</td>
<td>-</td>
<td>5, 7</td>
<td>6</td>
</tr>
<tr>
<td>7a</td>
<td>-</td>
<td>113.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3.76, 3H, s</td>
<td>60.3</td>
<td>3</td>
<td>2’, 3’</td>
</tr>
<tr>
<td>1’</td>
<td>6.59, 1H, m</td>
<td>117.2</td>
<td>2, 2’, 3’</td>
<td></td>
</tr>
<tr>
<td>2’</td>
<td>6.59, 1H, m</td>
<td>139.7</td>
<td>2, 1’, 3’</td>
<td>8</td>
</tr>
<tr>
<td>3’</td>
<td>2.27, 2H, m</td>
<td>34.5</td>
<td>1’, 2’, 4’, 5’</td>
<td>8</td>
</tr>
<tr>
<td>4’</td>
<td>1.48, 2H, sextet, 7.2</td>
<td>21.4</td>
<td>2’, 3’, 5’</td>
<td></td>
</tr>
<tr>
<td>5’</td>
<td>0.91, 3H, t, 7.5</td>
<td>13.6</td>
<td>3’, 4’</td>
<td></td>
</tr>
</tbody>
</table>
**JBIR-74**

**Journal Number:** LKJ032403
**Amount:** 11.0 mg
**Solvent:** DMSO-$d_6$

![Chemical Structure]

<table>
<thead>
<tr>
<th>Atom assignment</th>
<th>$^1$H chemical shift [ppm], multiplicity, $J$ [Hz]</th>
<th>$^{13}$C chemical shift [ppm]</th>
<th>HMBC correlations</th>
<th>NOE correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>165.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4.03, 1H, bd, 2.4</td>
<td>58.2</td>
<td>1, 3, 4, 5, 1'</td>
<td></td>
</tr>
<tr>
<td>2'-NH</td>
<td>8.20, 1H, bs</td>
<td>-</td>
<td>1, 2'</td>
<td>3, 4</td>
</tr>
<tr>
<td>3</td>
<td>1.90, 1H, m</td>
<td>39.6</td>
<td>1, 2, 4, 5, 6</td>
<td>2-NH</td>
</tr>
<tr>
<td>4</td>
<td>1.42, 1H, m</td>
<td>24.4</td>
<td>2, 3, 5, 6</td>
<td>2-NH</td>
</tr>
<tr>
<td>5</td>
<td>1.22, 1H, m</td>
<td>13.5</td>
<td>2, 3, 4</td>
<td>3'</td>
</tr>
<tr>
<td>6</td>
<td>0.78, 3H, d, 6.7</td>
<td>11.6</td>
<td>3, 4</td>
<td></td>
</tr>
<tr>
<td>1'</td>
<td>-</td>
<td>159.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2'</td>
<td>-</td>
<td>124.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2'-NH</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3'</td>
<td>0.87, 3H, m</td>
<td>11.6</td>
<td>3, 4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6.51, 1H, s</td>
<td>103.2</td>
<td>1', 4', 5', (1)</td>
<td>5', 5</td>
</tr>
<tr>
<td>4'</td>
<td>-</td>
<td>136.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4'-NH</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>7.43, 1H, s</td>
<td>118.6</td>
<td>(2'), 3', 6'</td>
<td>3', 6'</td>
</tr>
<tr>
<td>6</td>
<td>7.89, 1H, s</td>
<td>137.8</td>
<td>4', 5'</td>
<td>5'</td>
</tr>
</tbody>
</table>
Decumbenone A

Journal Number: LKJ032608
Amount: 0.8 mg
Solvent: DMSO-$d_6$

All spectral data have been referenced to external 1,4-dioxane with $\delta_H$ 3.75 ppm and $\delta_C$ 67.4 ppm.

Chapter 4
3’-sialyllactose (3’-SL)

Journal Number: LKJ020001a
Amount:
Solvent: D$_2$O

<table>
<thead>
<tr>
<th>Atom assignment</th>
<th>$^1$H chemical shift [ppm], multiplicity, $J$ [Hz]</th>
<th>$^{13}$C chemical shift [ppm]</th>
<th>HMBC correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Glc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1α</td>
<td>5.21, 1H, d, 3.6</td>
<td>92.7</td>
<td>2α, 4α</td>
</tr>
<tr>
<td>2α</td>
<td>3.58, 1H, m</td>
<td>72.0</td>
<td></td>
</tr>
<tr>
<td>3α</td>
<td>3.82, 1H, m</td>
<td>72.2</td>
<td></td>
</tr>
<tr>
<td>4α</td>
<td>3.67, 1H, m</td>
<td>79.1</td>
<td></td>
</tr>
<tr>
<td>5α</td>
<td>3.95, 1H, m</td>
<td>70.9</td>
<td>3α</td>
</tr>
<tr>
<td>6α</td>
<td>3.87, 1H, m</td>
<td>60.7</td>
<td></td>
</tr>
<tr>
<td>6α</td>
<td>3.92, 1H, m</td>
<td>60.7</td>
<td></td>
</tr>
<tr>
<td>β-Glc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1β</td>
<td>4.65, 1H, d, 8.1</td>
<td>96.6</td>
<td></td>
</tr>
<tr>
<td>2β</td>
<td>3.27, 1H, dd, 8.1, 9.0</td>
<td>74.6</td>
<td>1β, 3β</td>
</tr>
<tr>
<td>3β</td>
<td>3.64, 1H, m</td>
<td>75.2</td>
<td>5β</td>
</tr>
<tr>
<td>4β</td>
<td>3.65, 1H, m</td>
<td>79.0</td>
<td></td>
</tr>
<tr>
<td>5β</td>
<td>3.59, 1H, m</td>
<td>75.6</td>
<td>4β</td>
</tr>
<tr>
<td>6β</td>
<td>3.82, 1H, m</td>
<td>60.9</td>
<td>5β</td>
</tr>
<tr>
<td>6β</td>
<td>3.97, 1H, m</td>
<td>60.9</td>
<td></td>
</tr>
<tr>
<td>β-Gal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1’</td>
<td>4.52, 1H, d, 7.9</td>
<td>103.5</td>
<td>5α/5β</td>
</tr>
<tr>
<td>2’</td>
<td>3.57, 1H, m</td>
<td>70.2</td>
<td>1’, 3’</td>
</tr>
<tr>
<td>3’</td>
<td>4.10, 1H, bd, 9.8</td>
<td>76.3</td>
<td>2”, 2’</td>
</tr>
<tr>
<td>4’</td>
<td>3.95, 1H, m</td>
<td>68.3</td>
<td>2’, 3’</td>
</tr>
<tr>
<td>5’</td>
<td>3.70, 1H, m</td>
<td>76.0</td>
<td>6’</td>
</tr>
<tr>
<td>6’</td>
<td>3.71, 1H, m</td>
<td>61.9</td>
<td></td>
</tr>
<tr>
<td>6’</td>
<td>3.76, 1H, m</td>
<td>61.9</td>
<td>5’</td>
</tr>
<tr>
<td>α-NAcNeu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1”</td>
<td>-</td>
<td>174.7</td>
<td>-</td>
</tr>
<tr>
<td>2”</td>
<td>-</td>
<td>100.6</td>
<td>-</td>
</tr>
<tr>
<td>3”</td>
<td>1.79, 1H, dd, 12.5, 11.7</td>
<td>40.5</td>
<td>1”, 2”, 4”, 5”</td>
</tr>
<tr>
<td>3”</td>
<td>2.75, 1H, dd, 12.5, 4.5</td>
<td>40.5</td>
<td>2”, 4”, 5”</td>
</tr>
<tr>
<td>4”</td>
<td>3.68, 1H, m</td>
<td>69.2</td>
<td></td>
</tr>
<tr>
<td>5”</td>
<td>3.84, 1H, m</td>
<td>52.5</td>
<td>CO, 4”, 6”</td>
</tr>
<tr>
<td>CO</td>
<td>-</td>
<td>175.8</td>
<td>-</td>
</tr>
<tr>
<td>CH$_3$</td>
<td>2.02, 3H, s</td>
<td>22.9</td>
<td>CO</td>
</tr>
<tr>
<td>6”</td>
<td>3.62, 1H, m</td>
<td>73.7</td>
<td>2”, 4”, 8”</td>
</tr>
<tr>
<td>7”</td>
<td>3.59, 1H, m</td>
<td>68.9</td>
<td>8”, 9”</td>
</tr>
<tr>
<td>8”</td>
<td>3.88, 1H, m</td>
<td>72.6</td>
<td>7”</td>
</tr>
<tr>
<td>9”</td>
<td>3.64, 1H, m</td>
<td>63.4</td>
<td></td>
</tr>
<tr>
<td>9”</td>
<td>3.86, 1H, m</td>
<td>63.4</td>
<td></td>
</tr>
</tbody>
</table>
6’-sialyllactose (6’-SL)

Journal Number: LKJ020003a
Amount: 6.0 mg
Solvent: D₂O

<table>
<thead>
<tr>
<th>Atom assignment</th>
<th>¹H chemical shift [ppm], multiplicity, J [Hz]</th>
<th>¹³C chemical shift [ppm]</th>
<th>HMBC correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Glc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1α</td>
<td>5.20, 1H, d, 3.7</td>
<td>92.2</td>
<td>3α, 4α</td>
</tr>
<tr>
<td>2α</td>
<td>3.58, 1H, m</td>
<td>71.4</td>
<td>3α</td>
</tr>
<tr>
<td>3α</td>
<td>3.82, 1H, m</td>
<td>72.0</td>
<td>2α, 5α</td>
</tr>
<tr>
<td>4α</td>
<td>3.60, 1H, m</td>
<td>80.1</td>
<td>3α, 5α, 6α</td>
</tr>
<tr>
<td>5α</td>
<td>3.93, 1H, m</td>
<td>70.3</td>
<td>4α</td>
</tr>
<tr>
<td>6α</td>
<td>3.83, 1H, m</td>
<td>60.5</td>
<td></td>
</tr>
<tr>
<td>6α</td>
<td>3.86, 1H, m</td>
<td>60.5</td>
<td>5α</td>
</tr>
<tr>
<td>β-Glc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1β</td>
<td>4.65, 1H, d, 8.1</td>
<td>96.0</td>
<td>2β, 3β/4β</td>
</tr>
<tr>
<td>2β</td>
<td>3.29, 1H, m</td>
<td>74.1</td>
<td>1β, 3β/4β</td>
</tr>
<tr>
<td>3β</td>
<td>3.63, 1H, m</td>
<td>75.0</td>
<td>2β, 5β</td>
</tr>
<tr>
<td>4β</td>
<td>3.61, 1H, m</td>
<td>80.0</td>
<td>1’, 3β/5β</td>
</tr>
<tr>
<td>5β</td>
<td>3.59, 1H, m</td>
<td>75.0</td>
<td>4β</td>
</tr>
<tr>
<td>6β</td>
<td>3.78, 1H, m</td>
<td>60.6</td>
<td>4β, 5β</td>
</tr>
<tr>
<td>6β</td>
<td>3.93, 1H, m</td>
<td>60.6</td>
<td>4β, 5β</td>
</tr>
<tr>
<td>β-Gal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1’</td>
<td>4.41, 1H, d, 7.8</td>
<td>103.6</td>
<td>5α/5β, 3’, 5’</td>
</tr>
<tr>
<td>2’</td>
<td>3.51, 1H, m</td>
<td>71.2</td>
<td>1’, 3’</td>
</tr>
<tr>
<td>3’</td>
<td>3.64, 1H, m</td>
<td>72.7</td>
<td>2’, 4’</td>
</tr>
<tr>
<td>4’</td>
<td>3.91, 1H, m</td>
<td>68.9</td>
<td>2’, 3’, 5’, 6’</td>
</tr>
<tr>
<td>5’</td>
<td>3.79, 1H, m</td>
<td>74.1</td>
<td>1’, 4’, 6’</td>
</tr>
<tr>
<td>6’</td>
<td>3.58, 1H, m</td>
<td>64.0</td>
<td>2”, 4’</td>
</tr>
<tr>
<td>6’</td>
<td>3.96, 1H, m</td>
<td>64.0</td>
<td>2”, 4’, 5’</td>
</tr>
<tr>
<td>α-NAcNeu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1”</td>
<td>-</td>
<td>173.9</td>
<td>-</td>
</tr>
<tr>
<td>2”</td>
<td>-</td>
<td>100.6</td>
<td>-</td>
</tr>
<tr>
<td>3”</td>
<td>1.72, 1H, dd, 12.5, 11.7</td>
<td>40.5</td>
<td>1”, 2”, 4”, 5”</td>
</tr>
<tr>
<td>3”</td>
<td>2.69, 1H, dd, 12.5, 4.7</td>
<td>40.5</td>
<td>2”, 4”, 5”</td>
</tr>
<tr>
<td>4”</td>
<td>3.63, 1H, m</td>
<td>68.7</td>
<td>3”, 5”</td>
</tr>
<tr>
<td>5”</td>
<td>3.84, 1H, m</td>
<td>52.1</td>
<td>3”, 4”, CO, 6”</td>
</tr>
<tr>
<td>CO</td>
<td>-</td>
<td>175.3</td>
<td>-</td>
</tr>
<tr>
<td>CH₃</td>
<td>2.01, 3H, s</td>
<td>22.4</td>
<td>CO, 5”, 6”</td>
</tr>
<tr>
<td>6”</td>
<td>3.70, 1H, m</td>
<td>72.9</td>
<td>2”, 5”, 7”</td>
</tr>
<tr>
<td>7”</td>
<td>3.54, 1H, m</td>
<td>68.7</td>
<td>6”, 8”, 9”</td>
</tr>
<tr>
<td>8”</td>
<td>3.87, 1H, m</td>
<td>72.2</td>
<td>6”, 7”, 9”</td>
</tr>
</tbody>
</table>
### 2'-fucosyllactose (2'-FL)

**Journal Number:** LKJ020006a  
**Amount:** 2.0 mg  
**Solvent:** D$_2$O

<table>
<thead>
<tr>
<th>Atom assignment</th>
<th>$^1$H chemical shift [ppm], multiplicity, $J$ [Hz]</th>
<th>$^{13}$C chemical shift [ppm]</th>
<th>HMBC correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Glc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1α</td>
<td>5.21, 1H, d, 3.7</td>
<td>92.6</td>
<td>2α</td>
</tr>
<tr>
<td>2α</td>
<td>3.57, 1H, m</td>
<td>72.1</td>
<td></td>
</tr>
<tr>
<td>3α</td>
<td>3.78, 1H, m</td>
<td>76.1</td>
<td>5α</td>
</tr>
<tr>
<td>4α</td>
<td>3.71, 1H, m</td>
<td>76.7</td>
<td>3α, 6α</td>
</tr>
<tr>
<td>5α</td>
<td>3.89, 1H, m</td>
<td>71.2</td>
<td></td>
</tr>
<tr>
<td>6α</td>
<td>3.79, 1H, m</td>
<td>60.9</td>
<td>5α</td>
</tr>
<tr>
<td>6α</td>
<td>3.89, 1H, m</td>
<td>60.9</td>
<td></td>
</tr>
<tr>
<td>β-Glc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1β</td>
<td>4.62, 1H, d, 8.0</td>
<td>96.6</td>
<td></td>
</tr>
<tr>
<td>2β</td>
<td>3.28, 1H, dd, 9.4, 8.0</td>
<td>74.7</td>
<td>1β, 3β</td>
</tr>
<tr>
<td>3β</td>
<td>3.57, 1H, m</td>
<td>75.1</td>
<td>1β, 2β, 4β</td>
</tr>
<tr>
<td>4β</td>
<td>3.71, 1H, m</td>
<td>76.6</td>
<td>3β, 5β</td>
</tr>
<tr>
<td>5β</td>
<td>3.46, 1H, ddd, 9.8, 5.5, 1.7</td>
<td>76.1</td>
<td>1β, 3β</td>
</tr>
<tr>
<td>6β</td>
<td>3.76, 1H, m</td>
<td>61.0</td>
<td></td>
</tr>
<tr>
<td>6β</td>
<td>3.94, 1H, dd, 12.0, 1.7</td>
<td>61.0</td>
<td></td>
</tr>
<tr>
<td>β-Gal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1''</td>
<td>4.51, 1H, d, 7.7</td>
<td>101.0</td>
<td>5''</td>
</tr>
<tr>
<td>2''</td>
<td>3.66, 1H, dd, 9.3, 8.0</td>
<td>77.0</td>
<td>3'', 6''</td>
</tr>
<tr>
<td>3''</td>
<td>3.86, 1H, m</td>
<td>74.4</td>
<td>2''</td>
</tr>
<tr>
<td>4''</td>
<td>3.88, 1H, m</td>
<td>69.9</td>
<td>2'', 3'', 5''</td>
</tr>
<tr>
<td>5''</td>
<td>3.70, 1H, m</td>
<td>76.3</td>
<td></td>
</tr>
<tr>
<td>6''</td>
<td>3.73, 1H, m</td>
<td>61.9</td>
<td></td>
</tr>
<tr>
<td>6''</td>
<td>3.80, 1H, m</td>
<td>61.9</td>
<td>5''</td>
</tr>
<tr>
<td>α-Fuc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1'''</td>
<td>5.29, 1H, m</td>
<td>100.1</td>
<td>2'', 2''', 3''', 5''</td>
</tr>
<tr>
<td>2'''</td>
<td>3.78, 1H, m</td>
<td>68.9</td>
<td>3'''</td>
</tr>
<tr>
<td>3'''</td>
<td>3.77, 1H, m</td>
<td>70.4</td>
<td>2'''</td>
</tr>
<tr>
<td>4'''</td>
<td>3.80, 1H, m</td>
<td>72.4</td>
<td>2''', 3'''</td>
</tr>
<tr>
<td>5'''</td>
<td>4.22, 1H, m</td>
<td>67.7</td>
<td>4''', 6'''</td>
</tr>
<tr>
<td>6'''</td>
<td>1.21, 3H, bd, 6.6</td>
<td>16.1</td>
<td>4''', 5'''</td>
</tr>
</tbody>
</table>
3-fucosyllactose (3-FL)

Journal Number: LKJ020005a
Amount: 2.0 mg
Solvent: D$_2$O

![Chemical Structure of 3-fucosyllactose (3-FL)](image)

<table>
<thead>
<tr>
<th>Atom assignment</th>
<th>$^1$H chemical shift [ppm], multiplicity, $J$ [Hz]</th>
<th>$^{13}$C chemical shift [ppm]</th>
<th>HMBC correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$-Glc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1$\alpha$</td>
<td>5.16, 1H, d, 3.7</td>
<td>92.9</td>
<td>3$\alpha$, 4$\alpha$</td>
</tr>
<tr>
<td>2$\alpha$</td>
<td>3.75, 1H, m</td>
<td>73.4</td>
<td></td>
</tr>
<tr>
<td>3$\alpha$</td>
<td>3.93, 1H, m</td>
<td>75.5</td>
<td>1$''$, 4$\alpha$, 5$\alpha$</td>
</tr>
<tr>
<td>4$\alpha$</td>
<td>3.85, 1H, m</td>
<td>73.48</td>
<td>1$'$, 2$\alpha$, 5$\alpha$, 6$\alpha$</td>
</tr>
<tr>
<td>5$\alpha$</td>
<td>3.95, 1H, m</td>
<td>71.7</td>
<td>4$\alpha$</td>
</tr>
<tr>
<td>6$\alpha$</td>
<td>3.86, 1H, m</td>
<td>60.5</td>
<td></td>
</tr>
<tr>
<td>6$\alpha$</td>
<td>3.89, 1H, m</td>
<td>60.5</td>
<td></td>
</tr>
<tr>
<td>$\beta$-Glc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1$\beta$</td>
<td>4.63, 1H, d, 8.0</td>
<td>96.6</td>
<td>3$\beta$, 5$\beta$</td>
</tr>
<tr>
<td>2$\beta$</td>
<td>3.45, 1H, m</td>
<td>76.3</td>
<td>1$\beta$, 3$\beta$</td>
</tr>
<tr>
<td>3$\beta$</td>
<td>3.76, 1H, m</td>
<td>77.8</td>
<td>1$''$, 2$\beta$, 4$\beta$</td>
</tr>
<tr>
<td>4$\beta$</td>
<td>3.85, 1H, m</td>
<td>73.50</td>
<td>1$'$, 3$\beta$, 5$\beta$, 6$\beta$</td>
</tr>
<tr>
<td>5$\beta$</td>
<td>3.57, 1H, m</td>
<td>76.2</td>
<td>1$\beta$, 3$\beta$</td>
</tr>
<tr>
<td>6$\beta$</td>
<td>3.80, 1H, m</td>
<td>60.6</td>
<td></td>
</tr>
<tr>
<td>6$\beta$</td>
<td>3.95, 1H, m</td>
<td>60.6</td>
<td>4$\beta$, 5$\beta$</td>
</tr>
<tr>
<td>$\beta$-Gal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1$'$</td>
<td>4.41, 1H, d, 7.8</td>
<td>102.6</td>
<td>4$\alpha$, 4$\beta$, 5$'$</td>
</tr>
<tr>
<td>2$'$</td>
<td>3.47, 1H, m</td>
<td>72.0</td>
<td>1$'$, 3$'$</td>
</tr>
<tr>
<td>3$'$</td>
<td>3.63, 1H, ddd, 9.9, 3.2, 1.5</td>
<td>73.2</td>
<td>1$'$, 2$'$, 4$'$</td>
</tr>
<tr>
<td>4$'$</td>
<td>3.89, 1H, m</td>
<td>69.1</td>
<td>2$'$, 3$'$, 5$'$</td>
</tr>
<tr>
<td>5$'$</td>
<td>3.58, 1H, m</td>
<td>75.8</td>
<td>1$'$, 3$'$, 4$'$, 6$'$</td>
</tr>
<tr>
<td>6$'$</td>
<td>3.71, 1H, m</td>
<td>62.3</td>
<td>5$'$</td>
</tr>
<tr>
<td>6$'$</td>
<td>3.74, 1H, m</td>
<td>62.3</td>
<td>4$'$, 5$'$</td>
</tr>
<tr>
<td>$\alpha$-Fuc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1$''$</td>
<td>5.36+5.42, 1H, 2*d, 4.0</td>
<td>99.2+99.3</td>
<td>3$\beta$, 2$''$, 3$''$, 5$''$</td>
</tr>
<tr>
<td>2$''$</td>
<td>3.77-3.78, 1H, m</td>
<td>68.8+68.9</td>
<td>3$''$</td>
</tr>
<tr>
<td>3$''$</td>
<td>3.95, 1H, m</td>
<td>70.0+70.1</td>
<td>2$''$, 4$''$</td>
</tr>
<tr>
<td>4$''$</td>
<td>3.79, 1H, m</td>
<td>72.8</td>
<td>2$''$, 3$''$, 5$''$, 6$''$</td>
</tr>
<tr>
<td>5$''$</td>
<td>4.81, 1H, dq, 6.6</td>
<td>67.3</td>
<td>1$''$, 3$''$, 4$''$, 6$''$</td>
</tr>
<tr>
<td>6$''$</td>
<td>1.17, 3H, dd, 6.6, 2.6</td>
<td>16.0</td>
<td>4$''$, 5$''$</td>
</tr>
</tbody>
</table>
Lacto-\(N\)-tetraose (LNT)

Journal Number: LKJ020004a  
Amount: 20 mg  
Solvent: \(D_2O\)

![Lacto-\(N\)-tetraose (LNT) structure](image)

<table>
<thead>
<tr>
<th>Atom assignment</th>
<th>(^1H) chemical shift [ppm], multiplicity, (J) [Hz]</th>
<th>(^13C) chemical shift [ppm]</th>
<th>HMBC correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha)-Glc</td>
<td>(5.20, 1H, d, 3.8)</td>
<td>92.7</td>
<td>(2\alpha, 3\alpha, 4\alpha)</td>
</tr>
<tr>
<td>1(\alpha)</td>
<td>3.56, 1H, m</td>
<td>72.0</td>
<td>3(\alpha)</td>
</tr>
<tr>
<td>2(\alpha)</td>
<td>3.81, 1H, m</td>
<td>72.3</td>
<td>2(\alpha, 5\alpha)</td>
</tr>
<tr>
<td>3(\alpha)</td>
<td>3.62, 1H, m</td>
<td>79.2</td>
<td>3(\alpha, 5\alpha)</td>
</tr>
<tr>
<td>4(\alpha)</td>
<td>3.93, 1H, m</td>
<td>71.0</td>
<td></td>
</tr>
<tr>
<td>5(\alpha)</td>
<td>3.81, 1H, m</td>
<td>70.8</td>
<td>5(\alpha)</td>
</tr>
<tr>
<td>6(\alpha)</td>
<td>3.86, 1H, m</td>
<td>60.0</td>
<td>5(\alpha)</td>
</tr>
<tr>
<td>(\beta)-Glc</td>
<td>4.64, 1H, d, 7.9</td>
<td>96.9</td>
<td>2(\beta, 3\beta, 4\beta)</td>
</tr>
<tr>
<td>1(\beta)</td>
<td>3.26, 1H, m</td>
<td>74.6</td>
<td>2(\beta, 3\beta)</td>
</tr>
<tr>
<td>2(\beta)</td>
<td>3.62, 1H, m</td>
<td>75.2</td>
<td>5(\beta)</td>
</tr>
<tr>
<td>3(\beta)</td>
<td>3.63, 1H, m</td>
<td>79.2</td>
<td></td>
</tr>
<tr>
<td>4(\beta)</td>
<td>3.58, 1H, m</td>
<td>75.7</td>
<td>4(\beta)</td>
</tr>
<tr>
<td>5(\beta)</td>
<td>3.77, 1H, m</td>
<td>60.9</td>
<td></td>
</tr>
<tr>
<td>6(\beta)</td>
<td>3.93, 1H, m</td>
<td>60.9</td>
<td>4(\beta, 5\beta)</td>
</tr>
<tr>
<td>(\beta)-Gal 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1(\prime)</td>
<td>4.42, 1H, d, 7.8</td>
<td>103.8</td>
<td>3(\prime, 4\alpha, 5\alpha, 5\beta)</td>
</tr>
<tr>
<td>2(\prime)</td>
<td>3.57, 1H, m</td>
<td>70.9</td>
<td>3(\prime)</td>
</tr>
<tr>
<td>3(\prime)</td>
<td>3.72, 1H, m</td>
<td>82.8</td>
<td>2(\prime, 4\prime)</td>
</tr>
<tr>
<td>4(\prime)</td>
<td>4.13, 1H, d, 3.1</td>
<td>69.2</td>
<td>2(\prime, 3\prime, 5\prime, 6\prime)</td>
</tr>
<tr>
<td>5(\prime)</td>
<td>3.69, 1H, m</td>
<td>75.8</td>
<td>4(\prime)</td>
</tr>
<tr>
<td>6(\prime)</td>
<td>3.73, 1H, m</td>
<td>61.8</td>
<td>5(\prime)</td>
</tr>
<tr>
<td>6(\prime)</td>
<td>3.76, 1H, m</td>
<td>61.8</td>
<td>5(\prime)</td>
</tr>
<tr>
<td>(\beta)-GlcNAc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1(\ast)</td>
<td>4.71, 1H, m</td>
<td>103.4</td>
<td>2(\ast, 3\ast, 4\ast)</td>
</tr>
<tr>
<td>2(\ast)</td>
<td>3.88, 1H, m</td>
<td>55.6</td>
<td>1(\ast, 3\ast, CO)</td>
</tr>
<tr>
<td>CO</td>
<td>-</td>
<td>175.8</td>
<td>-</td>
</tr>
<tr>
<td>CH(3)</td>
<td>2.01, 3H, s</td>
<td>23.1</td>
<td>CO, 1(\ast, 2\ast, 3\ast)</td>
</tr>
<tr>
<td>3(\ast)</td>
<td>3.79, 1H, m</td>
<td>82.9</td>
<td>2(\ast, 4\ast, 5\ast)</td>
</tr>
<tr>
<td>4(\ast)</td>
<td>3.55, 1H, m</td>
<td>69.3</td>
<td>2(\ast, 3\ast, 5\ast, 6\ast)</td>
</tr>
<tr>
<td>5(\ast)</td>
<td>3.46, 1H, ddd, 9.2, 5, 2</td>
<td>76.0</td>
<td>3(\ast, 4\ast, 6\ast)</td>
</tr>
<tr>
<td>6(\ast)</td>
<td>3.76, 1H, m</td>
<td>61.3</td>
<td>4(\ast, 5\ast)</td>
</tr>
<tr>
<td>6(\ast)</td>
<td>3.88, 1H, m</td>
<td>61.3</td>
<td>5(\ast)</td>
</tr>
<tr>
<td>(\beta)-Gal 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1(\prime\prime)</td>
<td>4.42, 1H, d, 7.8</td>
<td>104.4</td>
<td>2(\prime\prime, 3\prime\prime, 3\ast)</td>
</tr>
<tr>
<td>2(\prime\prime)</td>
<td>3.51, 1H, m</td>
<td>71.5</td>
<td>1(\prime\prime, 3\prime\prime)</td>
</tr>
<tr>
<td>3(\prime\prime)</td>
<td>3.62, 1H, ddd, 9.9, 3.2, 1.5</td>
<td>73.3</td>
<td>2(\prime\prime, 4\prime\prime)</td>
</tr>
<tr>
<td>4(\prime\prime)</td>
<td>3.89, 1H, m</td>
<td>69.4</td>
<td>2(\prime\prime, 3\prime\prime)</td>
</tr>
<tr>
<td>5(\prime\prime)</td>
<td>3.69, 1H, m</td>
<td>76.1</td>
<td>1(\prime\prime, 4\prime\prime, 6\prime\prime)</td>
</tr>
</tbody>
</table>
Lacto-N-neotetraose (LNnT)

Journal Number: LKJ020026  
Amount: 10 mg  
Solvent: D$_2$O

<table>
<thead>
<tr>
<th>Atom assignment</th>
<th>$^1$H chemical shift [ppm], multiplicity, $J$ [Hz]</th>
<th>$^{13}$C chemical shift [ppm]</th>
<th>HMBC correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Glc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1α</td>
<td>5.21, 1H, d, 3.8</td>
<td>92.6</td>
<td>3α, 5α</td>
</tr>
<tr>
<td>2α</td>
<td>3.56, 1H, m</td>
<td>72.0</td>
<td></td>
</tr>
<tr>
<td>3α</td>
<td>3.81, 1H, m</td>
<td>72.2</td>
<td>4α</td>
</tr>
<tr>
<td>4α</td>
<td>3.60, 1H, m</td>
<td>79.2</td>
<td></td>
</tr>
<tr>
<td>5α</td>
<td>3.93, 1H, m</td>
<td>70.9</td>
<td>4α</td>
</tr>
<tr>
<td>6α</td>
<td>3.88, 1H, m</td>
<td>60.8</td>
<td></td>
</tr>
<tr>
<td>6α</td>
<td>3.82, 1H, m</td>
<td>60.8</td>
<td></td>
</tr>
<tr>
<td>β-Glc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1β</td>
<td>4.65, 1H, d, 8.0</td>
<td>96.6</td>
<td>2β, 5β</td>
</tr>
<tr>
<td>2β</td>
<td>3.26, 1H, m</td>
<td>74.6</td>
<td>1β, 3β</td>
</tr>
<tr>
<td>3β</td>
<td>3.63, 1H, m</td>
<td>75.2</td>
<td>1’</td>
</tr>
<tr>
<td>4β</td>
<td>3.67, 1H, m</td>
<td>79.1</td>
<td></td>
</tr>
<tr>
<td>5β</td>
<td>3.58, 1H, m</td>
<td>75.6</td>
<td></td>
</tr>
<tr>
<td>6β</td>
<td>3.80, 1H, m</td>
<td>60.9</td>
<td></td>
</tr>
<tr>
<td>6β</td>
<td>3.91, 1H, m</td>
<td>60.9</td>
<td></td>
</tr>
<tr>
<td>β-Gal 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1’</td>
<td>4.42, 1H, d, 7.9</td>
<td>103.75</td>
<td>4α/β, 2’, 3’</td>
</tr>
<tr>
<td>2’</td>
<td>3.58, 1H, m</td>
<td>70.8</td>
<td>1’, 3’, 6’</td>
</tr>
<tr>
<td>3’</td>
<td>3.71, 1H, m</td>
<td>82.85</td>
<td></td>
</tr>
<tr>
<td>4’</td>
<td>4.14, 1H, d, 3.2</td>
<td>69.2</td>
<td>2’, 3’, (5’), (6’)</td>
</tr>
<tr>
<td>5’</td>
<td>3.70, 1H, m</td>
<td>75.7</td>
<td></td>
</tr>
<tr>
<td>6’</td>
<td>3.7-3.8, 1H, m</td>
<td>61.8</td>
<td></td>
</tr>
<tr>
<td>6’</td>
<td>3.7-3.8, 1H, m</td>
<td>61.8</td>
<td></td>
</tr>
<tr>
<td>β-GlcNAc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>4.69, 1H, d, 8.2</td>
<td>103.57</td>
<td>3’, 2*, 3*, 5*</td>
</tr>
<tr>
<td>2*</td>
<td>3.79, 1H, m</td>
<td>56.0</td>
<td>CO, 1*</td>
</tr>
<tr>
<td>CO</td>
<td>-</td>
<td>175.7</td>
<td></td>
</tr>
<tr>
<td>CH$_3$</td>
<td>2.02, 3H, s</td>
<td>23.0</td>
<td>CO, 2*</td>
</tr>
<tr>
<td>3*</td>
<td>3.72, 1H, m</td>
<td>73.0</td>
<td>2*</td>
</tr>
<tr>
<td>4*</td>
<td>3.71, 1H, m</td>
<td>79.0</td>
<td>1*, 2*</td>
</tr>
<tr>
<td>5*</td>
<td>3.58, 1H, m</td>
<td>75.4</td>
<td></td>
</tr>
<tr>
<td>6*</td>
<td>3.85, 1H, m</td>
<td>60.7</td>
<td>4*</td>
</tr>
<tr>
<td>6*</td>
<td>3.95, 1H, m</td>
<td>60.7</td>
<td>4*, 5*</td>
</tr>
<tr>
<td>β-Gal 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1”</td>
<td>4.47, 1H, d, 7.8</td>
<td>103.68</td>
<td>4*, 2”, 3”, 5”</td>
</tr>
<tr>
<td>2”</td>
<td>3.52, 1H, m</td>
<td>71.8</td>
<td>1”, 3”</td>
</tr>
</tbody>
</table>
Lacto-N-fucopentaose V (LNFP-V)

Journal Number: LKJ020032a
Amount: 12.5 mg
Solvent: D₂O

<table>
<thead>
<tr>
<th>Atom assignment</th>
<th>¹H chemical shift [ppm], multiplicity, J [Hz]</th>
<th>¹³C chemical shift [ppm]</th>
<th>HMBC correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Glc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1α</td>
<td>5.17, 1H, d, 3.8</td>
<td>92.9</td>
<td>3α, 5α</td>
</tr>
<tr>
<td>2α</td>
<td>3.74, 1H, m</td>
<td>73.5</td>
<td></td>
</tr>
<tr>
<td>3α</td>
<td>3.92, 1H, m</td>
<td>75.5</td>
<td>1°</td>
</tr>
<tr>
<td>4α</td>
<td>3.85, 1H, m</td>
<td>73.2</td>
<td>5α, 1’</td>
</tr>
<tr>
<td>5α</td>
<td>3.95, 1H, m</td>
<td>71.8</td>
<td></td>
</tr>
<tr>
<td>6α</td>
<td>3.8-3.94, 1H, m</td>
<td>60.5</td>
<td></td>
</tr>
<tr>
<td>6α</td>
<td>3.8-3.94, 1H, m</td>
<td>60.5</td>
<td></td>
</tr>
<tr>
<td>β-Glc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1β</td>
<td>4.64, 1H, d, 8.0</td>
<td>96.6</td>
<td>2β, 3β, 5β</td>
</tr>
<tr>
<td>2β</td>
<td>3.45, 1H, m</td>
<td>76.3</td>
<td>1β, 3β</td>
</tr>
<tr>
<td>3β</td>
<td>3.75, 1H, m</td>
<td>77.8</td>
<td>4β, 5β, 1°</td>
</tr>
<tr>
<td>4β</td>
<td>3.85, 1H, m</td>
<td>73.2</td>
<td>3β, 1’</td>
</tr>
<tr>
<td>5β</td>
<td>3.57, 1H, m</td>
<td>76.2</td>
<td></td>
</tr>
<tr>
<td>6β</td>
<td>~3.95, 1H, m</td>
<td>60.6</td>
<td></td>
</tr>
<tr>
<td>6β</td>
<td>~3.80, 1H, m</td>
<td>60.6</td>
<td></td>
</tr>
<tr>
<td>α-L-Fuc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1°</td>
<td>5.36-5.42, 1H, d, 4.0</td>
<td>99.2-99.4</td>
<td>3°, 5°, 3α, 3β</td>
</tr>
<tr>
<td>2°</td>
<td>3.78, 1H, m</td>
<td>68.8</td>
<td></td>
</tr>
<tr>
<td>3°</td>
<td>3.94, 1H, m</td>
<td>70.0</td>
<td>2°</td>
</tr>
<tr>
<td>4°</td>
<td>3.77, 1H, m</td>
<td>72.7</td>
<td></td>
</tr>
<tr>
<td>5°</td>
<td>4.80, 1H, m</td>
<td>67.3</td>
<td>1°, (3°), 4°, 6°</td>
</tr>
<tr>
<td>6°</td>
<td>1.15, 3H, d, 6.7</td>
<td>16.0</td>
<td>4°, 5°</td>
</tr>
<tr>
<td>β-Gal 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1’</td>
<td>4.41, 1H, d, 7.9</td>
<td>102.5</td>
<td>3’, 5’, 4α β</td>
</tr>
<tr>
<td>2’</td>
<td>3.50, 1H, m</td>
<td>71.5</td>
<td>1’, 3’</td>
</tr>
<tr>
<td>3’</td>
<td>3.71, 1H, m</td>
<td>82.3</td>
<td>2’, 1°</td>
</tr>
<tr>
<td>4’</td>
<td>4.08, 1H, d, 3.3</td>
<td>69.0</td>
<td>2’, 3’, 5’, 6’</td>
</tr>
</tbody>
</table>
Lacto-N-neofucopentaose V (LNnFP-V)

**Journal Number:** LKJ020031a

**Amount:** 11.5 mg

**Solvent:** D$_2$O

<table>
<thead>
<tr>
<th>Atom assignment</th>
<th>$^1$H chemical shift [ppm], multiplicity, $J$ [Hz]</th>
<th>$^{13}$C chemical shift [ppm]</th>
<th>HMBC correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$-Glc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1$\alpha$</td>
<td>5.17, 1H, d, 3.8</td>
<td>93.0</td>
<td>3$\alpha$, 5$\alpha$</td>
</tr>
<tr>
<td>2$\alpha$</td>
<td>3.74, 1H, m</td>
<td>73.4</td>
<td>3$\alpha$</td>
</tr>
<tr>
<td>3$\alpha$</td>
<td>3.93, 1H, m</td>
<td>75.5</td>
<td></td>
</tr>
<tr>
<td>4$\alpha$</td>
<td>3.85, 1H, m</td>
<td>73.2</td>
<td>3$\alpha$, 5$\alpha$, 6$\alpha$, 1$'$</td>
</tr>
<tr>
<td>5$\alpha$</td>
<td>3.95, 1H, m</td>
<td>71.8</td>
<td></td>
</tr>
<tr>
<td>6$\alpha$</td>
<td>3.8-3.94, 1H, m</td>
<td>60.5</td>
<td></td>
</tr>
<tr>
<td>6$\beta$</td>
<td>3.8-3.94, 1H, m</td>
<td>60.5</td>
<td></td>
</tr>
<tr>
<td>$\beta$-Glc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1$\beta$</td>
<td>4.64, 1H, d, 8.0</td>
<td>96.7</td>
<td>2$\beta$, 3$\beta$, 5$\beta$</td>
</tr>
<tr>
<td>2$\beta$</td>
<td>3.44, 1H, m</td>
<td>76.4</td>
<td>1$\beta$, 3$\beta$</td>
</tr>
<tr>
<td>Appendix</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3β</td>
<td>3.75, 1H, m</td>
<td>77.8</td>
<td>4β, 5β, 1°</td>
</tr>
<tr>
<td>4β</td>
<td>3.85, 1H, m</td>
<td>73.3</td>
<td>3β, 5β, 1°</td>
</tr>
<tr>
<td>5β</td>
<td>3.58, 1H, m</td>
<td>76.2</td>
<td></td>
</tr>
<tr>
<td>6β</td>
<td>~3.95, 1H, m</td>
<td>60.7</td>
<td></td>
</tr>
<tr>
<td>6β</td>
<td>~3.80, 1H, m</td>
<td>60.7</td>
<td></td>
</tr>
<tr>
<td>α-L-Fuc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1°</td>
<td>5.37-5.42, 1H, d, 4.0</td>
<td>99.2-99.4</td>
<td>3°, 5°, 3α, 3β</td>
</tr>
<tr>
<td>2°</td>
<td>3.78, 1H, m</td>
<td>68.8</td>
<td></td>
</tr>
<tr>
<td>3°</td>
<td>3.93, 1H, m</td>
<td>70.1</td>
<td>1°</td>
</tr>
<tr>
<td>4°</td>
<td>3.78, 1H, m</td>
<td>72.8</td>
<td></td>
</tr>
<tr>
<td>5°</td>
<td>4.80, 1H, m</td>
<td>67.3</td>
<td>1°, (3°), 4°, 6°</td>
</tr>
<tr>
<td>6°</td>
<td>1.15, 3H, d, 6.7</td>
<td>16.0</td>
<td>4°, 5°</td>
</tr>
<tr>
<td>β-Gal 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1'</td>
<td>4.41, 1H, d, 7.9</td>
<td>102.6</td>
<td>(3'), 5', 4α/β</td>
</tr>
<tr>
<td>2'</td>
<td>3.49, 1H, m</td>
<td>71.5</td>
<td>1', 3'</td>
</tr>
<tr>
<td>3'</td>
<td>3.70, 1H, m</td>
<td>82.4</td>
<td>2', 1*</td>
</tr>
<tr>
<td>4'</td>
<td>4.09, 1H, d, 3.3</td>
<td>69.1</td>
<td>2', 3', (5'), 6'</td>
</tr>
<tr>
<td>5'</td>
<td>3.56, 1H, m</td>
<td>75.4</td>
<td>1', 4', 6'</td>
</tr>
<tr>
<td>6'</td>
<td>3.73, 1H, m</td>
<td>62.3</td>
<td></td>
</tr>
<tr>
<td>6'</td>
<td>3.76, 1H, m</td>
<td>62.3</td>
<td></td>
</tr>
<tr>
<td>β-GlcNAc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>4.69, 1H, d, 8.2</td>
<td>103.5</td>
<td>3*, (2*), 3*, 5*</td>
</tr>
<tr>
<td>2*</td>
<td>3.78, 1H, m</td>
<td>55.9</td>
<td>CO, 1*</td>
</tr>
<tr>
<td>CO</td>
<td></td>
<td>175.7</td>
<td></td>
</tr>
<tr>
<td>CH₃</td>
<td>2.02, 3H, s</td>
<td>23.0</td>
<td>CO, 2*</td>
</tr>
<tr>
<td>3*</td>
<td>3.73, 1H, m</td>
<td>73.0</td>
<td>2*, 4*, 5*</td>
</tr>
<tr>
<td>4*</td>
<td>3.74, 1H, m</td>
<td>79.1</td>
<td></td>
</tr>
<tr>
<td>5*</td>
<td>3.57, 1H, m</td>
<td>75.4</td>
<td>3*</td>
</tr>
<tr>
<td>6*</td>
<td>3.86, 1H, m</td>
<td>60.7</td>
<td>4*</td>
</tr>
<tr>
<td>6*</td>
<td>3.94, 1H, m</td>
<td>60.7</td>
<td>4*, 5*</td>
</tr>
<tr>
<td>β-Gal 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1”</td>
<td>4.47, 1H, d, 7.8</td>
<td>103.7</td>
<td>4*, 2”, 3”, (5”)</td>
</tr>
<tr>
<td>2”</td>
<td>3.53, 1H, m</td>
<td>71.8</td>
<td>1”, 3”</td>
</tr>
<tr>
<td>3”</td>
<td>3.66, 1H, m</td>
<td>73.3</td>
<td>2”</td>
</tr>
<tr>
<td>4”</td>
<td>3.92, 1H, m</td>
<td>69.4</td>
<td>2”, 3”</td>
</tr>
<tr>
<td>5”</td>
<td>3.72, 1H, m</td>
<td>76.2</td>
<td>1”</td>
</tr>
<tr>
<td>6”</td>
<td>3.7-3.8, 1H, m</td>
<td>61.9</td>
<td></td>
</tr>
<tr>
<td>6”</td>
<td>3.7-3.8, 1H, m</td>
<td>61.9</td>
<td></td>
</tr>
</tbody>
</table>
pNP-Fuc-Fuc glycosides

\[ \text{p-nitrophenyl } \alpha-L\text{-fucopyranosyl-(1-2)}\alpha-L\text{-fucopyranoside} \]

\(^1\)H NMR: \( \delta 8.26 (d, 2H, J = 8.8 \text{ Hz}, H3''), 7.28 (d, 2H, J = 8.8 \text{ Hz}, H2''), 5.84 (d, 1H, J = 2.8 \text{ Hz}, H1), 5.11 (d, 1H, J = 3.9 \text{ Hz}, H1'), 4.34 (br. q, 1H, J = 6.5 \text{ Hz}, H5'), 4.14 (m, 1H, H2), 4.13 (m, 1H, H5), 4.11 (m, 1H, H3), 4.08 (m, 1H, H4), 3.97 (dd, 1H, J = 3.5 Hz, 10.5 Hz, H3''), 3.83 (m, 1H, H2'), 1.23 (d, 3H, J = 6.5 Hz, H6'), 1.15 (d, 3H, J = 6.7 Hz, H6).

\(^{13}\)C NMR derived from gHSQC and gHMBC: \( \delta 162.2 \) (C1''), 143.0 (C4''), 126.9 (C3''), 117.6 (C2''), 97.6 (C1), 96.3 (C1'), 75.3 (C2), 72.7 (C4'), 70.3 (C3'), 68.7 (C4), 68.8 (C2'), 68.6 (C3), 67.9 (C5'), 66.7 (C5), 16.3 (C6'), 16.2 (C6).

\[ \text{p-nitrophenyl } \alpha-L\text{-fucopyranosyl-(1-3)}\alpha-L\text{-fucopyranoside} \]

\(^1\)H NMR: \( \delta 8.26 (d, 2H, J = 9.0 \text{ Hz}, H3''), 7.28 (d, 2H, J = 9.0 \text{ Hz}, H2''), 5.98 (d, 1H, J = 3.5 \text{ Hz}, H1), 5.05 (d, 1H, J = 4.0 \text{ Hz}, H1'), 4.28 (br. q, 1H, J = 6.7 \text{ Hz}, H5'), 4.21 (dd, 1H, J = 3.4 \text{ Hz}, 10.4 Hz, H3), 4.14 (br. q, 1H, J = 6.7 \text{ Hz}, H5), 4.04 (dd, 1H, J = 3.5 Hz, 10.4 Hz, H2), 3.92 (m, 1H, H3'), 3.91 (m, 1H, H3), 3.80 (br. d, 1H, J = 3.3 Hz, H4'), 3.72 (dd, 1H, J = 3.4 Hz, 10.2 Hz, H2'), 1.21 (d, 3H, J = 6.7 Hz, H6'), 1.15 (d, 3H, J = 6.7 Hz, H6).

\(^{13}\)C NMR derived from gHSQC and gHMBC: \( \delta 162.5 \) (C1''), 143.2 (C4''), 126.9 (C3''), 117.7 (C2''), 97.0 (C1'), 95.1 (C1), 72.7 (C2), 72.5 (C4), 72.3 (C4'), 70.1 (C3'), 68.8 (C3), 68.6 (C2'), 67.9 (C5'), 66.7 (C5), 16.3 (C6'), 16.1 (C6).

\[ \text{p-nitrophenyl } \alpha-L\text{-fucopyranosyl-(1-4)}\alpha-L\text{-fucopyranoside} \]

\(^1\)H NMR: \( \delta 8.26 (d, 2H, J = 9.2 \text{ Hz}, H3''), 7.27 (d, 2H, J = 9.2 \text{ Hz}, H2''), 5.83 (d, 1H, J = 3.9 \text{ Hz}, H1), 4.95 (d, 1H, J = 4.0 \text{ Hz}, H1'), 4.54 (br. q, 1H, J = 6.6 \text{ Hz}, H5), 4.20 (m, 1H, H3), 4.18 (m, 1H, H5'), 4.04 (dd, 1H, J = 3.9 Hz, 10.6 Hz, H2), 3.94 (dd, 1H, J = 3.0 Hz, 10.8 Hz, H3'), 3.90 (br. d, 1H, J = 3.0 Hz, H4'), 3.83 (br. d, 1H, J = 3.0 Hz, H4), 3.80 (dd, 1H, J = 4.0 Hz, 10.8 Hz, H2'), 1.23 (d, 3H, J = 7 Hz, H6'), 1.17 (d, 3H, J = 6.6 Hz, H6).

\(^{13}\)C NMR derived from gHSQC and gHMBC: \( \delta 162.3 \) (C1''), 143.1 (C4''), 126.8 (C3''), 117.6 (C2''), 101.3 (C1'), 97.7 (C1), 80.5 (C4'), 70.2 (C3'), 69.7 (C3), 69.5 (C2'), 69.5 (C5'), 68.8 (C4), 68.4 (C2), 67.7 (C5), 16.2 (C6'), 16.2 (C6).
Appendix

Paper I


Inhibition of Virulence Gene Expression in *Staphylococcus aureus* by Novel Depsipeptides from a Marine *Photobacterium*.

Inhibition of Virulence Gene Expression in
Staphylococcus aureus by Novel Depsipeptides from a Marine
Photobacterium

Maria Mansson 1,*, Anita Nielsen 2, Louise Kjaerulff 3, Charlotte H. Gotfredsen 3,
Matthias Wietz 4, Hanne Ingmer 2, Lone Gram 4 and Thomas O. Larsen 1

1 Center for Microbial Biotechnology, Department of Systems Biology, Technical University of
   Denmark, DK-2800 Kgs. Lyngby, Denmark; E-Mail: tol@bio.dtu.dk
2 Department of Veterinary Disease Biology, Faculty of Life Sciences, University of Copenhagen,
   DK-1870 Frederiksberg C, Denmark; E-Mails: anini@life.ku.dk (A.N.); hi@life.ku.dk (H.I.)
3 Department of Chemistry, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark;
   E-Mails: lokja@kemi.dtu.dk (L.K.); chg@kemi.dtu.dk (C.H.G.)
4 National Food Institute, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark;
   E-Mails: mwie@food.dtu.dk (M.W.); gram@food.dtu.dk (L.G.)

* Author to whom correspondence should be addressed; E-Mail: maj@bio.dtu.dk;
   Tel.: +45-4525-2724; Fax: +45-4588-4148.

Received: 22 September 2011; in revised form: 22 November 2011 / Accepted: 30 November 2011 /
Published: 7 December 2011

Abstract: During a global research expedition, more than five hundred marine bacterial
strains capable of inhibiting the growth of pathogenic bacteria were collected. The purpose
of the present study was to determine if these marine bacteria are also a source of
compounds that interfere with the agr quorum sensing system that controls virulence gene
expression in Staphylococcus aureus. Using a gene reporter fusion bioassay, we recorded agr
interference as enhanced expression of spa, encoding Protein A, concomitantly with reduced
expression of hla, encoding α-hemolysin, and rnanIII encoding RNAIII, the effector molecule
of agr. A marine Photobacterium produced compounds interfering with agr in S. aureus
strain 8325-4, and bioassay-guided fractionation of crude extracts led to the isolation of
two novel cyclodepsipeptides, designated solonamide A and B. Northern blot analysis
confirmed the agr interfering activity of pure solonamides in both S. aureus strain 8325-4
and the highly virulent, community-acquired strain USA300 (CA-MRSA). To our
knowledge, this is the first report of inhibitors of the agr system by a marine bacterium.
Keywords: Photobacterium; Vibrionaceae; antivirulence; quorum sensing inhibition; agr

1. Introduction

Microorganisms are an attractive source of new natural products with antimicrobial properties [1,2], and the marine environment constitutes a prolific resource of bioactive microorganisms [3–5]. Many marine microenvironments stimulate the production of specific metabolites as a response to environmental factors [6]. It is likely that some of these metabolites mediate both intra- and interspecies microbial interactions, and can be seen as potential new scaffolds for development of drug lead candidates [6,7]. The increasing problem of antibiotic resistance among human pathogens highlights the need for novel therapeutic strategies [8]. The search for new avenues in microbial control has therefore been extended from traditional bacteriostatic or bacteriolytic compounds to compounds that target, for example, quorum sensing (QS) pathways [9–11]. Quorum sensing inhibitors (QSI) do not necessarily kill or inhibit the growth of a pathogen but rather modulate microbial phenotypes, for example by attenuating virulence [12,13]. In vivo studies with QS inhibitory compounds demonstrated how these can be used to slow the spread of infection [14] or enhance the clearance of pathogens from infected tissue [10].

Staphylococcus aureus is one of the main causes of nosocomial infections, and methicillin-resistant S. aureus (MRSA) are emerging at an alarming rate [15,16]. The virulence of S. aureus is ascribed to a number of virulence factors, including extracellular toxins such as α-hemolysin encoded by hla, and cell surface adhesion factors such as Protein A encoded by spa [17]. Their expression is coordinated through several key regulators, of which the agr (accessory gene regulator) QS system is central [18]. This system is a classical two-component system with a sensor histidine kinase, AgrC, and a response regulator, AgrA, in addition to AgrB and AgrD which are responsible for the production of the quorum sensing signal [19,20]. agr-dependent QS is mediated via autoinducing peptides (AIP), 16-membered thiolactone macrorcycles carrying a peptide tail that control virulence gene expression through the effector molecule RNAIII [21,22]. Structure-activity relationship studies (SAR) demonstrated that while the tail of the AIP is essential for agr activation [19,23], the macrocyclic ring is responsible for antagonistic activity [23]. This has led to the synthesis of global inhibitors based on truncated AIPs [23,24]; however, there are only few reports of natural antagonists of this system [24–27]. Nielsen et al. [28] recently developed a screening assay based on S. aureus lacZ reporter fusion strains, where the effect of compounds or extracts on expression of three key virulence genes (spa, hla, and rnaIII) and hence potential interference with the agr locus can be assessed. Subsequently, the assay was used to identify two xanthones as novel quorum sensing interfering compounds in S. aureus [28].

We recently established a global collection of marine bacteria with antibacterial activity [29]. The purpose of the present study was to determine if this strain collection also harbored organisms that produced other types of bioactive compounds and we screened pure cultures, crude extracts, and purified secondary metabolites from the collection for potential inhibitors of the agr system. One of the bacterial families we investigated was the Vibrionaceae. These bacteria are ubiquitous in marine and brackish environments and often associated with marine organisms [30]. The Vibrionaceae consist of seven genera, with the majority of species belonging to the Vibrio and Photobacterium genera. Vibrio spp. can
be pathogenic to humans [31–33] or marine animals [30], but also occur in the commensal microflora of zooplankton [30] or live as bioluminescent symbionts with squid or fish [34–36]. The *Photobacterium* genus similarly comprises symbiotic [37,38] and pathogenic species [39–41]. Members of the *Vibrionaceae* produce broad-range inhibitory compounds [7,29]; however, only few of the antibacterial compounds have been isolated to date [42]. Antimicrobial compounds from *Vibrio* species include the pyrrolidinones andrimid [43–45] and moiramide B [46] that inhibit fatty acid synthesis [47]. In addition, we recently reported the production of the potent pyrrothine antibiotic holomycin by a marine *Photobacterium* [45].

Herein, we report the isolation and chemical investigation of two novel depsipeptides produced by that same *Photobacterium* strain. The compounds, designated solonamides A and B, inhibit the *agr* QS system of *S. aureus* and therefore interfere with its virulence gene expression. This indicates that marine bacteria are a source of novel chemistry with potential use in antibacterial therapy.

2. Results and Discussion

2.1. Identification of QS Inhibitors from *Photobacterium* sp.

In an initial search for antimicrobial compounds we isolated strain S2753 related to *Photobacterium halotolerans* [29]. Subsequently, the known antibiotic, holomycin, was identified as responsible for its growth inhibitory activity [45]. When investigating ethyl acetate extracts of S2753 in an agar diffusion assay monitoring expression of the *S. aureus* virulence genes *hla*, *rnanIII*, and *spa* [28], we observed an increased expression of *spa* and decreased expression of *hla* and *rnanIII*. The inverse effect of the extracts on *spa* and *hla/rnanIII* expression, respectively, indicates the presence of at least one compound that interferes with the *S. aureus* *agr* QS system [28]. Secondary screening of the extract by explorative solid-phase extraction (E-SPE) [48] detected the potential QSI activity in a fraction that did not inhibit growth of *S. aureus* or *V. anguillarum* (data not shown). Bioassay-guided fractionation by diol and C-18 columns resulted in the isolation of two compounds active in the *S. aureus* agar diffusion assay (Figure 1). The activity of the pure compounds matched the initial activity of the extract, confirming that these compounds are responsible for the observed changes in gene expression.

**Figure 1.** Effect of solonamides (A and B) on *hla*, *rnanIII* and *spa* expression. Solonamides (5 mg mL<sup>−1</sup>) were added to wells in TSA plates containing the 8325-4 derived *lacZ* reporter strains PC322 (*hla::lacZ*), SH101F7 (*rnanIII::lacZ*), or PC203 (*spa::lacZ*). Incubation time was 15 h for plate I and II, and 35 h for plate III (plate numbering indicated with white letters). Solonamide B tested in two wells.
2.2. Structural Elucidation of the Solonamides

The solonamides were isolated as white powder with respective molecular formulas C$_{30}$H$_{46}$N$_{4}$O$_{6}$ (A) and C$_{32}$H$_{50}$N$_{4}$O$_{6}$ (B) as determined by HRMS (1 ppm mass accuracy). Analysis of NMR data characterized the structures of the solonamides as cyclodepsipeptides consisting of four amino acids and a 3-hydroxy fatty acid (Figure 2). The amino acid composition was elucidated as alanine, phenylalanine, and two leucines for both peptides based on DQF-COSY and HSQC NMR data.

**Figure 2.** Structures of solonamides A and B produced by *Photobacterium* sp. strain S2753 and structure of natural group I AIP [21].

The spin systems of the amino acids were confirmed through strong and unambiguous H2BC correlations [49] and the carbonyl signals assigned by HMBC correlations. Through careful inspection of the DQF-COSY and H2BC NMR data, solonamide A was found to contain a 3-hydroxyhexanoic acid (Hha), while solonamide B contained a 3-hydroxyoctanoic acid (Hoa). Long-range HMBC and NOESY correlation data allowed the sequence of amino acids to be established as fatty acid-Phe-Leu-Ala-Leu (Figure in Supplementary Information). This was corroborated by MS-MS experiments giving the exact molecular formula of the fragments (2 ppm mass accuracy, see Supplementary Information). The signal from one oxygen-bearing carbon with a high carbon shift indicated an ester linkage. The ring closure linkage was secured by HMBC correlations from H-3 in the fatty acid moiety to the carbonyl
in Leu and a weak NOESY correlation from H-3 to the Leu amide and Hα protons. In total, this accounted for the ten degrees of unsaturation resulting from the macrocyclic ring, five carbonyls, and the phenyl group.

The absolute configurations of the individual amino acids were established by acid hydrolysis and Marfey’s method with UHPLC analysis. Both peptides were found to contain L-Phe, D-Ala, and an enantiomeric pair of L-Leu and D-Leu. Acid hydrolysis of the reduced linear peptides and subsequent Marfey’s derivatization specified the stereochemistry of the two Leu, exchanging the L-Leu peak (RT 3.77 min) with a new peak (RT 3.73 min), attributable to the corresponding alcohol. Thus, L- and D-stereochemistry was assigned as fatty acid-L-Phe-D-Leu-D-Ala-L-Leu in both solonamide A and B.

The absolute configuration of the fatty acid was established by NMR spectroscopic analysis of the 1H and 19F chemical shift differences ($\Delta\delta$) in the $(R$- and $(S$)-Mosher’s esters analysis of solonamide A and B. The absolute stereochemistry of C-3 in the 3-hydroxy fatty acid was established as $(R$ in both depsipeptides (Figure 3).

**Figure 3.** Distribution of the $\Delta\delta$ values (ppm) calculated for the (a) 3-hydroxyhexanoic acid (Hha) and (b) 3-hydroxyoctanoic acid (Hoa) in the $(R$- and $(S$)-Mosher’s esters.

The yield of solonamides was up to 10 mg L$^{-1}$, which is a high organic yield compared to that of other $\gamma$-proteobacteria [50,51]. This suggests that they could be storage compounds accumulated during growth. However, the solonamides were also produced on a chitin based minimal medium (Supplementary Information) which indicated that these compounds may be produced in the natural habitat of vibrios, such as chitinous zooplankton. In addition, D-alanine and L-leucine are incorporated in the structure; amino acids that are not present in the laboratory medium. Thus, the *Photobacterium* seems to produce these specific stereoforms rather than incorporating the available amino acids.

2.3. Production of Solonamides by Related Photobacterium Strains

To test whether the solonamides are also produced by strains related to our *Photobacterium* isolate, we compared strain S2753 with *P. halotolerans* LMG 22194$^T$, *P. rosenbergii* LMG 22223$^T$, and
P. angustum S14 [52]. None of these strains produced solonamides (as confirmed by LC-UV/MS), and none affected virulence gene expression in the gene-reporter agar diffusion assay. None of the three strains inhibited growth of V. anguillarum, and holomycin, the antibiotic of S2753 [45], was not detected.

To the best of our knowledge, only two species of Photobacterium have been investigated for their secondary metabolites so far [51,53,54]. Oku et al. [51] isolated unnarmicin A and C from a marine Photobacterium strain MBIC06485 related to P. leiognathi. Like the solonamides, unnarmicin A and C consist of four amino acids (L-Phe, L-Leu, D-Phe, L-Leu) and a 3-hydroxyoctanoic and 3-hydroxyhexanoic fatty acid, respectively. The finding of the unnarmicins in another marine Photobacterium sp. indicates that production of such peptides could be a common feature in this group of bacteria, despite the absence of solonamides in any related strains that we investigated.

2.4. Solonamides Interfere with agr

To verify that the purified solonamides do in fact cause transcriptional changes in virulence gene expression and to assess if the effect is strain specific, we isolated mRNA from S. aureus 8325-4 and the community-acquired strain, USA300 at different stages of growth following solonamide exposure and monitored gene expression by Northern blot analysis. Solonamide B dramatically reduced the expression of both hla and rnaIII while increasing expression of spa, strongly indicating that the compound interferes with agr regulation (Figure 4). The decreased expression of rnaIII was even more pronounced in the highly virulent USA300 strain where high agr activity is suspected to be a main contributor to the virulence of the strain [55,56]. The solonamides did not affect the growth rate of the liquid S. aureus cultures. Solonamide A was able to increase spa expression, but caused only a marginal reduction of hla and rnaIII expression in both 8325-4 and USA300. The discrepancy between the Northern blot analysis and the agar diffusion assay (mainly with regard to hla) may be rooted in the much higher concentrations of solonamides that are used in the agar diffusion assay as compared to the Northern blot analysis. Also, the Northern blot analysis directly monitors the amount of mRNA shortly after solonamide addition, whereas the agar diffusion assay relies on the accumulation of β-galactosidase enzyme over a period of 15 or 35 h.

The structural similarity of the solonamides and the AIPs (Figure 2) suggest that they may be competitive inhibitors of the agr system. Unlike the AIPs, the solonamides are cyclized through a 3-hydroxy fatty acid forming a lactone rather than a thiolactone. However, synthetic lactone and lactam variations of natural AIPs have been found to have antagonistic activity [19,23,57], which our study corroborates. While inhibition of agr by AIPs is more tolerant of sequence and structural diversity than is activation [23], Mayville et al. [14] found that the presence of the hydrophobic leucine and phenylalanine residues is important for the inhibition of the agr response. Both solonamides contain a leucine and phenylalanine; however, the reduced activity of solonamide A indicates that the overall hydrophobicity of the depsipeptides affected by the varying length of the fatty acid moiety might be an important factor influencing activity.

The solonamides are the first reported antagonists produced by a natural source with a structure resembling that of native S. aureus AIPs. Kiran et al. (2008) [25] identified hamamelitannin from Hamamelis virginiana (witch hazel) as an inhibitor of RNAIII and δ-hemolysin production in S. aureus 8325-4, USA300, and clinical S. epidermidis isolate MH. Also, ambuic acid from an unidentified fungal
strain was found to attenuate \textit{agr} [26]. Given the relatively low abundance of staphylococci in the marine environment, it seems unlikely that the \textit{Photobacterium} sp. S2753 produces solonamides as part of a deliberate strategy to interfere with this specific type of bacteria. However, the solonamides might be targeted at other Gram-positive bacteria in the marine environment, such as bacilli and actinobacteria, though little is known about quorum sensing pathways in marine Gram-positive bacteria. A large number of different QS systems have been characterized from Gram-negative bacteria in the marine environment [57–62]. We speculate that the solonamides could also affect such systems despite sharing little structural similarity to agonists and antagonists of the systems identified to date [63]. Acylated homoserine lactones, the most widely researched type of QS molecules in Gram-negatives, can serve as both agonists and antagonists in different systems [27], and thus the solonamides may also serve as quorum sensing signals for the \textit{Photobacterium} itself. However, we did not detect solonamides or compounds with similar QSI activity in any of the related strains.

\textbf{Figure 4.} Effect of solonamide A and B on virulence gene expression in \textit{S. aureus} strain 8325-4 [18] and USA300 [55] examined by Northern blot analysis. Solonamides were added to exponentially growing cultures at OD\textsubscript{600} = 0.4, and RNA was purified at OD\textsubscript{600} = 0.7 and 1.7. The RNA was reacted with probes recognizing \textit{hla}, \textit{rnanIII}, and \textit{spa}, respectively. Solonamide B tested in duplicates. DMSO was used as negative control.

Our findings suggest that quorum sensing inhibition could be an alternative therapeutic strategy for treatment of MRSA \textit{S. aureus} infections; however, the effect in an \textit{in vivo} infection model needs to be tested. Many genes are under QS control both in Gram-positive and Gram-negative bacteria, and thus it is not a simple drug target [64,65]. For example, biofilm formation in \textit{S. aureus} has been linked to
low QS activity [66], and so there is a risk that the use of QS inhibitors could lead to decreased susceptibility of traditional antibiotics. Also, it is still unknown how QS inhibition will affect the overall fitness of a pathogen under in vivo conditions and thus pose a selective pressure for development of resistance [67] or enhanced virulence [68,69].

3. Experimental Section

3.1. Isolation and Identification of Strain S2753

Bacterial strain S2753 was isolated from a mussel surface collected in the tropical Pacific (9.1°S 156.8°E) during the Danish Galathea 3 expedition [29]. S2753 was assigned to the Vibrionaceae by 16S rRNA gene sequence similarity [29] and identified as being closely related to Photobacterium halotolerans based on recA and rpoA gene sequences, with homologies of 87% (recA) and 94% (rpoA) [45]. BLAST analyses showed that other closely related species were Photobacterium rosenbergii and Photobacterium angustum S14 [52].

3.2. Initial Screening for Anti-Virulence Compounds

S2753 was cultured in each 30 mL of (i) marine minimal medium (MMM) [70] containing 0.4% glucose and 0.3% casamino acids; (ii) Marine Broth 2216 (MB; Difco 2216); (iii) Sigma Sea Salts (SSS; Sigma S9883; 40 g L⁻¹) containing 0.4% glucose and 0.3% casamino acids, and (iv) MMM containing 0.2% colloidal chitin [71,72] to investigate the best conditions for production of antibacterial compounds. All cultures were incubated aerated (200 rpm) for 72 h at 25 °C. All cultures were extracted with an equal volume of ethyl acetate (EtOAc). The extract was evaporated under nitrogen until dryness and redissolved in 300 μL 80% v/v ethanol (EtOH) in water. 20 μL of the extract was tested in an agar diffusion assay where expression from promoters of hla, rnalIII, and spa is monitored [28].

For further screening, the culture broth of S2753 grown in MMM (1 L, 72 h, 25 °C, 100 rpm) was extracted directly with sterile Diaion HP20SS (12 g L⁻¹, 24 h) (Sigma-Aldrich, St. Louis, MO). The resin was filtered off and washed with 80% (v/v) acetonitrile (MeCN)/water (300 mL) and the extract evaporated until dryness on a rotary evaporator. From this extract, 10 mg dry material was subjected to an explorative solid-phase extraction (E-SPE) protocol [48]. This yielded 15 fractions for re-testing in the agar diffusion assay as described above. The E-SPE fractions were also tested for antibacterial activity against Vibrio anguillarum strain 90-11-287 and S. aureus 8325 in a well diffusion agar assay [73].

3.3. Isolation and Structural Elucidation of Solonamide A and B

Using 10 L glass fermentors, S2753 was cultured in 5 × 4 L SSS (iii, above) containing 0.4% glucose and 0.3% casamino acids (25 °C, 72 h, 100 rpm) as this medium gave comparable yields to that of MMM (i) but at a lower cost. The broth was extracted directly with Diaion HP20SS (12 g L⁻¹) as described above. The extract (3.4 g) was redissolved in EtOAc, absorbed onto 5 g Isolute diol (Biotage, Uppsala, Sweden), and added to a glass column with pure diol (95 g). A total of 12 fractions were collected from the diol column (100 g, 20 × 350 mm) ranging from heptane, dichloromethane (DCM), EtOAc, to pure methanol (MeOH), running under gravity. The fractions containing the QSI compounds (fraction 5, 20:80 (v/v) EtOAc/DCM and fraction 6, 40:60 (v/v) EtOAc/DCM) were...
separated on Sepra ZT C\textsubscript{18} (Phenomenex, Torrance, CA) (10 g SNAP) on an Isolera automated flash system (Biotage) using a MeCN/water gradient 25–75\% over 20 min (12 mL min\textsuperscript{−1}). Pure compounds were obtained directly: Solonamide A (17 mg) and B (201 mg). Activity of pure compounds was tested in the agar diffusion assay as described above (20 \& \mu L per well), with a final concentration of compounds of 5 mg mL\textsuperscript{−1} in dimethyl sulfoxide (DMSO).

NMR spectra were recorded on a Varian Unity Inova 500 MHz spectrometer equipped with a 5 mm probe using standard pulse sequences. The signals of the solvent were used as internal references (δ\textsubscript{H} 2.49 and δ\textsubscript{C} 39.5 ppm for DMSO). Carbonyl shifts were confirmed with \textsuperscript{13}C 1D on a Bruker Avance 800 MHz spectrometer with a 5 mm TCI cryoprobe at the Danish Instrument Center for NMR Spectroscopy of Biological Macromolecules.

LC-MS and LC-MS/MS analyses were performed on a maXis quadrupole time of flight mass spectrometer (Bruker Daltonics, Bremen, Germany) equipped with an electrospray (ESI) ion source. The MS was connected to an Ultimate 3000 UHPLC system (Dionex, Sunnyvale, CA) equipped with a diode-array detector. Separation was performed at 40 °C on a 150 mm × 2.1 mm ID, 2.6 μm Kinetex C\textsubscript{18} column (Phenomenex) using a linear water/MeCN (both buffered with 20 mM formic acid) gradient starting from 15\% MeCN and increased to 100\% in 13 min at a flow of 0.4 mL min\textsuperscript{−1}. The MS and MS/MS experiments were performed in ESI\textsuperscript{+} with a data acquisition range of m/z 100–1200 with collision energy of 27 V. The MS was calibrated using sodium formate automatically infused prior to each analytical run, providing a mass accuracy of below 1 ppm in MS mode and 2 ppm in MS/MS mode.

The absolute configuration of the amino acids were found using acid hydrolysis (6 M HCl, 110 °C, 20 h) \cite{74} and derivatisation with Marfey’s reagent (1-fluoro-2,4-dinitrophenyl-5-L-alanine amide, FDAA, Sigma-Aldrich) following the protocol by Bonnard et al. \cite{75}. Ultra-high liquid chromatography-diode array (UHPLC-UV) analyses of the amino acids were done on a Dionex RSLC Ultimate 3000 (Dionex) equipped with a diode-array detector. Separation was obtained on a Kinetex C\textsubscript{18} column (150 × 2.10 mm, 2.6 μm, Phenomenex) maintained at 60 °C using a linear gradient starting from 25\% MeCN in water (both buffered with 50 ppm TFA) increasing to 27\% MeCN over 6 min at a flow rate of 0.8 mL min\textsuperscript{−1}. Retention times of the FDAA amino acid derivatives used as standards were as follows (maximum standard deviation ± 0.002 min): FDAA (1.50 min), L-Ala (1.14 min), D-Ala (1.61 min), L-Phe (3.58 min), D-Phe (5.04 min), L-Leu (3.77 min), D-Leu (5.49 min), comparable to the observed retention times from the solonamide-derived amino acids.

To specify the stereochemistry of enantiomeric amino acids, the depsipeptides were reduced by LiBH\textsubscript{4}. The resulting linear peptides were subjected to the above mentioned acid hydrolysis and Marfey’s derivatisation. Details are given in the Supplementary Information.

For the absolute configuration of the fatty acid residues, the (R)- and (S)-Mosher’s esters were prepared for both depsipeptides, and the stereocenters were assigned based on their \textsuperscript{1}H and \textsuperscript{19}F chemical shift differences (Δδ\textsuperscript{SR}) \cite{76,77}. Details are given in the Supplementary Information.

\textit{Solonamide A}: white amorphous solids; UV (MeCN/H\textsubscript{2}O) λ\textsubscript{max} 200 nm (100\%); for NMR data refer to Table 1; HRESIMS m/z 558.3486 (calcd for C\textsubscript{30}H\textsubscript{46}N\textsubscript{4}O\textsubscript{6}, 558.3496).

\textit{Solonamide B}: white amorphous solids; UV (MeCN/H\textsubscript{2}O) λ\textsubscript{max} 200 nm (100\%); for NMR data refer to Table 1; HRESIMS m/z 586.3725 (calcd for C\textsubscript{32}H\textsubscript{50}N\textsubscript{4}O\textsubscript{6}, 586.3730).
Table 1. NMR spectroscopic data (DMSO-$d_6$) of solonamide A and B.

<table>
<thead>
<tr>
<th>Atom</th>
<th>δ$_C$ (ppm)</th>
<th>δ$_H$ (ppm) (multiplicity, $J$ (Hz))</th>
<th>HMBC</th>
<th>Atom</th>
<th>δ$_C$ (ppm)</th>
<th>δ$_H$ (ppm) (multiplicity, $J$ (Hz))</th>
<th>HMBC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hha/Hoa</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>L-Phe</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>170.6</td>
<td></td>
<td></td>
<td>1</td>
<td>170.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>40.7</td>
<td>2.67 (1H, dd, 13.5, 3.8)</td>
<td>1,3,4</td>
<td>2</td>
<td>40.5</td>
<td>2.69 (1H, dd, 13.5, 3.7)</td>
<td>1,3,4</td>
</tr>
<tr>
<td>3</td>
<td>72.1</td>
<td>5.14 (1H, m)</td>
<td>1,5</td>
<td>3</td>
<td>72.0</td>
<td>5.13 (1H, m)</td>
<td>1,4,5,36</td>
</tr>
<tr>
<td>4</td>
<td>36.2</td>
<td>1.62 (1H, m)</td>
<td>2,3,6</td>
<td>4</td>
<td>33.8</td>
<td>1.65 (1H, m)</td>
<td>5,6</td>
</tr>
<tr>
<td>5</td>
<td>17.7</td>
<td>~1.22 (1H, m)</td>
<td>3,4</td>
<td>5</td>
<td>23.8</td>
<td>~1.20 (2H, m)</td>
<td>6,7</td>
</tr>
<tr>
<td>6</td>
<td>13.8</td>
<td>0.85 (3H, t, 7)</td>
<td>4,5</td>
<td>6</td>
<td>30.8</td>
<td>~1.22 (2H, m)</td>
<td>ND</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>8.61 (1H, d, 3.2)</td>
<td></td>
<td>7</td>
<td>21.8</td>
<td>1.25 (2H, m)</td>
<td>ND</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td></td>
<td></td>
<td>8</td>
<td>13.6</td>
<td>0.84 (3H, m)</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td></td>
<td></td>
<td>9</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>135.8</td>
<td></td>
<td></td>
<td>10</td>
<td>136.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11,15</td>
<td>129.1</td>
<td>7.18 (2H, m)</td>
<td>9,13</td>
<td>11,15</td>
<td>128.7</td>
<td>7.20 (2H, m)</td>
<td>11,15</td>
</tr>
<tr>
<td>12,14</td>
<td>128.4</td>
<td>7.25 (2H, m)</td>
<td>10</td>
<td>12,14</td>
<td>128.0</td>
<td>7.26 (2H, m)</td>
<td>12,14/16</td>
</tr>
<tr>
<td>13</td>
<td>126.5</td>
<td>7.18 (1H, m)</td>
<td>11</td>
<td>13</td>
<td>126.2</td>
<td>7.20 (1H, m)</td>
<td>13/17</td>
</tr>
<tr>
<td>14</td>
<td>16-CO</td>
<td>174.4</td>
<td></td>
<td>14</td>
<td>174.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>-</td>
<td></td>
<td></td>
<td>15</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>-</td>
<td></td>
<td></td>
<td>16</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>d-Leu</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>d-Ala</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>-</td>
<td>8.63 (1H, d, 5.6)</td>
<td>16,18,19</td>
<td>17</td>
<td>-</td>
<td>8.64 (1H, d, 5.6)</td>
<td>18,20,21</td>
</tr>
<tr>
<td>18</td>
<td>53.1</td>
<td>3.56 (1H, m)</td>
<td>ND</td>
<td>18</td>
<td>52.7</td>
<td>3.57 (1H, m)</td>
<td>21,22,25</td>
</tr>
<tr>
<td>19</td>
<td>38.9</td>
<td>1.36 (1H, m)</td>
<td>18,21,22</td>
<td>19</td>
<td>38.6</td>
<td>1.31–1.37 (2H, m)</td>
<td>20,22,24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.29 (1H, m)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>25.7</td>
<td>0.97 (1H, m)</td>
<td>21</td>
<td>20</td>
<td>23.3</td>
<td>0.99 (1H, m)</td>
<td>ND</td>
</tr>
<tr>
<td>21</td>
<td>23.2</td>
<td>0.69 (3H, d, 6.6)</td>
<td>19,20</td>
<td>21</td>
<td>22.9</td>
<td>0.70 (3H, d, 6.6)</td>
<td>21,22,24</td>
</tr>
<tr>
<td>22</td>
<td>20.7</td>
<td>0.52 (3H, d, 6.5)</td>
<td>19,20</td>
<td>22</td>
<td>20.4</td>
<td>0.53 (3H, d, 6.4)</td>
<td>21,23</td>
</tr>
<tr>
<td>23-CO</td>
<td>171.8</td>
<td></td>
<td></td>
<td>23-CO</td>
<td>171.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>-</td>
<td>7.53 (1H, d, 8.9)</td>
<td>27,29</td>
<td>24</td>
<td>-</td>
<td>7.48 (1H, d, 8.9)</td>
<td>25,27,28</td>
</tr>
<tr>
<td>25</td>
<td>48.2</td>
<td>4.18 (1H, m)</td>
<td>26,27</td>
<td>25</td>
<td>47.8</td>
<td>4.18 (1H, m)</td>
<td>28,29</td>
</tr>
<tr>
<td>26-CO</td>
<td>16.6</td>
<td>1.32 (3H, d, 7.4)</td>
<td>27</td>
<td>26-CO</td>
<td>16.4</td>
<td>1.33 (3H, d, 7.3)</td>
<td>27,29</td>
</tr>
<tr>
<td>27-CO</td>
<td>171.3</td>
<td></td>
<td></td>
<td>27-CO</td>
<td>170.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>-</td>
<td>7.08 (1H, d, 10)</td>
<td>27,29</td>
<td>28</td>
<td>-</td>
<td>7.05 (1H, d, 10)</td>
<td>29,31</td>
</tr>
<tr>
<td>29</td>
<td>49.0</td>
<td>4.46 (1H, dt, 10, 4.4)</td>
<td>30,31,34</td>
<td>29</td>
<td>48.6</td>
<td>4.47 (1H, dt, 10, 4.3)</td>
<td>32,33,29/36</td>
</tr>
<tr>
<td>30-CO</td>
<td>39.3</td>
<td>1.63 (1H, m)</td>
<td>29,31,32,37/34</td>
<td>30-CO</td>
<td>39.0</td>
<td>1.64 (1H, m)</td>
<td>31,33,35,29/36</td>
</tr>
<tr>
<td>31</td>
<td>23.9</td>
<td>1.50 (1H, m)</td>
<td>29,30,33</td>
<td>31</td>
<td>23.8</td>
<td>1.50 (1H, m)</td>
<td>32,34,35</td>
</tr>
<tr>
<td>32-CO</td>
<td>23.2</td>
<td>0.86 (3H, d, 6.5)</td>
<td>30,31,33</td>
<td>32-CO</td>
<td>23.0</td>
<td>0.86 (3H, d, 6.6)</td>
<td>32,33,35</td>
</tr>
<tr>
<td>33-CO</td>
<td>21.3</td>
<td>0.81 (3H, d, 6.5)</td>
<td>30</td>
<td>33-CO</td>
<td>21.1</td>
<td>0.82 (3H, d, 6.4)</td>
<td>32,33</td>
</tr>
<tr>
<td>34-CO</td>
<td>171.2</td>
<td></td>
<td></td>
<td>34-CO</td>
<td>170.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.4. LC-UV/MS Analyses of Related Photobacterium Strains

To investigate the potential production of agr inhibitors in related Photobacterium strains, the metabolite profile of S2753 was compared by liquid chromatography-diode array/mass spectrometry (LC-UV/MS) to P. halotolerans (LMG 22194\(^T\)), P. rosenbergii (LMG 22223\(^T\)), and P. angustum (S14) described by de Nys et al. [52]. All strains were grown in 30 mL MMM containing 0.4% glucose and 0.3% casamino acids (25 °C, 72 h, 200 rpm). Cultures were extracted with an equal volume of EtOAc and evaporated under nitrogen. Residues were redissolved in MeOH for LC-UV/MS analyses and in 80% EtOH for bioassay testing. Inhibition of agr was tested as described above. Extracts were also tested against V. anguillarum strain 90-11-287 for growth inhibition. LC-UV/MS analyses were performed on an Agilent 1100 liquid chromatograph with a diode array detector (Agilent, Waldbronn, Germany) coupled to an LCT TOF mass spectrometer (Micromass, Manchester, UK) using a Z-spray ESI source. Separation was obtained on a Luna II C\(_{18}\) column (50 × 2 mm, 3 μm, Phenomenex) fitted with a security guard system using a linear gradient starting from 15% MeCN in water (both buffered with 20 mM formic acid) increasing to 100% MeCN over 20 min at a flow rate of 0.3 mL min\(^{-1}\).

3.5. Northern Blot Analysis

Northern blot analysis was performed as described previously [78]. The strains used were S. aureus FPR 3757 [55], a CA-MRSA USA300 obtained from ATCC (Boras, Sweden), and 8325-4 [17]. Samples for RNA purification were taken from cultures in Tryptone Soya Broth (TSB, Oxoid, Greve, Denmark) shaking at 185 rpm at 37 °C in a water bath (10 mL culture in 100 mL Erlenmeyer flask). Growth was monitored by measuring optical density at OD\(_{600}\). Start inoculum was OD\(_{600}\) = 0.03. Solonamides were added at OD\(_{600}\) = 0.4. Samples for RNA purification were taken at OD\(_{600}\) = 0.7 and 1.7. Probes targeting rnaIII, spa, and hla transcripts were amplified by PCR using primers hla forward (5′-GGG TTA GCC TGG CCT TCA GCC-3′), hla reverse (5′-GGG TGC CAT ATA CCG GGT TC-3′), spa forward (5′-GGG GGT GTA GGT ATT GCA TCT G-3′), spa reverse (5′-GGG GCT CCT GAA GGA TCG TC-3′), rnalII forward (5′-GGG GAT CAT ACA AGA GAT GTG ATG-3′), and rnalIII reverse (5′-GGG CAT AGC ACT GAG TCC AAG G-3′)(TAG Copenhagen A/S, Denmark). The resulting PCR fragments were 311 bp (hla), 719 bp (spa), and 316 bp (rnalIII), respectively.

4. Conclusions

The rapid, worldwide increase in antibiotic-resistant S. aureus [15] has led to an intense search for compounds with potential use in alternative therapeutic strategies [9]. Virulence of S. aureus involves a complex set of proteins, with the agr QS system controlling expression of several of the virulence genes. The investigation of crude extracts and fractions from a marine Photobacterium led to the identification of two novel depsipeptides, solonamides A and B, as potent inhibitors of this system. Interestingly, we found that solonamide B interfered with agr not only in S. aureus 8325-4 but also in strain USA300, which is a predominant community-acquired MRSA (CA-MRSA) strain in the US [79]. This finding suggests that quorum sensing inhibition could be an option for treatment of S. aureus USA300 infections. Future experiments will reveal the extent to which the solonamides are effective in treating S. aureus infections. In combination with other recent work from our laboratory [42,45,71]
the present study underlines that vibrios are a promising potential source of novel bioactive secondary metabolites.

Acknowledgments

We thank Anita Iversen for ultra high-resolution MS data and the Danish Center for NMR Spectroscopy of Biological Macromolecules for NMR time. Funding from the Programme Committee for Food, Health and Welfare under the Danish Strategic Research Council is acknowledged. The present work was carried out as part of the Galathea 3 expedition under the auspices of the Danish Expedition foundation. This is Galathea 3 contribution p85.

References


Samples Availability: Available from the authors.

© 2011 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).
Supplementary Information

A1: LC-MS selective ion-trace of solonamide A (m/z 558, RT = 11.43 min) and solonamide B (m/z 586, RT = 13.25 min) production on different growth media, including marine minimal media with glucose (MMM), marine broth (MB), sigma sea salts (SSS), and marine minimal media with chitin (Chitin). Slight shift in retention times due to inter-batch variations on the LC-MS system.
NMR

Assignments of the depsipeptides were performed using a conventional assignment approach. HMBC correlations between 3-H/1-CO, 3-H/36-CO and NOEs between 3-H/31-Hα-Leu confirms the lactone closure of the structure. The figure (A3) below shows key correlation for the lactone ring closure.
A4. 1D $^1$H spectrum of Solonamide A
A5. 1D $^1$H spectrum of Solonamide B
MS/MS data

**A6:** MS and MS/MS spectrum of solonamide A ($C_{30}H_{46}N_4O_6$, calcd. 558.3496, $\Delta m/z$ 0.8 ppm) with characteristic fragments and their molecular formulas ($\Delta m/z$ 0.2-0.6 ppm):

**A8:** MS and MS/MS spectrum of solamides B ($C_{32}H_{50}N_4O_6$, calcd. 586.3730, $\Delta m/z$ 0.0 ppm) with characteristic fragments and their molecular formulas ($\Delta m/z$ 0.2-0.6 ppm):
A7: Characteristic fragments for solonamide A (left) and B (right):

Establishment of absolute configuration

The absolute configuration of C3 at 3-hydroxyhexanoic acid (Hha) and 3-hydroxyoctanoic acid (Hoa) was determined using the Mosher’s method. Here with the Mosher’s method two enantiomers of an acid chloride (α-methoxy-α-trifluoromethylphenylacetic acid chloride, MTPA-Cl) are reacted with the secondary alcohol at the stereogenic center in order to reveal the configuration of this center. The reaction yields two diastereoisomers with known configuration around one of the stereogenic centers enabling the determination of the other stereogenic center.

The sign of the Δδ values can be seen in the figures below, all pointing toward the same configuration at the C3 stereocenter in Hha and Hoa. The values for Δδ are positive for the alkane end of Hha/Hoa and negative towards the carbonyl end, which points towards an R-configuration of the C3 stereocenter. The difference seen in the fluorine NMR data supports this conclusion when
the alkane chain of the molecule is regarded as least bulky compared to the remaining part of the molecule with a $\Delta\delta(SR) \sim -0.68$ ppm.

A8. Distribution of the $\Delta\delta$ values calculated for the 3-hydroxyhexanoic acid (Hha) of solonamide A

A9. Distribution of the $\Delta\delta$ values calculated for the 3-hydroxyoctanoic acid (Hoa) of solonamide B

**Experimental procedure used for Mosher’s:**

**Methanolysis of solonamide A (and B)**

5 mg (8.52 μmol) of the depsipeptide (solonamide A) is dissolved in 1.5 mL of 0.5 M NaOMe in MeOH and stirred at room temperature for 1 h. The reaction mixture is neutralised by careful addition of 1M HCl (aq) and the pH is monitored with universal indicator paper. The reaction mixture is concentrated to dryness on a rotary evaporator, and partitioned in EtOAc and H$_2$O.
Methanolysis of solonamide B was performed using the same procedure. The organic phase is concentrated to dryness on a rotary evaporator and the methyl ester product verified by LC-MS ([M+H]^+ A: 591.3803 Da, B: 619.4053 Da) and purified by RP-HPLC on a Waters HPLC with a 600 controller and a 996 photodiode array detector using a water/MeCN gradient elution from 40 to 60 % MeCN over 20 min (5 mL/min) with a Luna C_{18} column (5 μm, 250x10 mm, Phenomenex). Pure methyl esters of solonamide A (1.1 mg) and B (2.7 mg).

**Preparation of the R-MTPA ester of solonamide A (and B)**

To a stirred solution of the purified methanolysis product (0.55 mg, 0.889 μmol) in dry pyridine (150 μL) is added 5 μL (catalytic amount, 0.1 eq) of a solution of dimethylaminopyridine (DMAP) in dry pyridine (0.4 mg, 50 μL). After 10 min 9.1 μL S-(+)-MTPA-Cl (48.6 μmol) is added and the reaction is left overnight at room temperature. The reaction mixture is dried on a speedvac, redissolved in MeOH and purified on a Waters HPLC with a 600 controller and a 996 photodiode array detector using a water/MeCN gradient elution from 45 to 100 % MeCN over 20 min (5 mL/min) with a Luna C_{18} (5 μm, 250x10 mm, Phenomenex) column to give pure R-MTPA ester (A: 0.5 mg, B: ~1.2 mg).

**Preparation of the S-MTPA ester of solonamide A (and B)**

Using the same procedure as for the R-MTPA ester, the S-MTPA ester is produced by reaction with R-(−)-MTPA-Cl. Pure R-MTPA esters (A: 0.4 mg, B: ~1.1 mg).
Establishing stereochemistry of enantiomeric Leu by reduction and Marfey’s

2 mg of solonamide B is reduced in 2 M LiBH₄ in THF at 0°C for 30 min, then at rt overnight and lastly heated to 50°C for 30 min. EtOAc is added and the reaction is quenched with sat. NH₄Cl (aq) before the product is extracted with EtOAc, CH₂Cl₂ and Et₂O concentrated and purified by RP-HPLC on a Waters HPLC with a 600 controller and a 996 photodiode array detector using a water/MeCN gradient elution from 25 to 80% MeCN over 20 min (5 mL/min) with a Luna C₁₈ column (5 µm, 250x10 mm, Phenomenex). Yields 0.7 mg of reduced open-chain solonamide B. The same procedure is used for solonamide A, only with a gradient of 35-70 %, yielding 0.8 mg.

300 µg of each peptide is hydrolysed with 300 µL 6 M HCl at 110°C for 20 h. To the hydrolysis products is added 75 µL water, 20 µL 1 M NaHCO₃ solution and 100 µL 1% FDAA in acetone, followed by reaction at 40°C in 1 h. The vial is removed from the heat, neutralised with 10 µL 2 M HCl and the solution is diluted with 395 µL MeOH to a total volume of 0.5 mL. The FDAA derivatives are analysed by UPLC on a Dionex Ultimate 3000 with a diode array detector and a Kinetex C₁₈ column (2.6 µm, 150x2.10 mm, Phenomenex). The analyses are run with a gradient elution of water/MeCN from 25 to 37 % MeCN over 6 min (60°C, 0.8 mL/min) and the FDAA derivatives of the hydrolysates are compared to retention times of the standard amino acid derivatives: D-Ala (1.61 min), L-Ala (1.14 min), D-Leu (5.49 min), L-Leu (3.77 min), D-Phe (5.04 min), L-Phe (3.585 min). FDAA elutes at 1.50 min.
Paper II


Identification of four new *agr* quorum sensing-interfering cyclodepsipeptides from a marine *Photobacterium*.

Marine Drugs 2013, 11, 5051–5062.
Identification of Four New agr Quorum Sensing-Interfering Cyclodepsipeptides from a Marine Photobacterium

Louise Kjaerulff 1, Anita Nielsen 2, Maria Mansson 3, Lone Gram 3, Thomas O. Larsen 3, Hanne Ingmer 2,* and Charlotte H. Gotfredsen 1,*

1 Department of Chemistry, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark; E-Mail: lokja@kemi.dtu.dk
2 Department of Veterinary Disease Biology, Faculty of Health and Medical Sciences, University of Copenhagen, DK-1870 Frederiksberg C, Denmark; E-Mail: anini@sund.ku.dk
3 Department of Systems Biology, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark; E-Mails: maj@bio.dtu.dk (M.M.); gram@bio.dtu.dk (L.G); tol@bio.dtu.dk (T.O.L.)

* Authors to whom correspondence should be addressed; E-Mails: hi@sund.ku.dk (H.I.); chg@kemi.dtu.dk (C.H.G.); Tel.: +45-3533-2773 (H.I.); Fax: +45-3533-2755 (H.I.); Tel.: +45-4525-2148 (C.H.G.); Fax: +45-4593-3968 (C.H.G.).

Received: 8 October 2013; in revised form: 21 November 2013 / Accepted: 2 December 2013 / Published: 12 December 2013

Abstract: During our search for new natural products from the marine environment, we discovered a wide range of cyclic peptides from a marine Photobacterium, closely related to P. halotolerans. The chemical fingerprint of the bacterium showed primarily non-ribosomal peptide synthetase (NRPS)-like compounds, including the known pyrothine antibiotic holomycin and a wide range of peptides, from diketopiperazines to cyclodepsipeptides of 500–900 Da. Purification of components from the pellet fraction led to the isolation and structure elucidation of four new cyclodepsipeptides, ngercheumicin F, G, H, and I. The ngercheumicins interfered with expression of virulence genes known to be controlled by the agr quorum sensing system of Staphylococcus aureus, although to a lesser extent than the previously described solonamides from the same strain of Photobacterium.

Keywords: Photobacterium; depsipeptide; structure elucidation; quorum sensing; antivirulence; agr
1. Introduction

The marine environment is a rich and vastly underexplored resource in many aspects. Most of the Earth’s surface is covered by water, inhabited by an incredible diversity of species, many of which have yet to be discovered. Microbial species are an important source for marine chemodiversity and it is believed that marine microorganisms will provide valuable drug candidates in the future [1,2]. There are however certain challenges with respect to sampling and culturing, and the usually low yields of metabolites can hamper full structural characterization and biological explorations.

A marine Photobacterium was selected from 500 bacterial strains collected during a global marine research cruise in 2006–2007 [3]. It was sampled from a mussel surface near the Solomon Islands and this bacterial strain was singled out as particularly interesting because of a two-pronged inhibitory effect of growth and quorum sensing (QS) in important human pathogens. The known pyrrothine antibiotic holomycin was responsible for the growth inhibition of V. anguillarum and S. aureus [4], while the solonamides were identified as the major contributors to the interference with agr quorum sensing in S. aureus [5]. Inhibition of virulence factor production and activity has been suggested as a new therapeutic approach suitable for antibiotic resistant pathogens [6]. In S. aureus, one of the possible targets is the agr quorum sensing system that in response to autoinducing peptide (AIP) accumulation at high cell densities induces expression of numerous extracellular toxins including hla encoding α-hemolysin while repressing expression of surface factors such as the spa encoded Protein A [7].

In this study, we report the identification of four cyclodepsipeptides in the 800–900 Da size range from the cell-associated (pellet) fraction of this marine Photobacterium that modulate expression of agr controlled virulence genes. The depsipeptides ngercheumicin F, G, H, and I are new additions to the structural family that initially constituted ngercheumicin A and B [8]. Three other depsipeptides ngercheumicin C, D, and E were isolated from the same bacterial strain as ngercheumicin A and B, but they are structurally different. All ngercheumicins reported to date have been isolated from Photobacterium spp. and a biological activity reported by Shizuri et al. for ngercheumicins A–E was against infections by Pseudovibrio denitrificans [8]. Like the solonamides [5], the ngercheumicins are 16-membered macrocyclic depsipeptides with some structural resemblance to the AIPs of S. aureus. Generally, AIPs consist of a cyclopentapeptide moiety cyclized through a cysteine residue by a thiolactonization, and with an exocyclic peptide chain of variable length extended from the cysteine residue in the N-terminal direction [9]. The exocyclic chain appears to be closely related to agonistic activity as truncated AIPs are known to have antagonistic properties [10]. Structure-activity relationship studies by Mayville et al. [11] indicated that adjacent Leu and Phe residues are required for inhibitory activity; however, structural comparisons of 24 natural staphylococcal AIPs later showed that they consistently have bulky, hydrophobic amino acid side chains in the C-terminus [12]. This may instead be the structural requirement for activity.

Here, we describe the isolation and structure elucidation of the four new ngercheumicins and discuss their role in QS.
2. Results and Discussion

2.1. Isolation and Structure Elucidation of Ngercheumicins F–I

The pellet of the Photobacterium sp. was extracted with organic solvents (see Experimental Section 3.2) and fractionated on a diol column. Mass spectrometric analysis revealed a series of peptide-like analogues which display good ionization in ESI+ MS and end absorption in UV spectroscopy. The fractions containing these analogues were pooled and subjected to further purification, first on a smaller diol column and then by preparative reversed phase HPLC, which gave four fractions of ngercheumicin F, G, H, and I, respectively. Ngercheumicins A and B were also detected by LC-MS, but they were not purified in sufficient amounts for structural or biological screening studies. The four new ngercheumicins were found to inhibit transcription of the regulatory rnaIII in S. aureus, which is the effector molecule of the agr QS system (See Section 2.2 and Supplementary Information). The ngercheumicins were isolated as white solids with the respective exact masses (HR-ESI-TOF) and molecular formulae: Ngercheumicin F (m/z: [M + H]+ 853.5685, calculated for C43H77N6O11 as 853.5650), ngercheumicin G (m/z: [M + H]+ 855.5907, calculated for C43H79N6O11 as 855.5807), ngercheumicin H (m/z: [M + H]+ 881.6033, calculated for C48H81N6O11 as 881.5963), and ngercheumicin I (m/z: [M + H]+ 883.6255, calculated for C48H83N6O11 as 883.6120).

Analysis of 1D and 2D nuclear magnetic resonance (NMR) spectroscopic data obtained for the four compounds characterized the structures as cyclodepsipeptides consisting of six amino acids and a 3-hydroxy fatty acid (six NH signals and seven carbonyl resonances) (Figure 1). All four ngercheumicins were elucidated as having identical amino acid sequences, consisting of three leucines, two threonines, and one serine, as established by DQF-COSY, gHSQC, gHMBC, gH2BC, and TOCSY 2D NMR spectroscopic analyses. The closure of the macrocyclic ring through an ester linkage between the C-terminus and the hydroxyl group in one of the threonine residues was verified by the low field chemical shift of the β-proton of Thr2 (H26, Table 1) and a HMBC correlation between H26 and the carbonyl (C1) of the C-terminal Leu1 residue (for HMBC and H2BC correlations, see Supplementary Information). This formed a 16-membered macrocycle with an exocyclic chain continuing in the N-terminal direction from Thr2, the chain constituting a Leu residue and a 3-hydroxy fatty acid (Figure 1). The structural difference between the four analogues was found in the length and saturation of the unbranched fatty acid chain. Ngercheumicin F and H each have one double bond in the 3-hydroxy fatty acid chain, whereas the fatty acids in ngercheumicin G and I are fully saturated, but with the same lengths as F and H, respectively. The previously isolated ngercheumicin A and B have a similar relationship, but with 12-carbon long fatty acid chains [8], whereas ngercheumicin F and H have fatty acid chains with 14 and 16 carbon atoms, respectively. The longer chains result in increasing overlap of 1H and 13C resonances in the aliphatic regions on both sides of the double bond. By thorough examination of HMBC and H2BC correlations in the chains, the structures of ngercheumicin F and H were elucidated and assigned as shown in Figure 1 and Table 1, respectively. The chemical shifts of ngercheumicin H were virtually symmetrical around the double bond because of the long chain, whereas there was a slight shift in ngercheumicin F, which has a shorter alkyl chain. The position of the double bond in ngercheumicin H was tentatively assigned based on the NMR data, as the data left an ambiguity of a CH2-group between position 41 and 47 due to the high degree of
symmetry and the overlapping resonances that 2D NMR was unable to resolve. However, assuming a correct assignment, a structural pattern emerged where ngercheumicin A, F, and H had their double bonds positioned seven carbon atoms from the end of the fatty acid chain, also known as n-7 fatty acids. Counting from the peptide-end of the fatty acid chain, the double bonds in ngercheumicin F and H are thus situated further into the fatty acid chains.

**Figure 1.** Structures of ngercheumicin A, B, F, G, H, and I, where ngercheumicin A, F, and H have an unsaturation (cis) in the unbranched alkyl chain, R.

![Structures of ngercheumicins](image)

**Table 1.** $^1$H and $^{13}$C NMR spectroscopic data (800 MHz, DMSO-$d_6$) for ngercheumicins F–I. More elaborate NMR tables can be found in the Supplementary Information.

<table>
<thead>
<tr>
<th>Position, Type</th>
<th>Ngercheumicin F</th>
<th>Ngercheumicin G</th>
<th>Ngercheumicin H</th>
<th>Ngercheumicin I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu$^1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1—CO</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2—CH$_a$</td>
<td>4.37</td>
<td>51.1</td>
<td>4.37</td>
<td>51.1</td>
</tr>
<tr>
<td>3—CH$_b$</td>
<td>1.62, 1.55</td>
<td>40.4</td>
<td>1.61, 1.55</td>
<td>40.4</td>
</tr>
<tr>
<td>4—CH$_t$</td>
<td>1.49</td>
<td>24.5</td>
<td>1.49</td>
<td>24.5</td>
</tr>
<tr>
<td>5—CH$_{a,1}$</td>
<td>0.88</td>
<td>22.3</td>
<td>0.88</td>
<td>22.3</td>
</tr>
<tr>
<td>6—CH$_{a,2}$</td>
<td>0.85</td>
<td>22.5</td>
<td>0.85</td>
<td>22.5</td>
</tr>
<tr>
<td>7—NH</td>
<td>7.73</td>
<td>–</td>
<td>7.73</td>
<td>–</td>
</tr>
<tr>
<td>Leu$^2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8—CO</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>9—CH$_a$</td>
<td>4.26</td>
<td>50.4</td>
<td>4.26</td>
<td>50.4</td>
</tr>
<tr>
<td>10—CH$_b$</td>
<td>1.49</td>
<td>38.8</td>
<td>1.49</td>
<td>38.8</td>
</tr>
<tr>
<td>11—CH$_t$</td>
<td>1.53</td>
<td>24.2</td>
<td>1.53</td>
<td>24.2</td>
</tr>
<tr>
<td>12—CH$_{a,1}$</td>
<td>0.87</td>
<td>22.9</td>
<td>0.87</td>
<td>22.9</td>
</tr>
<tr>
<td>13—CH$_{a,2}$</td>
<td>0.80</td>
<td>21.4</td>
<td>0.80</td>
<td>21.4</td>
</tr>
<tr>
<td></td>
<td>14—NH</td>
<td>8.12</td>
<td>8.11</td>
<td>8.11</td>
</tr>
<tr>
<td>----------------</td>
<td>--------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>L-allo-Thr(^1)</td>
<td>15—CO</td>
<td>–</td>
<td>170.7</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>16—CH(_3)</td>
<td>3.86</td>
<td>59.9</td>
<td>3.86</td>
</tr>
<tr>
<td></td>
<td>17—CH(_2)</td>
<td>3.88</td>
<td>65.1</td>
<td>3.88</td>
</tr>
<tr>
<td></td>
<td>17—OH</td>
<td>4.70</td>
<td>–</td>
<td>4.69</td>
</tr>
<tr>
<td></td>
<td>18—CH(_3)</td>
<td>1.03</td>
<td>20.0</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>19—NH</td>
<td>8.09</td>
<td>–</td>
<td>8.09</td>
</tr>
<tr>
<td>d-Ser</td>
<td>20—CO</td>
<td>–</td>
<td>170.0</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>21—CH(_3)</td>
<td>4.38</td>
<td>54.5</td>
<td>4.38</td>
</tr>
<tr>
<td></td>
<td>22—CH(_2)</td>
<td>3.54</td>
<td>61.6</td>
<td>3.54</td>
</tr>
<tr>
<td></td>
<td>22—OH</td>
<td>4.86</td>
<td>–</td>
<td>4.86</td>
</tr>
<tr>
<td></td>
<td>23—NH</td>
<td>7.53</td>
<td>–</td>
<td>7.53</td>
</tr>
<tr>
<td>L-allo-Thr(^2)</td>
<td>24—CO</td>
<td>–</td>
<td>168.4</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>25—CH(_3)</td>
<td>4.40</td>
<td>56.2</td>
<td>4.39</td>
</tr>
<tr>
<td></td>
<td>26—CH(_2)</td>
<td>5.35</td>
<td>69.6</td>
<td>5.34</td>
</tr>
<tr>
<td></td>
<td>27—CH(_3)</td>
<td>1.10</td>
<td>17.0</td>
<td>1.10</td>
</tr>
<tr>
<td>Leu(^3)</td>
<td>29—CO</td>
<td>–</td>
<td>173.1</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>30—CH(_3)</td>
<td>4.37</td>
<td>51.7</td>
<td>4.37</td>
</tr>
<tr>
<td></td>
<td>31—CH(_2)</td>
<td>1.48</td>
<td>40.3</td>
<td>1.48</td>
</tr>
<tr>
<td></td>
<td>32—CH(_3)</td>
<td>1.62</td>
<td>24.2</td>
<td>1.62</td>
</tr>
<tr>
<td></td>
<td>33—CH(_3)</td>
<td>0.92</td>
<td>22.6</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>34—CH(_2)</td>
<td>0.87</td>
<td>21.9</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>35—NH</td>
<td>8.20</td>
<td>–</td>
<td>8.20</td>
</tr>
<tr>
<td>Fatty acid</td>
<td>36—CO</td>
<td>–</td>
<td>171.8</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>~2.24</td>
<td>43.3</td>
<td>2.24</td>
</tr>
<tr>
<td></td>
<td>38</td>
<td>3.78</td>
<td>67.4</td>
<td>3.77</td>
</tr>
<tr>
<td></td>
<td>38—OH</td>
<td>4.62</td>
<td>–</td>
<td>4.60</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>1.34</td>
<td>36.2</td>
<td>1.31</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>1.42</td>
<td>25.1</td>
<td>1.34</td>
</tr>
<tr>
<td></td>
<td>41</td>
<td>1.96</td>
<td>26.6</td>
<td>1.23</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>5.31</td>
<td>129.6</td>
<td>~1.2</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>5.31</td>
<td>129.6</td>
<td>~1.2</td>
</tr>
<tr>
<td></td>
<td>44</td>
<td>1.96</td>
<td>26.5</td>
<td>~1.2</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>1.28</td>
<td>29.0</td>
<td>~1.2</td>
</tr>
<tr>
<td></td>
<td>46</td>
<td>1.24</td>
<td>28.3</td>
<td>1.21</td>
</tr>
<tr>
<td></td>
<td>47</td>
<td>1.22</td>
<td>31.1</td>
<td>1.22</td>
</tr>
<tr>
<td></td>
<td>48</td>
<td>1.25</td>
<td>22.1</td>
<td>1.25</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>0.84</td>
<td>13.9</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>1.25</td>
<td>22.0</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>51</td>
<td>0.84</td>
<td>13.9</td>
<td>0.84</td>
</tr>
</tbody>
</table>
The resonances originating from the double bond were very close in chemical shift, leading to severe second order effects in the $^1$H multiplet patterns. Due to the second order spin systems, it was not possible to determine the size of the $J$ coupling constant, however the total span of the multiplets at 5.31 ppm for both F and H was below 15 Hz, and assuming complete symmetry, a trans coupling seemed unlikely. The chemical shifts for the allylic carbons at 26.5–26.6 ppm are also consistent with cis configuration, as allylic carbons in trans fatty acids are about 5 ppm further downfield [13]. Therefore the double bonds in ngercheumicin F and H were assigned as cis.

Attempts to obtain absolute stereochemical assignment of the amino acids were done by Marfey’s method, using existing methods for acid hydrolysis [14] and derivatization with Marfey’s reagent [15]. Complete stereochemical assignments were not obtained as Marfey’s method revealed the presence of both L- and D-Leu, and unambiguous resolution of the hydroxyamino acids Ser and Thr is known to be challenging [16]. Pure enantiomers of Leu, Ser and Thr were used to synthesise single diastereomers with Marfey’s reagent for comparison with the Marfey’s derivatives of the hydrolysed ngercheumicins. This also included allo-Thr. By comparison to the pure amino acid derivatives, L-Ser, D-Thr and D-allo-Thr were dismissed. Therefore the configuration of the Ser residue was firmly assigned as D-Ser. Although the peaks of L-Thr, D-Ser, and L-allo-Thr eluted within a narrow spectral window, the elution order together with MS detection verified the presence of L-allo-Thr and no L-Thr. This was supported by the size of the $J$ coupling constant from Thr$^2$ H26 to H25 and H27. The absolute configuration of the three Leu residues was ambiguous, however both L- and D-Leu were present. Due to the minute amounts available, (0.5–1.1 mg of each analogue) the configuration of the 3-hydroxy fatty acid was not determined.

Organic synthesis could be a solution to the supply problem of these peptides, as synthesis of cyclic peptides is often a relatively straightforward procedure. Many cyclic peptides from marine microbial sources contain non-proteinogenic and D-amino acids as well as polyketide-derived structural motifs or fatty acids [17], which is also the case for the solonamides and the ngercheumicins. Undoubtedly, these traits complicate synthesis of both natural products and analogues. However, a study by Molhoek et al. showed that cyclization of a peptide antibiotic and substitution of L- with D-amino acids improved stability towards bacterial proteases (including those in S. aureus) and decreased cytotoxicity while retaining the antibacterial activity [18]. The ngercheumicin macrocycle is closed through an ester bond between the C-terminus and the side chain of a Thr residue. This is a common trait in depsipeptides, and e.g., the cyclodepsipeptide plitidepsin [2], which is undergoing clinical trials for treatment of several cancers, has this feature in common with the ngercheumicins.

2.2. Ngercheumicins Interfere with agr

Ngercheumicin F, G, H, and I were examined in the S. aureus lacZ reporter assays described by Nielsen et al. [19] monitoring transcriptional activity of the hla, spa, and P3 (rnaIII) promortors. All four ngercheumicins increased transcription of spa and reduced expression of hla and rnaIII, compared to a solvent control (see Supplementary Information). The inverse effect of the ngercheumicins on spa expression compared to rnaIII and hla indicates that the ngercheumicins interfere with agr activation.
To confirm these initial results, cultures of a highly virulent, community-acquired strain of *S. aureus* USA300 were exposed to ngercheumicins F–I and RNAIII expression was monitored in exponential and stationary growth phase by Northern blot analysis. Data confirmed that ngercheumicins reduced expression of *rnaIII* and thus interfere with *agr* in CA-MRSA strain USA300. At the ngercheumicin concentrations used in the culture samples for Northern blotting, there was a minor down regulation of *rnaIII* by ngercheumicin G and H at 5 μg mL\(^{-1}\) (see Supplementary Information). At 20 μg mL\(^{-1}\) (23 μM, see Figure 2), the *rnaIII* inhibiting effect was more pronounced for ngercheumicin G and H, but also F showed activity here and appeared to have the highest *rnaIII* inhibiting activity. Ngercheumicin I did not show any significant effect at either 5 or 20 μg mL\(^{-1}\).

**Figure 2.** Northern blot of ngercheumicin F, G, H and I treated *S. aureus* USA300 wt (FPR3757) cells in stationary phase (OD\(_{600}\) = 3) at ngercheumicin concentration of 20 μg mL\(^{-1}\).

The ngercheumicins share structural traits with the AIPs of *S. aureus*. *S. aureus* strains express one of a least four variants of AIPs and each variant induces *agr* expression in strains of the same type but repress *agr* expression in strains of other types [20]. Interestingly, the ngercheumicins resemble the type II and III AIPs, whereas the solonamides resemble type I and II the most [21]. All have 16-membered macrocyclic rings and flexible exocyclic chains (Figure 3). The AIPs have purely peptidic exocyclic chains, whereas the depsipeptides all contain fatty acid chains, but both the AIPs and the cyclodepsipeptides have variations in the length of the exocyclic chains, which perhaps relate to receptor specificity. Looking at the amino acid sequence of the ngercheumicins (Figure 3) starting at the C-terminus, the two Leu residues are also found in AIP-III, and the Thr-Ser sequence closely resembles the Ser-Ser sequence in AIP-II. It should be noted that this comparison is made without taking 3D configuration of the side chains into account, because only the 2D structures are known. It is however clear that part of the backbone structure is highly homologous to the native AIPs of *S. aureus*. 
**Figure 3.** Schematic structures of *S. aureus* autoinducing peptides (AIPs) I–III, solonamide B and ngercheumicin G, with amino acid sequences and type of ring closure. AIP structures are reproduced from [21]. AIP-IV also exists, where the aspartic acid (D) of AIP-I is replaced by a tyrosine residue.

Little is known about peptide signaling in the marine environment and whether other groups of peptide signal molecules exist there. It can be speculated that these depsipeptides interfere with QS pathways present in the marine environment or even act as alternative quorum sensing molecules. Existing QS systems include the N-acyl homoserine lactones in Gram-negative bacteria and the autoinducing peptides in the Gram-positive staphylococci [22]. That ngercheumicins F–I were isolated in low yields as intracellular metabolites, unlike the solonamides that were excreted in large amounts, could mean that they do not act as intercellular signaling molecules in the natural environment and therefore there could be an unexplored biological role for the ngercheumicins. Analyses of co-cultures of this *Photobacterium* with naturally co-occurring or competitive strains could provide further insight into the biological function of these compounds.

3. Experimental Section

3.1. Isolation and Identification of Strain s2753

Bacterial strain s2753 was isolated from a mussel surface collected in the tropical Pacific (9.1° S 156.8° E) during the Danish Galathea 3 expedition [3]. S2753 was assigned to the *Vibrionaceae* and sequence similarity of household genes identified it as a *Photobacterium halotolerans* as previously described [3–5].
3.2. Isolation and Structure Elucidation of Four New Ngercheumicins

S2753 was cultured in 10 L glass fermentors in 5 × 4 L SSS containing 0.4% glucose and 0.3% casamino acids in (25 °C, 72 h, 100 rpm). The liquid culture was centrifuged (15 min, 3500 × g) to isolate the pellet from the broth and Diaion HP20SS (12 g L⁻¹) which was used to extract the bioactive compounds holomycin and solonamides A–B as previously described [4,5]. The pellet was extracted with 1 L 1:9 (v/v) MeOH/EtOAc (25 °C, 24 h, 100 rpm) and filtered off through a Watman 1 filter. The pellet extract was concentrated on a rotary evaporator and absorbed onto 5 g Isolute diol (Biotage, Uppsala, Sweden) for dry loading onto a 50 g SNAP column packed with Isolute diol and eluted on an Isolera automated flash system (Biotage, Uppsala, Sweden) using solvents ranging from heptane, dichloromethane, EtOAc to pure MeOH (30 mL min⁻¹, 72 min). A total of 33 fractions were collected and subjected to LC-UV/MS. Fractions 7 to 10 (dichloromethane/EtOAc) were pooled and absorbed onto 1.5 g Isolute diol and further fractionated on a 10 g diol column run by gravity with heptanes, dichloromethane, EtOAc and MeOH as above. This yielded 10 fractions (again subjected to LC-UV/MS) of which fractions 6 and 7 (10%–30% MeOH in EtOAc) were purified on a Luna II column (5 μm C₁₈, 250 × 10 mm ID, Phenomenex) in a Gilson 322 liquid chromatograph with a 215 liquid handler/injector (BioLab, Risskov, Denmark) going from 70% to 100% aqueous MeCN, 20 mM formic acid, over 10 min followed by 6 min isocratic elution. This yielded 12 fractions of which pure compounds were obtained directly: Ngercheumicin F (0.5 mg), Ngercheumicin G (1.0 mg), Ngercheumicin H (0.5 mg), and Ngercheumicin I (1.1 mg). Selected chromatograms are available in the Supplementary Information.

LC-UV/MS analyses were performed on an Agilent 1100 HPLC system with a diode array detector coupled to an LCT TOF mass spectrometer (Micromass, Manchester, UK) using a Z-spray ESI source. Separation was performed at 40 °C with a Luna II C₁₈ column (50 × 2 mm ID, 3 μm, Phenomenex, Torrance, CA, USA), applying a linear gradient of 15%–100% aqueous MeCN, 20 mM formic acid (LC-MS-grade), over 20 minutes at a flow rate of 0.3 mL/min. MS experiments were performed in ESI⁺ with a data acquisition range of m/z 100–2000. Accurate masses of ammonium adducts were measured for Ngercheumicin F (m/z: [M + NH₄]⁺ 870.5979, calculated for C₄₃H₈₀N₂O₁₁ as 870.5916), ngercheumicin G (m/z: [M + NH₄]⁺ 872.6206, calculated for C₄₃H₈₂N₂O₁₁ as 872.6072), ngercheumicin H (m/z: [M + NH₄]⁺ 898.6359, calculated for C₄₅H₈₄N₂O₁₁ as 898.6229), and ngercheumicin I (m/z: [M + NH₄]⁺ 900.6450, calculated for C₄₅H₈₆N₂O₁₁ as 900.6385). The [M + H]⁺ adducts were reported in Section 2.1.

To solve the absolute configuration of the amino acids Marfey’s method was applied: 100 μg of each cyclospesipeptide was subjected to acid hydrolysis (200 μL 6 M HCl, 110 °C, 20 h), redissolved in 50 μL water and added 20 μL 1 M aqueous NaHCO₃, then derivatised with 100 μL 1% w/v Marfey’s reagent in acetone (1-fluoro-2,4-dinitrophenyl-5-L-alanine amide, FDAA, Sigma-Aldrich, St. Louis, MO, USA) at 40 °C for 1 h as described by Bonnard et al. [15]. The reaction mixtures were neutralised with 10 μL 2 M aqueous HCl and diluted with 820 μL MeOH. Pure amino acid standards were derivatised by the same procedure using 50 μL 50 mM aqueous amino acid solution. Ultra-high performance liquid chromatography-diode array (UHPLC-DAD) separation and detection of the amino acid derivatives was done on a Dionex RSLC Ultimate 3000 (Dionex, Sunnyvale, CA, USA) equipped with a diode array detector. The separation was done in a Kinetex C₁₈ column (150 × 2.10 mm,
2.6 µm, Phenomenex) at 60 °C with a flow rate of 0.8 mL min⁻¹ using two different linear gradient methods. Method A included all L- and D-amino acids in the structure, whereas method B was run with shallow gradient and included L- and D-allo-Thr but not L- and D-Leu. This was an attempt to distinguish derivatives of D-Ser, L-Thr and L-allo-Thr which had very similar chromatographic properties on the column.

Method A: From 8% to 15% aqueous MeCN, 0.65 mM TFA, over 22 min followed by an increase from 15% to 100% for 8.5 min. Retention times for the FDAA-amino acid derivatives were: L-Ser (4.21 min), L-Thr (5.46 min), D-Ser (5.56 min), D-Thr (12.38 min), L-Leu (25.23 min), D-Leu (26.00 min). Unreacted FDAA eluted at 11.0 min.

Method B: From 8% to 10% aqueous MeCN, 0.65 mM TFA, over 25 min followed by an increase from 10% to 100% over 5.5 min. Retention times for the FDAA-amino acid derivatives were: L-Ser (4.46 min), L-Thr (5.90 min), D-Ser (5.97 min), L-allo-Thr (6.15 min), D-allo-Thr (10.28 min), D-Thr (15.82 min). Unreacted FDAA eluted at 10.4 min.

NMR spectra were recorded on a Bruker Avance 800 MHz spectrometer equipped with a 5 mm TCI Cryoprobe using standard pulse sequences. The NMR data used for the structural assignments were acquired in DMSO-\(d_6\) (\(\delta_H\) 2.49 ppm and \(\delta_C\) 39.5 ppm). \(^1\)H and \(^{13}\)C NMR spectra are available in the Supplementary Information.

3.3. Antivirulence Activity Testing and Northern Blotting

The S. aureus lacZ reporter assays were performed as described by Nielsen et al. [19]. Ngercheumicin F, G, H, and I were dissolved in DMSO, and DMSO and H_2O was included as negative controls in the assay. Pictures were taken after 11, 13 and 34 h of incubation for the rnaIII-, hla- and spa-reporter strains respectively (Supplementary Information).

RNA for Northern blotting was purified from USA300 (FPR3757) samples from cultures grown in 100 mL Erlenmeyer flasks containing 10 mL Tryptone Soya Broth (TSB, Oxoid, Greve, Denmark) shaking vigorously (200 rpm) in a water bath at 37 °C. Ngercheumicins (5 µg mL⁻¹ and 20 µg mL⁻¹) and DMSO (control) were added at OD\(_{500}\) = 0.4, and samples were taken at OD\(_{500}\) = 0.8 and 3.0. Northern blotting was performed as previously described [5] using an RNAIII-probe constructed using primer rnaIII forward (5′-GGG GAT CAC AGA GAT GTG ATG-3′), and rnaIII reverse (5′-GGG CAT AGC ACT GAG TCC AAG G-3′)(TAG Copenhagen A/S, Frederiksberg, Denmark).

4. Conclusions

Four new ngercheumicins were isolated on the mg-scale and their structure elucidated; however complete stereochemical assignments were not obtained. Although 20 L of bacterial culture was extracted, low isolated yields restricted the possibilities in both the structure elucidation and in the assessment of biological properties. Ngercheumicins were found to inhibit transcription of the regulatory rnaIII in S. aureus, the effector molecule of the agr QS system. These findings will aid in the future work to understand quorum sensing in bacteria, as more structural knowledge about QS inhibitors is valuable in the design of novel inhibitors. The cyclodepsipeptides isolated from this marine Photobacterium have some resemblance to the AIPs of S. aureus, and it can be speculated as to whether these molecules are a new class of peptide signaling molecules in the marine environment.
Acknowledgments

We gratefully thank the Danish Instrument Center for NMR Spectroscopy of Biological Macromolecules for NMR time on the 800 MHz. Funding from the Danish Council for Strategic Research (DSF) is acknowledged. The present work was carried out as part of the Galathea 3 expedition under the auspices of the Danish Expedition Foundation and this is Galathea 3 contribution P103.

Conflicts of Interest

The authors declare no conflict of interest.

References


© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).
Supplementary Information

Figure S1. HPLC-UV chromatogram from final RP purification step (Gilson).
Figure S2. LC-UV-MS diode array chromatograms for ngercheuminic F–I.
Figure S3. Selected regions of ESI+ mass spectra for purified ngercheuminic F–I.
Figure S4. 1H NMR spectrum for ngercheuminic F.
Figure S5. DQF-COSY for ngercheuminic F.
Figure S6. Multiplicity edited gHSQC for ngercheuminic F.
Figure S7. gHMBC for ngercheuminic F.
Figure S8. 1H NMR spectrum for ngercheuminic G.
Figure S9. 13C NMR spectrum for ngercheuminic G.
Figure S10. DQF-COSY for ngercheuminic G.
Figure S11. Multiplicity edited gHSQC for ngercheuminic G.
Figure S12. gHMBC for ngercheuminic G.
Figure S13. 1H NMR spectrum for ngercheuminic H.
Figure S14. 13C NMR spectrum for ngercheuminic H.
Figure S15. DQF-COSY for ngercheuminic H.
Figure S16. Multiplicity edited gHSQC for ngercheuminic H.
Figure S17. gHMBC for ngercheuminic H.
Figure S18. 1H NMR spectrum for ngercheuminic I.
Figure S19. 13C NMR spectrum for ngercheuminic I.
Figure S20. DQF-COSY for ngercheuminic I.
Figure S21. Multiplicity edited gHSQC for ngercheuminic I.
Figure S22. gHMBC for ngercheuminic I.
Figure S23. Key HMBC and H2BC correlations in the fatty acid chain of Ngercheuminic F.
Figure S24. Northern blot results for ngercheuminic F–I.
Figure S25. Plate assay results in the colorimetric S. aureus lacZ reporter assay against S. aureus 8325-4.

Table S1. NMR table for ngercheuminic F.
Table S2. NMR table for Ngercheuminic G.
Table S3. NMR table for Ngercheuminic H.
Table S4. NMR table for Ngercheuminic I.
Figure S1. Detector response from HPLC purification of ngercheumicins, where fractions 6, 8, 9 and 11 gave ngercheumicin F, G, H and I, respectively. Fraction size is not shown, as these are pooled fractions from the automatic fractionation. Fraction 7 and 10 contains isomers of ngercheumicin F and H.
Figure S2. Diode array detected (200–700 nm) chromatograms for purified samples of ngercheumicin F, G, H and I (from the top).
**Figure S3.** Selected region of the total ion chromatograms (TIC) of ngercheumaticin F, G, H and I (from the top), showing \([\text{M + H}]^+\), \([\text{M + NH}_4]^+\) and weak \([\text{M + Na}]^+\) adducts. Outside the range are also \([2\text{M + NH}_4]^+\) at \(m/z\) 1723–1784.
Figure S4. $^1$H spectrum for ngercheumicin F.

Figure S5. DQF-COSY for ngercheumicin F.
Figure S6. Multiplicity edited gHSQC for ngercheuminicin F.

Figure S7. gHMBC for ngercheuminicin F.
Figure S8. $^1$H spectrum for ngercheumicin G.

Figure S9. $^{13}$C spectrum for ngercheumicin G.
Figure S10. DQF-COSY for ngercheumcin G.

Figure S11. Multiplicity edited gHSQC for ngercheumcin G.
Figure S12. gHMBC for ngercheumicin G.

Figure S13. $^1$H spectrum for ngercheumicin H.
Figure S14. $^{13}$C spectrum for ngercheumicin H.

Figure S15. DQF-COSY for ngercheumicin H.
Figure S16. Multiplicity edited gHSQC for ngercheumicin H.

Figure S17. gHMBC for ngercheumicin H.
Figure S18. $^1$H spectrum for ngercheumicin I.

Figure S19. $^{13}$C spectrum for ngercheumicin I.
Figure S20. DQF-COSY for ngercheumatic I.

Figure S21. Multiplicity edited gHSQC for ngercheumatic I.
Figure S22. gHMBC for ngercheumicin I.

Figure S23. Key HMBC and H2BC correlations in the fatty acid chain of Ngercheumicin F.
**Figure S24.** Northern blot results for ngercheумicin F, G, H and I in DMSO at 5 μg/mL (~6 μM) and 20 μg/mL (~23 μM) against *S. aureus* USA300/FPR3757 at OD$_{600}$ = 0.8 and OD$_{600}$ = 3.0. DMSO is used as negative control.
Figure S25. Plate assay results in the colorimetric *S. aureus lacZ* reporter assay with ngercheumicins F, G, H and I in DMSO at 1 mg/mL against *S. aureus* 8325-4. DMSO and water is used as negative controls. Inhibitors of the *agr* locus enhance transcription of *spa* and reduce expression of *hla* and *rnlIII*.

**Table S1.** NMR table for Ngercheumicin F.

<table>
<thead>
<tr>
<th>Position, Type</th>
<th>δH (ppm), #H, Multiplicity, J (Hz)</th>
<th>δC (ppm)</th>
<th>HMBC Connectivity</th>
<th>H2BC Connectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Leu</em>&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1—CO</td>
<td>–</td>
<td>170.8</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>2—CH&lt;sub&gt;a&lt;/sub&gt;</td>
<td>4.37, 1H, m</td>
<td>51.1</td>
<td>1, 3</td>
<td></td>
</tr>
<tr>
<td>3—CH&lt;sub&gt;b&lt;/sub&gt;</td>
<td>1.62, 1H, m</td>
<td>40.4</td>
<td>1, 2, 4, 5, 6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.55, 1H, m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4—CH&lt;sub&gt;γ&lt;/sub&gt;</td>
<td>1.49, 1H, m</td>
<td>24.5</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>5—CH&lt;sub&gt;δ,1&lt;/sub&gt;</td>
<td>0.88, 3H, d, 6.5</td>
<td>22.3</td>
<td>3, 4, 6</td>
<td>4</td>
</tr>
<tr>
<td>6—CH&lt;sub&gt;δ,2&lt;/sub&gt;</td>
<td>0.85, 3H, d, 6.5</td>
<td>22.5</td>
<td>3, 4, 5</td>
<td>4</td>
</tr>
<tr>
<td>7—NH</td>
<td>7.73, 1H, d, 6.5</td>
<td>–</td>
<td>(2), 8</td>
<td>2</td>
</tr>
<tr>
<td><em>Leu</em>&lt;sup&gt;2&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8—CO</td>
<td>–</td>
<td>171.0</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>9—CH&lt;sub&gt;a&lt;/sub&gt;</td>
<td>4.26, 1H, q, 8.0</td>
<td>50.4</td>
<td>8, 10</td>
<td>10</td>
</tr>
<tr>
<td>10—CH&lt;sub&gt;b&lt;/sub&gt;</td>
<td>1.49, 2H, m</td>
<td>38.8</td>
<td>9, 11, 13</td>
<td>9</td>
</tr>
<tr>
<td>11—CH&lt;sub&gt;c&lt;/sub&gt;</td>
<td>1.53, 1H, m</td>
<td>24.2</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>12—CH&lt;sub&gt;δ,1&lt;/sub&gt;</td>
<td>0.87, 3H, d, 6.5</td>
<td>22.9</td>
<td>10, 11, 13</td>
<td>11</td>
</tr>
<tr>
<td>13—CH&lt;sub&gt;δ,2&lt;/sub&gt;</td>
<td>0.80, 3H, d, 6.5</td>
<td>21.4</td>
<td>10, 11, 12</td>
<td>11</td>
</tr>
<tr>
<td>14—NH</td>
<td>8.12, 1H, d, 8.5</td>
<td>–</td>
<td>(9), 15</td>
<td>9</td>
</tr>
</tbody>
</table>
Table S1. Cont.

<table>
<thead>
<tr>
<th>L-\textit{allo-Thr}$^1$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15—CO</td>
<td>–</td>
<td>170.7</td>
<td>–</td>
</tr>
<tr>
<td>16—CH$_a$</td>
<td>3.86, 1H, m</td>
<td>59.9</td>
<td>15, 17</td>
</tr>
<tr>
<td>17—CH$_b$</td>
<td>3.88, 1H, m</td>
<td>65.1</td>
<td>(16)</td>
</tr>
<tr>
<td>17—OH</td>
<td>4.70, 1H, d, 4.6</td>
<td>–</td>
<td>16, 17, 18</td>
</tr>
<tr>
<td>18—CH$_r$</td>
<td>1.03, 3H, d, 6.0</td>
<td>20.0</td>
<td>16, 17</td>
</tr>
<tr>
<td>19—NH</td>
<td>8.09, 1H, d, 6.0</td>
<td>–</td>
<td>(16), 20</td>
</tr>
</tbody>
</table>

\textit{D-Ser}

| 20—CO | – | 170.0 | – |
| 21—CH$_a$ | 4.38, 1H, m | 54.5 | 20, 22, 24 | 22 |
| 22—CH$_b$ | 3.54, 2H, m | 61.6 | 20, (21) | 21 |
| 22—OH | 4.86, 1H, t, 6.0 | – | (21), 22 | 22 |
| 23—NH | 7.53, 1H, d, 8.0 | – | 24 | 21 |

L-\textit{allo-Thr}$^2$

| 24—CO | – | 168.4 | – |
| 25—CH$_a$ | 4.40, 1H, d, 9.0 | 56.2 | 24, 29 |
| 26—CH$_b$ | 5.35, 1H, qd, 6.0, <1 | 69.6 | 1 | 27 |
| 27—CH$_r$ | 1.10, 3H, d, 6.5 | 17.0 | 25, 26 | 26 |
| 28—NH | 8.40, 1H, d, 9.5 | – | 29 | 25 |

\textit{Leu}$^3$

| 29—CO | – | 173.1 | – |
| 30—CH$_a$ | 4.37, 1H, m | 51.7 | 29, 31 | 31 |
| 31—CH$_b$ | 1.48, 2H, m | 40.3 | 29, 30, 32 |
| 32—CH$_r$ | 1.62, 1H, m | 24.2 | 33 |
| 33—CH$_{b,1}$ | 0.92, 3H, d, 6.5 | 22.6 | 31, 32, 34 | 32 |
| 34—CH$_{b,2}$ | 0.87, 3H, d, 6.5 | 21.9 | 31, 32, 33 | 32 |
| 35—NH | 8.20, 1H, d, 9.5 | – | 30, 31, 36 | 30 |

\textit{Fatty acid}

| 36—CO | – | 171.8 | – |
| 37 | ~2.24, 2H, m | 43.3 | 36, 38, (39) | 38 |
| 38 | 3.78, 1H, m | 67.4 | – | 37 |
| 38—OH | 4.62, 1H, d, 5.0 | – | 37, 38, 39 | 38 |
| 39 | 1.34, 1H, m | 36.2 | (38), 40 | (38), 40 |
| 1.30, 1H, m | | | | |
| 40 | 1.42, 1H, m | 25.1 | 41, 42 | 41 |
| 1.28, 1H, m | | | | |
| 41 | 1.96, 2H, m | 26.6 | 43 | (40), 42 |
| 42 | 5.31, 1H, m | 129.6 | 44 | 41, 43 |
| 43 | 5.31, 1H, m | 129.6 | 43 | 42, 44 |
| 44 | 1.96, 2H, m | 26.5 | 42 | 43, 45 |
| 45 | 1.28, 2H, m | 29.0 | – | 44 |
| 46 | 1.24, 2H, m | 28.3 | – | 47 |
| 47 | 1.22, 2H, m | 31.1 | 45 | 46 |
| 48 | 1.25, 2H, m | 22.1 | 47, 49 | 47, 49 |
| 49 | 0.84, 3H, t, 7.0 | 13.9 | 47, 48 | 48 |
Table S2. NMR table for Ngercheumicin G.

<table>
<thead>
<tr>
<th>Position</th>
<th>Type</th>
<th>( \delta_H ) (ppm), #H, Multiplicity, ( J ) (Hz)</th>
<th>( \delta_C ) (ppm)</th>
<th>HMBC Connectivity</th>
<th>H2BC Connectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1—CO</td>
<td>–</td>
<td>170.8</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2—( \text{CH}_2 )</td>
<td>4.37, 1H, m</td>
<td>51.1</td>
<td>1, 3, 4, 8</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3—( \text{CH}_3 )</td>
<td>1.61, 1H, m</td>
<td>40.4</td>
<td>1, 2, 4, 5, 6</td>
<td>2, 4</td>
<td></td>
</tr>
<tr>
<td>4—( \text{CH}_3 )</td>
<td>1.49, 1H, m</td>
<td>24.5</td>
<td>3, 5, 6</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>5—( \text{CH}_3 )</td>
<td>0.88, 3H, d, 6.5</td>
<td>22.3</td>
<td>3, 4, 6</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>6—( \text{CH}_3 )</td>
<td>0.85, 3H, d, 6.5</td>
<td>22.5</td>
<td>3, 4, 5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>7—NH</td>
<td>7.73, 1H, d, 6.5</td>
<td>–</td>
<td>–</td>
<td>1, 2, 3, 8</td>
<td>2</td>
</tr>
<tr>
<td>Leu&lt;sup&gt;2&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8—CO</td>
<td>–</td>
<td>171.0</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9—( \text{CH}_3 )</td>
<td>4.26, 1H, q, 8.0</td>
<td>50.4</td>
<td>8, 10, 11</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>10—( \text{CH}_3 )</td>
<td>1.49, 2H, m</td>
<td>38.8</td>
<td>9, 11, 12, 13</td>
<td>9, 11</td>
<td></td>
</tr>
<tr>
<td>11—( \text{CH}_3 )</td>
<td>1.53, 1H, m</td>
<td>24.2</td>
<td>(9), 10, 12, 13</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>12—( \text{CH}_3 )</td>
<td>0.87, 3H, d, 6.5</td>
<td>22.9</td>
<td>10, 11, 13</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>13—( \text{CH}_3 )</td>
<td>0.80, 3H, d, 6.5</td>
<td>21.4</td>
<td>10, 11, 12</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>14—NH</td>
<td>8.11, 1H, d, 8.5</td>
<td>–</td>
<td>9, (10), 15</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>L-allo-Thr&lt;sup&gt;1&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15—CO</td>
<td>–</td>
<td>170.7</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16—( \text{CH}_3 )</td>
<td>3.86, 1H, m</td>
<td>59.9</td>
<td>15, 17, (18), 20</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>17—( \text{CH}_3 )</td>
<td>3.88, 1H, m</td>
<td>65.1</td>
<td>16</td>
<td>16, 18</td>
<td></td>
</tr>
<tr>
<td>17—OH</td>
<td>4.69, 1H, d, 4.6</td>
<td>–</td>
<td>16, 17, 18</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>18—( \text{CH}_3 )</td>
<td>1.03, 3H, d, 6.0</td>
<td>20.1</td>
<td>16, 17</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>19—NH</td>
<td>8.09, 1H, d, 6.0</td>
<td>–</td>
<td>16, 17, 20</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>D-Ser</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20—CO</td>
<td>–</td>
<td>170.0</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21—( \text{CH}_3 )</td>
<td>4.38, 1H, m</td>
<td>54.6</td>
<td>20, 22, (24)</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>22—( \text{CH}_3 )</td>
<td>3.54, 2H, m</td>
<td>61.5</td>
<td>20, 21</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>22—OH</td>
<td>4.86, 1H, t, 6.0</td>
<td>–</td>
<td>21, 22</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>23—NH</td>
<td>7.53, 1H, d, 8.0</td>
<td>–</td>
<td>(20), 21, 22, 24</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>L-allo-Thr&lt;sup&gt;2&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24—CO</td>
<td>–</td>
<td>168.4</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25—( \text{CH}_3 )</td>
<td>4.39, 1H, d, 9.0</td>
<td>56.2</td>
<td>24, 26, 27, 29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26—( \text{CH}_3 )</td>
<td>5.34, 1H, qd, 6.0, &lt;1</td>
<td>69.6</td>
<td>1, 24, 27</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>27—( \text{CH}_3 )</td>
<td>1.10, 3H, d, 6.5</td>
<td>17.0</td>
<td>25, 26</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>28—NH</td>
<td>8.39, 1H, d, 9.5</td>
<td>–</td>
<td>25, (26), 29</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Leu&lt;sup&gt;3&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29—CO</td>
<td>–</td>
<td>173.2</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30—( \text{CH}_3 )</td>
<td>4.37, 1H, m</td>
<td>51.8</td>
<td>29, 31, 32, 36</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>31—( \text{CH}_3 )</td>
<td>1.48, 2H, m</td>
<td>40.2</td>
<td>29, 30, 32, 33</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>32—( \text{CH}_3 )</td>
<td>1.62, 1H, m</td>
<td>24.2</td>
<td>31, 33</td>
<td>31, 33, 34</td>
<td></td>
</tr>
<tr>
<td>33—( \text{CH}_3 )</td>
<td>0.92, 3H, d, 6.5</td>
<td>22.6</td>
<td>31, 32, 34</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>34—( \text{CH}_3 )</td>
<td>0.87, 3H, d, 6.5</td>
<td>21.9</td>
<td>31, 32, 33</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>35—NH</td>
<td>8.20, 1H, d, 6.5</td>
<td>–</td>
<td>(29), 30, 31, 36</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>

Fatty acid
### Table S2. Cont.

<table>
<thead>
<tr>
<th>Position</th>
<th>Type</th>
<th>δ_H (ppm)</th>
<th>#H</th>
<th>Multiplicity, J (Hz)</th>
<th>δ_C (ppm)</th>
<th>HMBC Connectivity</th>
<th>H2BC Connectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>36—CO</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td></td>
<td>2.24, 1H, dd, 14.0, 6.0</td>
<td></td>
<td></td>
<td>43.4</td>
<td>36, 38, 39</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.22, dd, 14.0, 7.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td></td>
<td>3.77, 1H, m</td>
<td></td>
<td></td>
<td>67.5</td>
<td>36, 40</td>
<td>37, 39</td>
</tr>
<tr>
<td>38—OH</td>
<td></td>
<td>4.60, 1H, d, 5.0</td>
<td></td>
<td></td>
<td></td>
<td>37, 38, 39</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td></td>
<td>1.31, 2H, m</td>
<td></td>
<td></td>
<td>36.6</td>
<td>37, 38, 41</td>
<td>38, 40</td>
</tr>
<tr>
<td>40</td>
<td></td>
<td>1.34, 1H, m</td>
<td></td>
<td></td>
<td>24.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.22, 1H, m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td></td>
<td>1.23, 2H, m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td></td>
<td>~1.2, 2H, m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td></td>
<td>~1.2, 2H, m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td></td>
<td>~1.2, 2H, m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td></td>
<td>~1.2, 2H, m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td></td>
<td>1.21, 2H, m</td>
<td></td>
<td></td>
<td>29.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td></td>
<td>1.22, 2H, m</td>
<td></td>
<td></td>
<td>31.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td></td>
<td>1.25, 2H, m</td>
<td></td>
<td></td>
<td>22.1</td>
<td>47, 49</td>
<td>47, 49</td>
</tr>
<tr>
<td>49</td>
<td></td>
<td>0.84, 3H, t, 7.0</td>
<td></td>
<td></td>
<td>13.9</td>
<td>47, 48</td>
<td>48</td>
</tr>
</tbody>
</table>

### Table S3. NMR table for ngercheumicin H.

<table>
<thead>
<tr>
<th>Position, Type</th>
<th>δ_H (ppm), #H, Multiplicity, J (Hz)</th>
<th>HMBC Connectivity</th>
<th>H2BC Connectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Leu</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1—CO</td>
<td></td>
<td>170.8</td>
<td></td>
</tr>
<tr>
<td>2—CH_α</td>
<td></td>
<td>51.1</td>
<td>1, 3, 4, 8</td>
</tr>
<tr>
<td>3—CH_β</td>
<td></td>
<td>40.4</td>
<td>1, 2, 4, 5, 6</td>
</tr>
<tr>
<td>4—CH_γ</td>
<td></td>
<td>24.5</td>
<td>2, 3, 5, 6</td>
</tr>
<tr>
<td>5—CH_α,1</td>
<td></td>
<td>22.3</td>
<td>3, 4, 6</td>
</tr>
<tr>
<td>6—CH_α,2</td>
<td></td>
<td>22.5</td>
<td>3, 4, 5</td>
</tr>
<tr>
<td>7—NH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.73, 1H, d, 6.5</td>
<td></td>
</tr>
<tr>
<td><strong>Leu</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8—CO</td>
<td></td>
<td>171.0</td>
<td></td>
</tr>
<tr>
<td>9—CH_α</td>
<td></td>
<td>50.4</td>
<td>8, 10, 11</td>
</tr>
<tr>
<td>10—CH_β</td>
<td></td>
<td>38.8</td>
<td>9, 11, 13</td>
</tr>
<tr>
<td>11—CH_γ</td>
<td></td>
<td>24.1</td>
<td>9, 10, 13</td>
</tr>
<tr>
<td>12—CH_α,1</td>
<td></td>
<td>22.9</td>
<td>10, 11, 13</td>
</tr>
<tr>
<td>13—CH_α,2</td>
<td></td>
<td>21.4</td>
<td>10, 11, 12</td>
</tr>
<tr>
<td>14—NH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.11, 1H, d, 8.5</td>
<td></td>
</tr>
<tr>
<td><strong>L-allo-Thr</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15—CO</td>
<td></td>
<td>170.7</td>
<td></td>
</tr>
<tr>
<td>16—CH_α</td>
<td></td>
<td>59.9</td>
<td>15, 17, 18, 20</td>
</tr>
<tr>
<td>17—CH_β</td>
<td></td>
<td>65.1</td>
<td>16</td>
</tr>
<tr>
<td>18—OH</td>
<td></td>
<td></td>
<td>16, 17, 18</td>
</tr>
<tr>
<td>18—CH_γ</td>
<td></td>
<td>20.1</td>
<td>16, 17</td>
</tr>
<tr>
<td>19—NH</td>
<td></td>
<td></td>
<td>16, 17, 20</td>
</tr>
<tr>
<td><strong>D-Ser</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table S3. Cont.

<table>
<thead>
<tr>
<th>L-**allo-**Thr</th>
<th>20—CO</th>
<th>170.0</th>
<th></th>
<th>21—CH(_2)</th>
<th>4.38, 1H, m</th>
<th>54.5</th>
<th>20, 22, 24</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>22—CH(_2)</td>
<td>3.54, 2H, m</td>
<td>61.5</td>
<td>20, 21</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22—OH</td>
<td>4.86, 1H, t, 6.0</td>
<td>–</td>
<td>21, 22</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23—NH</td>
<td>7.53, 1H, d, 8.0</td>
<td>–</td>
<td>(20), 21, 22, 24</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24—CO</td>
<td>–</td>
<td>168.4</td>
<td>–</td>
<td>25—CH(_2)</td>
<td>4.39, 1H, d, 9.0</td>
<td>56.2</td>
<td>24, 26, 27, 29</td>
<td>29</td>
</tr>
<tr>
<td>26—CH(_2)</td>
<td>5.34, 1H, qd, 6.0, &lt;1</td>
<td>69.6</td>
<td>1, 24, 27</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27—CH(_2)</td>
<td>1.10, 3H, d, 6.5</td>
<td>16.9</td>
<td>25, 26</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28—NH</td>
<td>8.39, 1H, d, 9.5</td>
<td>–</td>
<td>25, 26, 29</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29—CO</td>
<td>–</td>
<td>173.1</td>
<td>–</td>
<td>30—CH(_2)</td>
<td>4.37, 1H, m</td>
<td>51.8</td>
<td>29, 31, 32, 36</td>
<td>31</td>
</tr>
<tr>
<td>31—CH(_2)</td>
<td>1.48, 2H, m</td>
<td>40.2</td>
<td>29, 30, 32, 33</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32—CH(_2)</td>
<td>1.62, 1H, m</td>
<td>24.2</td>
<td>30, 31, 34</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33—CH(_3)</td>
<td>0.91, 3H, d, 6.5</td>
<td>22.6</td>
<td>31, 32, 34</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34—CH(_3)</td>
<td>0.87, 3H, d, 6.5</td>
<td>21.9</td>
<td>31, 32, 33</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35—NH</td>
<td>8.20, 1H, d, 9.5</td>
<td>–</td>
<td>(29), 30, 31, 36</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36—CO</td>
<td>–</td>
<td>171.9</td>
<td>–</td>
<td>37</td>
<td>2.25, 1H, dd, 14.0, 6.0</td>
<td>43.4</td>
<td>36, 38, 39</td>
<td>38</td>
</tr>
<tr>
<td>38</td>
<td>3.78, 1H, m</td>
<td>67.5</td>
<td>36, 40</td>
<td>37, 39</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38—OH</td>
<td>4.60, 1H, d, 5.0</td>
<td>–</td>
<td>37, 38, 39</td>
<td>38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>1.32, 2H, m</td>
<td>36.6</td>
<td>38, 41</td>
<td>38, 40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>1.36, 1H, m</td>
<td>24.8</td>
<td>–</td>
<td>39</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>~1.2, 2H, m</td>
<td>~29</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>1.28, 2H, m</td>
<td>~29</td>
<td>44/45</td>
<td>43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>1.97, 2H, m</td>
<td>26.6</td>
<td>41?</td>
<td>42, 44</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>5.31, 1H, m</td>
<td>129.6</td>
<td>43/46</td>
<td>43, 45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>5.31, 1H, m</td>
<td>129.6</td>
<td>43/46</td>
<td>44, 46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>1.97, 2H, m</td>
<td>26.6</td>
<td>48</td>
<td>45, 47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>1.28, 2H, m</td>
<td>~29</td>
<td>44/45</td>
<td>46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>1.25, 2H, m</td>
<td>28.2</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>1.22, 2H, m</td>
<td>31.1</td>
<td>47, (51)</td>
<td>48, 50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>1.25, 2H, m</td>
<td>22.0</td>
<td>48, 49, 51</td>
<td>49, 51</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>0.84, 3H, t, 7.0</td>
<td>13.9</td>
<td>49, 50</td>
<td>50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table S4. NMR table for ngercheumicin I.

<table>
<thead>
<tr>
<th>Position, Type</th>
<th>δ_H (ppm), #H, Multiplicity, J (Hz)</th>
<th>δ_C (ppm)</th>
<th>HMBC Connectivity</th>
<th>H2BC Connectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu₁</td>
<td>1—CO</td>
<td>--</td>
<td>170.8</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>2—CHₐ</td>
<td>4.37, 1H, m</td>
<td>51.1</td>
<td>1, 3, 4, 8</td>
</tr>
<tr>
<td></td>
<td>3—CH₈</td>
<td>1.61, 1H, m</td>
<td>40.4</td>
<td>1, 2, 4, 5, 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.56, 1H, m</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4—CH₇</td>
<td>1.49, 1H, m</td>
<td>24.5</td>
<td>2, 3, 5, 6</td>
</tr>
<tr>
<td></td>
<td>5—CH₈₁</td>
<td>0.88, 3H, d, 6.5</td>
<td>22.3</td>
<td>3, 4, 6</td>
</tr>
<tr>
<td></td>
<td>6—CH₈₂</td>
<td>0.85, 3H, d, 6.5</td>
<td>22.5</td>
<td>3, 4, 5</td>
</tr>
<tr>
<td></td>
<td>7—NH</td>
<td>7.73, 1H, d, 6.5</td>
<td>--</td>
<td>(1), 2, 3, 8</td>
</tr>
<tr>
<td>Leu₁</td>
<td>8—CO</td>
<td>--</td>
<td>171.0</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>9—CHₐ</td>
<td>4.26, 1H, q, 8.0</td>
<td>50.4</td>
<td>8, 10, 11</td>
</tr>
<tr>
<td></td>
<td>10—CH₈</td>
<td>1.49, 2H, m</td>
<td>38.8</td>
<td>9, 11, 13</td>
</tr>
<tr>
<td></td>
<td>11—CH₇</td>
<td>1.54, 1H, m</td>
<td>24.1</td>
<td>9, 10, 13</td>
</tr>
<tr>
<td></td>
<td>12—CH₈₁</td>
<td>0.87, 3H, d, 6.5</td>
<td>22.9</td>
<td>10, 11, 13</td>
</tr>
<tr>
<td></td>
<td>13—CH₈₂</td>
<td>0.80, 3H, d, 6.5</td>
<td>21.4</td>
<td>10, 11, 12</td>
</tr>
<tr>
<td></td>
<td>14—NH</td>
<td>8.11, 1H, d, 8.5</td>
<td>--</td>
<td>9, 10, 15</td>
</tr>
<tr>
<td>L-allo-Thr₁</td>
<td>15—CO</td>
<td>--</td>
<td>170.7</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>16—CHₐ</td>
<td>3.86, 1H, m</td>
<td>59.9</td>
<td>15, 17, 18, 20</td>
</tr>
<tr>
<td></td>
<td>17—CH₈</td>
<td>3.87, 1H, m</td>
<td>65.1</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>17—OH</td>
<td>4.69, 1H, d, 4.6</td>
<td>--</td>
<td>16, 17, 18</td>
</tr>
<tr>
<td></td>
<td>18—CH₇</td>
<td>1.03, 3H, d, 6.0</td>
<td>20.1</td>
<td>16, 17</td>
</tr>
<tr>
<td></td>
<td>19—NH</td>
<td>8.09, 1H, d, 6.0</td>
<td>--</td>
<td>16, 17, 20</td>
</tr>
<tr>
<td>D-Ser</td>
<td>20—CO</td>
<td>--</td>
<td>170.0</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>21—CHₐ</td>
<td>4.38, 1H, m</td>
<td>54.5</td>
<td>20, 22, 24</td>
</tr>
<tr>
<td></td>
<td>22—CH₈</td>
<td>3.54, 2H, m</td>
<td>61.5</td>
<td>20, 21</td>
</tr>
<tr>
<td></td>
<td>22—OH</td>
<td>4.86, 1H, t, 6.0</td>
<td>--</td>
<td>21, 22</td>
</tr>
<tr>
<td></td>
<td>23—NH</td>
<td>7.53, 1H, d, 8.0</td>
<td>--</td>
<td>(20), 21, 22, 24</td>
</tr>
<tr>
<td>L-allo-Thr₁</td>
<td>24—CO</td>
<td>--</td>
<td>168.4</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>25—CHₐ</td>
<td>4.39, 1H, d, 9.0</td>
<td>56.2</td>
<td>24, 26, 27, 29</td>
</tr>
<tr>
<td></td>
<td>26—CH₈</td>
<td>5.35, 1H, qd, 6.0, &lt;1</td>
<td>69.6</td>
<td>1, 24, 27</td>
</tr>
<tr>
<td></td>
<td>27—CH₇</td>
<td>1.10, 3H, d, 6.5</td>
<td>17.0</td>
<td>25, 26</td>
</tr>
<tr>
<td></td>
<td>28—NH</td>
<td>8.39, 1H, d, 9.5</td>
<td>25, 26, 29</td>
<td>25</td>
</tr>
<tr>
<td>Leu₁</td>
<td>29—CO</td>
<td>--</td>
<td>173.2</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>30—CHₐ</td>
<td>4.37, 1H, m</td>
<td>51.8</td>
<td>29, 31, 32, 36</td>
</tr>
<tr>
<td></td>
<td>31—CH₈</td>
<td>1.48, 2H, m</td>
<td>40.2</td>
<td>29, 30, 32, 33</td>
</tr>
<tr>
<td></td>
<td>32—CH₇</td>
<td>1.63, 1H, m</td>
<td>24.2</td>
<td>30, 31, 34</td>
</tr>
<tr>
<td></td>
<td>33—CH₈₁</td>
<td>0.92, 3H, d, 6.5</td>
<td>22.6</td>
<td>31, 32, 34</td>
</tr>
</tbody>
</table>
Table S4. Cont.

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>34—CH$_{2}$</td>
<td>0.87, 3H, d, 6.5</td>
<td>21.9</td>
<td>31, 32, 33</td>
<td>32</td>
</tr>
<tr>
<td>35—NH</td>
<td>8.20, 1H, d, 6.5</td>
<td>—</td>
<td>(29), 30, 31, 36</td>
<td>30</td>
</tr>
<tr>
<td>Fatty acid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36—CO</td>
<td>—</td>
<td>171.9</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>37</td>
<td>2.25, 1H, dd, 14.0, 7.0 2.22, 1H, dd, 14.0, 6.0</td>
<td>43.4</td>
<td>36, 38, 39</td>
<td>38</td>
</tr>
<tr>
<td>38</td>
<td>3.78, 1H, m</td>
<td>67.5</td>
<td>36, 40</td>
<td>37, 39</td>
</tr>
<tr>
<td>38—OH</td>
<td>4.59, 1H, d, 5.0</td>
<td>—</td>
<td>37, 38, 39</td>
<td>38</td>
</tr>
<tr>
<td>39</td>
<td>1.32, 2H, m</td>
<td>36.6</td>
<td>38, 41</td>
<td>38, 40</td>
</tr>
<tr>
<td>40</td>
<td>1.35, 1H, m</td>
<td>24.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.22, 1H, m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>1.23, 2H, m</td>
<td>29.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>~1.2, 2H, m</td>
<td>~29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>~1.2, 2H, m</td>
<td>~29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>~1.2, 2H, m</td>
<td>~29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>~1.2, 2H, m</td>
<td>~29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>~1.2, 2H, m</td>
<td>~29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>~1.2, 2H, m</td>
<td>~29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>~1.2, 2H, m</td>
<td>~29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>1.22, 2H, m</td>
<td>31.3</td>
<td></td>
<td>48, 50</td>
</tr>
<tr>
<td>50</td>
<td>1.25, 2H, m</td>
<td>22.1</td>
<td>48, 49, 51</td>
<td>49, 51</td>
</tr>
<tr>
<td>51</td>
<td>0.84, 3H, t, 7.0</td>
<td>13.9</td>
<td>49, 50</td>
<td>50</td>
</tr>
</tbody>
</table>

Enzyme catalysed production of sialylated human milk oligosaccharides and galactooligosaccharides by *Trypanosoma cruzi* trans-sialidase.

Enzyme catalysed production of sialylated human milk oligosaccharides and galactooligosaccharides by *Trypanosoma cruzi* trans-sialidase

Jesper Holck¹,⁵, Dorte M. Larsen¹,⁵, Malwina Michalak¹, Haiying Li², Louise Kjærluff³, Finn Kirpekar², Charlotte H. Gotfredsen³, Sofia Forssten⁴, Arthur C. Ouwehand⁴, Jørn D. Mikkelsen¹ and Anne S. Meyer¹

¹ Centre for BioProcess Engineering, Department of Chemical and Biochemical Engineering, Building 229, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
² Department of Biochemistry and Molecular Biology, University of Southern Denmark, 5230 Odense M, Denmark
³ Department of Chemistry, Building 201, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
⁴ DuPont Nutrition and Health, Active Nutrition, Sokeritehtaantie 20, 02460 Kantvik, Finland

A *Trypanosoma cruzi* trans-sialidase (E.C. 3.2.1.18) was cloned into *Picha pastoris* and expressed. The pH and temperature optimum of the enzyme was determined as pH 5.7 and 30°C. Using casein glycomacropeptide (CGMP) and lactose as sialyl-donor and acceptor respectively, the optimal donor/acceptor ratio for the trans-sialidase catalysed 3'-sialyllactose production was found to be 1:4. Quantitative amounts of 3'-sialyllactose were produced from CGMP and lactose at a yield of 40 mg/g CGMP. The 3'-sialyllactose obtained exerted a stimulatory effect on selected probiotic strains, including different *Bifidobacterium* strains in single culture fermentations. The trans-sialidase also catalysed the transfer of sialic acid from CGMP to galacto-oligosaccharides (GOS) and to the human milk oligosaccharide (HMO) backbone lacto-N-tetrose (LNT) to produce 3'-sialyl-GOS, including doubly sialylated GOS products, and 3'-sialyl-LNT, respectively. This work thus provides proof of the concept of producing 3'-sialyllactose and potentially other sialylated HMOs as well as sialylated GOS enzymatically by trans-sialidase activity, while at the same time providing valorisation of CGMP, a co-processing product from cheese manufacture.

Introduction

Human milk oligosaccharides (HMOs) are a group of complex glycans that are abundant in human breast milk. With few exceptions, all known HMOs have a lactose core and are elongated via linkage to one or more N-acetylgalactosamine and galactose (N-acetyllactosamine) moieties and can be decorated with sialic acid (N-acetylneuraminic acid) and/or fucose residues [1]. Feeding of breast milk to infants is linked to several beneficial effects, among them promotion of bifidogenic growth [2,3], anti-adhesive effects by blocking pathogens like *Campylobacter jejuni* [4] and immunomodulatory effects [5]. The sialylated oligosaccharides are a particularly interesting subgroup of HMOs. The sialylated compounds constitute 28% of total oligo-saccharides in human milk [6]. 3'-Sialyllactose has been shown to induce growth of various *Bifidobacterium* strains [7], inhibit the binding of cholera toxin *in vitro* [8], and alter the gut microbiota in murine models [9], while sialylated HMOs are moreover believed to be involved in infant brain development [10]. Besides their presence in breast milk, HMOs are in general not found in larger concentrations in nature or in mammalian milk of other species and particularly not in bovine milk, where only trace amounts are found [11]. The scarcity in nature combined with the relatively expensive commercial 3'-sialyllactose (270€/g, www.carbosynth.com) and other sialylated
HMOs and sialylated glycans, encourage the development of new methods for economic large scale production of 3’-sialyllactose, both for further biological studies and commercial value.

Sialylated oligosaccharides can be produced enzymatically by trans-sialidases, sialidases of sialyltransferases. Sialyltransferases can transfer sialic acid from cytidine monophosphate-sialic acid to an acceptor, but this process requires the activated CMP-sialic acid [12]. Sialidases are hydrolytic enzymes that catalyse the removal of sialic acid from a glycan residue. Under some conditions or by mutation of the enzyme, their hydrolytic activity can be repressed and sialic acid is transferred to an acceptor [13]. Trans-sialidases are structurally related to sialidases, but lack the hydrolytic activity completely. The Trypanosoma cruzi trans-sialidase (E.C. 3.2.1.18), an enzyme adapted from the parasite responsible for Chagas disease [14], catalyses the transfer of α(2,3)-linked sialic acid from one β-galactosyl residue to another [15]. Previous studies have shown that sialidases and trans-sialidases can be used to successfully transfer sialic acid from a sialic acid rich biopolymer like whey protein or fetuin to a di- or oligosaccharide, for example lactose, generating sialyllactose [16–18].

In this study, casein glycomacropeptide (CGMP), a dairy process byproduct from cheese production with a sialic acid content of 4–7% (w/w), was used as sialic acid donor for enzymatic synthesis of 3’-sialyllactose. The primary objective of this study thus was to assess the ability of the (cloned) Trypanosoma cruzi enzyme to catalyse the trans-sialidase reaction and to optimise reaction parameters for an efficient production of gram quantities of 3’-sialyllactose from CGMP and lactose. By utilising an industrial, but food grade, byproduct stream such as CGMP (and lactose), the costs can be kept low and value is added to these byproducts from cheese manufacture. In order also to specifically target novelty in relation to enzymatic synthesis of other sialylated glycans, and based on the hypothesis that the trans-sialidase might not be selective with respect to the acceptor, an additional objective was to assess the applicability of the enzyme also to catalyse the production of other HMOs and potentially HMO-like compounds such as sialylated lacto-N-tetraose (LNT) and sialylated galacto-oligosaccharides (GOS) using CGMP as the sialyl-donor.

Materials and methods

Chemicals

Ammonium formate, citric acid, β-o-lactose, sodium acetate, and sodium hydroxide were purchased from Sigma–Aldrich (Steinheim, Germany). 3’- and 6’-Sialyllactose (α(2,3)- and α(2,6)-sialyllactose) standards were from Carbosynth Limited (Compton Bershire, United Kingdom). LNT was purchased from Elicytl SA (Crolles, France). GOS were purchased from Di Giovanni (Bologna, Italy). SepharoseQ anion exchange resin and chromatography columns were obtained from GE Healthcare Biosciences (Uppsala, Sweden). All chemicals used were analytical grade.

Production of the trans-sialidase from T. cruzi (TcTS) in P. pastoris

Vector and strain construction

The T. cruzi trans-sialidase amino acid sequence was obtained from Uniprot, AC Q26966 [19]. For the cloning, the amino acid sequence (Supplementary material, Fig. S1) was back-translated into its gene sequence and the nucleotide sequence was optimised for expression in Pichia pastoris by Geneart® (Regensburg, Germany) using the GeneOptimiser® [20]. The codon-optimised gene sequence, that is, the TcTS gene, was then cloned into the pPICZAlphaC vector (Invitrogen) using XbaI and XhoI restriction sites. The final coding gene sequence consisted of (i) the yeast alpha factor signal sequence, (ii) the TcTS gene, and (iii) the C-terminal c-myc- and 6xhis-tag sequences (Supplementary material, Fig. S2). pPICZAlphaC_TcTS was cloned into Escherichia coli Top10 by electroporation and selected on low salt LB plates (20 g agar, 5 g NaCl, 5 g yeast extract, 10 g tryptone and zoecin 25 mg per liter). Colonies were checked by PCR using primers 5’AOX (5’GACTGGTTCCAATTGACAAGC3’) and 3’AOX (5’GCAAAATG GACTTTGACATCC3’). The target plasmid was purified from 5 mL of ON culture according to the protocol (Zyppo Plasmid Miniprep kit, ZYMO Research, CA, USA). Linearisation of the plasmid was performed using the restriction enzyme MssI according to the manufacturer’s protocol (Fermentas, Germany) and the plasmid was subsequently purified using a GFX column (GE Healthcare, Uppsala, Sweden). The plasmid solution was concentrated at 60°C to a final volume of 10 μl obtaining approximately 10 μg of plasmid. 5 μl was used for transformation. Transformation into P. pastoris X-33 and verification of the positive transformants were performed according to the Invitrogen protocol (Catalog V195-20, version E 2002) using the 3’- and 5’-AOX primers described above. Selection was performed onYPD (10 g yeast extract, 20 g peptone, 20 g dextrose) with 200 μg/ml and 500 μg/ml zoecin, respectively. Selective plates were incubated at room temperature for four days before re-streaking on the same selective media for another four days.

Screening, selection and production of P. pastoris transformants with TcTS activity

14 re-streaked colonies were transferred to 250 mL flasks each containing 40 mL BMGY (13.3 g YNB (yeast nitrogen base with ammonium sulphate without amino acids)), 10 g glycerol, 10 g yeast extract, 20 g peptone and 0.02 g biotin per liter and, buffered to pH 6.0 with 100 mL 1 M potassium phosphate per liter and incubated for 16 hours at 30°C at 200 rpm. The cultures were diluted into 15 mL BMMY (glycerol substituted with a final concentration of 0.5% methanol) and incubated for 24 hours. A final concentration of 0.5% methanol was added once a day for two consecutive days, before the supernatants were harvested by centrifugation and sterile filtered. Concentration was performed using vivaospin20, 30 kDa cut-off concentrators according to the protocol (GE Healthcare, Uppsala, Sweden). Purification of the TcTS enzymes on affinity column chromatography and the SDS silver staining was according to the procedure described by Silva et al. [21]. Dot-blot was performed according to the protocol from Abcam® (Abcam, Cambridge, UK). Hydrolase activity was measured in a reaction mixture containing 100 mM sodium phosphate buffer (pH 7.0), 0.75 mM p-nitrophenol-N-acetyl neuraminic acid (PNP-NeuAc) and enzyme. Reactions were initiated by addition of PNP-NeuAc and colour development was monitored by absorbance at 410 nm at 30°C. Trans-sialidase activity was assayed as described previously [22] but with several modifications: 1 mM CGMP was used as the donor substrate, and 0.5 mM 4-methylumbelliferol-β-o-galactopyranoside (MU-Gal) was used as the acceptor.
**Fermentation of a *P. pastoris* clone expressing the TcTS enzyme**

Based on the evaluation criteria described above, a *P. pastoris* strain (TcTS27.200) expressing the TcTS enzyme in reasonable amounts was selected for larger production in a 5 L Sartorius Biostat Aplus fermentor (Sartorius, Company), principally as described previously [21], with the following modifications: the temperature was 30°C for the glycerol batch and glycerol fed-batch phases, but was decreased to 28°C during the induction of the methanol fed-batch.

**Sialic acid donor**

CGMP (Lactoprodan, CGMP-20) was obtained from Arla Foods Ingredients (VibyJ, Denmark). The sialic acid content was determined to be 5.7% (w/w), corresponding to 0.2 mmol/g dry matter, by a fluorimetric quantification method using 3’-sialyllactose as a standard [23]. The distribution of (2,3) linked and (2,6) linked sialic acid was estimated to be approximately 50:50 based on the literature [24].

**pH/temperature and donor/acceptor ratio optima**

The temperature and pH optimum and the optimal donor/acceptor ratio were determined at the same time via a randomised, quadratic central composite statistically designed experiment. The design contained 24 different combinations of pH, temperature, donor- and acceptor concentrations, and three replications of the centre point. The following factor limits were used: pH 5.5–7.5, temperature 15–35°C, CGMP concentration measured as total sialic acid equivalents 0.18–4.6 mM (corresponding to 0.9–23.4 g/L), and lactose concentration 1–25 mM (corresponding to 0.3–8.5 g/L). The pH was adjusted using a citric acid/disodium-phosphate buffer system, having a final concentration in reactions of 50 mM buffer capacity at each pH value. The enzyme concentration was 1.1 g/Lin all reactions. Temperature was maintained using an Eppendorf thermomixer with agitation at 700 rpm. After three hours reaction time, the reaction mixture was transferred to Vivaspin 500 spin columns (GE Healthcare Biosciences, Uppsala, Sweden), centrifuged at 15,000 × g for 20 min for removal of CGMP, and the 3’-sialyllactose concentration in the permeate was measured as response by HPAEC-PAD as described below.

**pH/temperature optimum improvement and level influence**

The temperature and pH optimum and influence of different levels of the optimal donor/acceptor ratio were determined at the same time via a randomised, quadratic central composite statistically designed experiment. The design included 15 different combinations of pH, temperature, concentration levels, and three replications of the centre point. The factor limits of pH and temperature were adjusted to fit the suggested optimal conditions based on the initial study. The following factor limits were used: pH 4.5–6.5, temperature 25–45°C. Three different concentration levels, all with the same ratio, were chosen based on the optimal ratio determined in the first experiment. Level 1: 1.5 mM CGMP and 8 mM lactose. Level 2: 3.1 mM CGMP and 16 mM lactose. Level 3: 6.2 mM CGMP and 32 mM lactose (7.7, 15.8 and 31.5 g/L CGMP and 2.7, 5.5, 10.9 g/L lactose respectively). Additional reaction parameters were as described above.

**Utilisation of alternative acceptors**

The ability of the (cloned) *T. cruzi* enzyme to utilise LNT or GOS as acceptor was investigated. Reaction conditions were maintained from the lactose experiments which were 6.2 mM CGMP, 32 mM LNT or GOS, 1.1 g/L enzyme, pH 6, temperature 30°C. The average molecular weight of the GOS was estimated to approximately 615 Da based on size distribution, measured by LC–MS. Reaction times were two days and three hours for LNT and GOS respectively, followed by removal of CGMP by spin column treatment as described above.

**Preparative scale production**

The purpose of the preparative scale procedure was to produce gram quantities of pure, well characterised 3’-sialyllactose for functional studies. The process involved enzymatic reaction, purification, structure verification, and functional analysis.

**Enzymatic reaction**

Prior to the enzymatic reaction the aqueous substrate solution containing 7.5 mM CGMP (38.3 g/L) and 41.2 mM lactose (14.1 g/L) was pasteurised in a steam heated, stirred batch reactor at 75°C for 2.5 min. A total volume of 5.8 L of the substrate solution was transferred to a 6 L Sartorius Biostat Aplus fermentor (Goettingen, Germany) equipped with monitors and controls for pH, agitator speed, and temperature. The temperature was adjusted to 30°C and the pH was adjusted to 5.8 using citric acid. The reaction was initiated by adding enzyme to a total of 0.7 g protein/L. The reaction progressed for 22 hours.

**Membrane separation**

CGMP and enzyme were removed from residual lactose and 3’-sialyllactose by crossflow membrane separation using two sets of Vivaflow 200 5 kDa MWCO Hydrosart (Sartorius, Goettingen Germany) in parallel equipped with a Masterflex L/S peristaltic pump. Filtration was maintained at room temperature with 1 bar pressure for 22 hours. After reducing the reaction volume to approximately 30%, 2 L of water was added for diafiltration of the retentate. The diafiltration was terminated after six hours.

**Preparative ion exchange chromatography**

Anion exchange chromatography for purification of 3’-sialyllactose from filtration permeate was performed at 23°C with an ÄKTA purifier 100 workstation equipped with a P-900 pump, P-960 sample pump, UV-900 monitor, pH/C monitor, and Frac-950 fraction collector, all controlled by UNICORN software. A HiScale 50 column with a 400 mL packed bed of SepharoseQ was used for separation. All instruments, column material and software were obtained from GE Healthcare Biosciences (Uppsala, Sweden). The elution of the 3’-sialyllactose was monitored by the UV absorption at 210 nm. Elution was performed at a flow rate of 70 mL/min with water and ammonium formate as eluents. Before injection the column was equilibrated with water for 2 column volumes (CV). After injection, unbound compounds were removed from the column by washing with water for 2.5 CV, followed by isocratic elution with 40 mM ammonium formate for 2.5 CV. After elution the column was regenerated with 400 mM ammonium formate for 4 CV and with water for 2 CV. Fractionation was performed using an elution volume as the fractionation parameter. The collected
fractions containing 3'-sialyllactose were pooled and lyophilized repeatedly five times to remove water and ammonium formate [25]. Residual ammonium formate was monitored using a Formic Acid Assay kit from Megazyme.

3'-Sialyllactose quantification

3'-Sialyllactose concentrations were quantified on HPAEC-PAD (High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection) using a DionexBioLC system (Dionex Corp., Sunnyvale, CA) equipped with a DionexCarboPaTm) PA20 analytical column (3 mm x 150 mm). Lactose and 3'-sialyllactose were isocratically eluted using 16 mm sodium hydroxide and 55 mm sodium acetate. The following pulse potentials and durations were used for detection: $E_1 = 0.1$ V, $t_1 = 400$ ms; $E_2 = -2$ V, $t_2 = 20$ ms; $E_3 = 0.6$ V, $m_3 = 10$ ms; $E_4 = -0.1$ V, $t_4 = 60$ ms. Data were collected and analyzed with the program Chromelon 6.80 SF Build 2361 software (Dionex Corp., Sunnyvale, CA).

Capillary-LC/MS analyses

For liquid chromatography/mass spectrometry (LC/MS) analyses, an Agilent 1100 LC/Agilent 6340 ion trap MS system was used. Oligosaccharides were separated using a Hypercarb porous graphitic carbon (PGC) column (0.32 mm x 150 mm, 5 μm, Thermo Scientific) at 30°C. Samples in 10 mM ammonium bicarbonate were loaded onto the column. Gradient elution was achieved using a binary solvent system consisting of (A) 10 mM ammonium bicarbonate, adjusted to pH 8.5 with ammonium hydroxide, and (B) 100% acetonitrile at a flow rate of 5 μL/min. The gradient was initially at 98.2% (A:B) for 5 min, followed by a linear increase to 42:58% (A:B) at 33 min. This concentration of B was held for 3 min. Subsequently the eluent was returned to 98.2% (A:B) at 40 min and the system was allowed to equilibrate for 10 min prior to the next injection. All solvents used were of the highest HPLC grade. The mass spectrometry was performed in negative ion mode, and was scanned in the range m/z 150-2200, followed by data-dependent MS2 scans of the four most abundant ions in each MS1 scan.

NMR spectroscopy

Nuclear magnetic resonance (NMR) spectra were recorded on 10 mg of 3'-sialyllactose after anion exchange chromatography and lyophilization in 500 μL D_2O (99.9% D, Sigma–Aldrich) at 25°C. Experiments were performed on a Bruker Avance 800 MHz spectrometer with a 5 mm TCI cryoprobe, using standard 1D and 2D pulse sequences. External 1,4-dioxane was used as chemical shift reference (δ 3.75 ppm and δ: 67.4 ppm) and all spectra were processed and analyzed using Topspin 3.0.

Functionality

The ability of the produced 3'-sialyllactose to function as a sole carbon source for seven probiotic strains and a potential pathogen was assessed in a pure culture fermentation test. The tested bacterial strains were Bifidobacterium longum subsp. longum (Danisco Global Culture Collection DGCC 232), B. longum subsp. infantis (DGCC 233), B. longum subsp. infantis (DGCC 1497), B. longum subsp. infantis (DGCC 2238), Lactobacillus acidophilus (NCFM, ATCC 700396), B. longum subsp. longum (BI-05, DGCC 9917), Bifidobacterium lactis (HN019, DGCC 2013), and Clostridium perfringens (ATCC 13124). The strains were pre-cultured from stocks stored at −80°C, in MRS media (de Man, Rogosa and Sharpe medium, Lab M, Bury, United Kingdom) containing glucose for 24 hours at 37°C. After pre-culturing, microbes were inoculated in fresh MRS growth medium containing glucose for 24 hours at 37°C. A cell suspension (1%, v/v) was prepared in MRS medium without glucose and used immediately. Substrates sterilized by UV radiation for 20 s were added to the cell suspension in a final concentration of 1% (w/v). Bacterial growth was measured over 12 hours with the automatic Bioscreen® C system as described by Måkeläinen et al. [26] and measured as the area under the growth curve (OD_600 × min) obtained from the Bioscreen® data. Potato galactan (Megazyme International Ltd., Bray Co., Wicklow, Ireland) was used as a bifidogenic control.

Statistical design and data analysis

For experimental designs the software MODDE Version 7.0.0.1 (UmetricsAB, Umeå, Sweden) was used as an aid to design the experimental template and to evaluate the effects and the interactions by multiple linear regressions. All insignificant interactions were removed from models, unless otherwise stated. For single culture growth experiments one way analyses of variances 95% confidence intervals were compared as Tukey–Kramer intervals calculated from pooled standard deviations (Minitab Statistical Software, Addison-Wesley, Reading, MA, USA).

Results and discussion

The optimal reaction conditions, being pH, temperature, donor and acceptor ratio, and substrate concentration were determined after accomplishing two factorial designs. The first design was performed to determine the preliminary pH/temperature optimum and donor and acceptor ratio. The second design was performed to determine the exact pH/temperature optimum and the influence of substrate concentration.

pH/temperature optimum

From the available literature (Table 1), the exact pH/temperature optimum for the T. cruzi trans-sialidase enzyme is unclear; the pH at which enzyme assays have been run has generally been in the range of 5–9, typically pH 7, whereas assay temperatures have varied from 10 to 37°C, including ‘room temperature’ (Table 1).

The experimental design was made to cover the mid-range of reported assay conditions. With the factors in the first design set at pH 5.5–7.5 and temperature 15–35°C, the different experiments gave yields in the range of 0.18–0.25 mm 3'-sialyllactose.

Multiple linear regression analysis showed that both temperature and pH had a significant influence on the 3'-sialyllactose production (Table 2). A lower pH and a higher temperature were favourable, with an apparent maximum at pH 5.5 and 35°C, but it was neither possible to determine a pH nor a temperature maximum, since both the pH × pH and temperature × temperature quadratic interactions were insignificant. The validity of the model was confirmed by the value of the centre points (0.22 ± 0.01) being close to the coefficient of the constant (Table 2).

In the second factorial design, the experimental frame was adjusted for better determination of pH and temperature optimum. The intervals were adjusted to pH 4.5–6.5 and temperature 25–45°C. The different experiments gave yields in the range of
0.02–0.29 mM 3'-sialyllactose. Multiple linear regression analysis showed that both pH and temperature were significant factors (Table 3). The validity of the model was confirmed by the value of the centre points (0.22 ± 0.02) being close to the coefficient of the constant (Table 3). The pH × pH and temperature × temperature quadratic interactions were also both significant, and pH and temperature optima were identified as pH 5.8 and 29.5°C respectively (Fig. 1). The determined optima were included in the frame of the first factorial design, but the data were inconclusive to establish the optima. The temperature optimum at 29.5°C was in accordance with the temperatures ranging from room temperature to 37°C previously reported for T. cruzi trans-sialidase (Table 1). The finding of the pH optimum at pH 5.8 was notably different from the pH values previously reported, where the T. cruzi trans-sialidase has usually been applied around pH 7 (Table 1).

Optimal donor/acceptor ratio
Apart from pH and temperature being factors in the first factorial design, CGMP (as sialic acid equivalents) and lactose concentration, 0.18–4.60 mM and 1–25 mM, respectively, were also included with yields in the range of 0.02–0.25 mM 3'-sialyllactose. Multiple linear regression analysis showed that the CGMP concentration and the CGMP quadratic interaction were significant, with an optimum at 3.5 mM (Table 2) (Fig. 2). The lactose concentration was not found to be significant, but the significance of the quadratic coefficient indicated the presence of an optimal lactose concentration, estimated to 13.9 mM (Fig. 2). These results suggested that the optimal donor/acceptor ratio was approximately 1:4. The tendency towards a higher acceptor concentration than donor concentration does correspond well with previously reported ratios, when the sialic acid donor is a glycoprotein type [16,18,27], where the studied cases tend to have a higher donor concentration. An opposite tendency is observed in cases involving radioactive labeled acceptors, where donor concentrations are 100 times higher than acceptor concentrations [28–30], presumably due to the availability and costs of radioactive labeled materials.

Optimal concentration levels
Like the pH and temperature, a second factorial design regarding the donor and acceptor concentrations was established. While maintaining the optimal donor/acceptor ratio determined in the

### TABLE 1
Donor and acceptor, ratios, pH, temperature and yields for Trypanosoma cruzi trans-sialidase reactions. Assay in yield indicates purpose of reaction is screening

<table>
<thead>
<tr>
<th>Donor</th>
<th>Acceptor</th>
<th>Ratio D:A (w/w%)</th>
<th>pH</th>
<th>Temperature</th>
<th>Yield</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whey</td>
<td>Lactose</td>
<td>1:5</td>
<td>5–9</td>
<td>25°C</td>
<td>25 mg/g DM</td>
<td>[16]</td>
</tr>
<tr>
<td>Fetuin</td>
<td>Lactose</td>
<td>1:4</td>
<td>7.0</td>
<td>30°C</td>
<td>80 mol%</td>
<td>[18]</td>
</tr>
<tr>
<td>Fetuin</td>
<td>Lactose</td>
<td>1:3.3</td>
<td>7.4</td>
<td>37°C</td>
<td>Assay</td>
<td>[27]</td>
</tr>
<tr>
<td>3'-Sialyllactose</td>
<td>MUGal</td>
<td>2:1</td>
<td>7.0</td>
<td>20°C</td>
<td>Assay</td>
<td>[47]</td>
</tr>
<tr>
<td>Fetuin</td>
<td>Lactose derivative</td>
<td>1.6:1</td>
<td>7.0</td>
<td>Room</td>
<td>Assay</td>
<td>[48]</td>
</tr>
<tr>
<td>3'-Sialyllactose</td>
<td>Lactose derivative</td>
<td>1:1</td>
<td>7.0</td>
<td>Room</td>
<td>Assay</td>
<td>[48]</td>
</tr>
<tr>
<td>Sialyl-para-nitrophenol</td>
<td>Branched α and β galactosides</td>
<td>1:6</td>
<td>6.0</td>
<td>30°C</td>
<td>80–90%</td>
<td>[49]</td>
</tr>
<tr>
<td>3'-Sialyllactose</td>
<td>[C14]N-acetyllactosamin</td>
<td>2:1</td>
<td>4.4–8.9, 7.9 optimum</td>
<td>37°C</td>
<td>5–40 mol%</td>
<td>[50]</td>
</tr>
<tr>
<td>3'-Sialyllactose</td>
<td>[C14]lactose</td>
<td>125:1</td>
<td>5–9.7 optimum</td>
<td>Room</td>
<td>Assay</td>
<td>[28]</td>
</tr>
<tr>
<td>3'-Sialyllactose</td>
<td>[C14]lactose</td>
<td>83:1</td>
<td>7.6</td>
<td>25°C</td>
<td>Assay</td>
<td>[30]</td>
</tr>
<tr>
<td>3'-Sialyllactose</td>
<td>[C14]lactose</td>
<td>135:1</td>
<td>7.0</td>
<td>10–29°C</td>
<td>Assay</td>
<td>[29]</td>
</tr>
</tbody>
</table>

a Benzyl β-D-fucopyranosyl(1 → 6)-2-acetamido-2-deoxy-α-D-glucopyranoside.
b 4-Methyl-umbelliferyl-β-D-galactoside.
*c But optimum was found to be 13°C.

### TABLE 2
Multiple linear regression results on the pH/temperature and ratio. The validity of the model was confirmed by the value of the centre points being close to the coefficient of the constant

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>p³</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.2093</td>
</tr>
<tr>
<td>CGMP</td>
<td>0.0653</td>
</tr>
<tr>
<td>Lactose</td>
<td>0.0047</td>
</tr>
<tr>
<td>Temperature</td>
<td>0.0233</td>
</tr>
<tr>
<td>pH</td>
<td>0.0123</td>
</tr>
<tr>
<td>CGMP × CGMP</td>
<td>-0.0676</td>
</tr>
<tr>
<td>Lactose × lactose</td>
<td>-0.0375</td>
</tr>
</tbody>
</table>

Q² = 0.91 and R² = 0.95.

* P < 0.05 indicates significance at the 95% level.

### TABLE 3
Multiple linear regression results on the pH/temperature improvement and level

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>p³</th>
</tr>
</thead>
<tbody>
<tr>
<td>3'-Sialyllactose</td>
<td>0.217</td>
</tr>
<tr>
<td>Level</td>
<td>0.043</td>
</tr>
<tr>
<td>pH</td>
<td>0.030</td>
</tr>
<tr>
<td>Temperature</td>
<td>0.073</td>
</tr>
<tr>
<td>pH × pH</td>
<td>0.052</td>
</tr>
<tr>
<td>Temperature × temperature</td>
<td>0.067</td>
</tr>
</tbody>
</table>

Q² = 0.69 and R² = 0.89.

* P < 0.05 indicates significance at the 95% level.
first factorial design, the impact of total substrate concentration was investigated using three different levels, the initial concentrations from the first design, 50% lower concentration and 100% higher concentration, giving the levels: Level 1: 1.5 mM CGMP and 8 mM lactose. Level 2: 3.1 mM CGMP and 16 mM lactose. Level 3:

6.2 mM CGMP and 32 mM lactose. All with a maintained donor/acceptor ratio. The formation of 3’-sialyllactose went from 0.16 mm to 0.29 mm when substrate concentration was increased from level 1 to level 3 (Fig. 3). Multiple linear regression analysis confirmed that the substrate level was a significant factor (Table 3). Even though an increased substrate level resulted in an increased 3’-sialyllactose formation, the overall yield measured as sialic acid transfer, or conversion, was decreased from 10.5% to 4.7% with an increased substrate concentration from level 1 to level 3 (Fig. 3). This controversy in product formation in relation to conversion yield must be taken into account when large scale production of 3’-sialyllactose or other similar sialylated products is considered. While the high substrate concentration, and hence high product formation is favourable under circumstances where much product as possible is desirable, lower substrate concentrations, and hence a higher conversion yield might be favourable if one or more reactants are scarce or expensive.

**Preparative scale production**

The preparative scale reaction reached an approximate 3’-sialyllactose concentration of 2.4 mm corresponding to 1.54 g/L. 3’-Sialyllactose was purified from the ultrafiltration and dialysis permeates by anionic exchange chromatography and the peak of interest was detected by UV monitoring at 210 nm. Lactose was eluted in pure water and 3’-sialyllactose was eluted in 40 mM ammonium formate. Fractions containing 3’-sialyllactose from repeated runs were collected, analyzed for 3’-sialyllactose content, and pooled. The fractions at around 35–40 min. were rich in 3’-sialyllactose and contained no traces of lactose. Typical injections of 500 mL ultrafiltration permeate with 2.4 mm 3’-sialyllactose resulted in approximately 315 mL total fraction containing 3–3.5 mm pure 3’-sialyllactose resulting in a recovery for the anionic exchange chromatography step at around 80–90%. The column was loaded close to maximum capacity, since the lactose fraction contained only trace amounts of 3’-sialyllactose and increasing the
sample load with 50% did neither result in a notable increase in the 3′-sialyllactose fraction nor increased product concentration. In turn, the amount of 3′-sialyllactose found in the lactose fraction was increased.

Addressing the overall yield based on mass, the final concentration of 3′-sialyllactose was 1.54 g/L, corresponding to the yield of approximately 40.2 mg/g CGMP. This yield exceeded the 3′-sialyllactose yield of 25 mg/g and 6 mg/g reported for mozzarella whey and κ-casein as sialic acid donors in a previously described method using the T. cruzi trans-sialidase as enzymatic catalyst [16]. In a similar reaction scheme comprised of CGMP and lactose, sialidases showing trans-sialidase activity from Arthrobacter ureafaciens and Bifidobacterium infantis were investigated [17]. In three hours the A. ureafaciens trans-sialidase reached a 3′-sialyllactose yield at 5.5 mg/g CGMP similar to the 5.8 mg/g CGMP (0.29 mM, level 3 Fig. 3) obtained in this study after three hours. A drawback of applying a sialidase was an observed reduction in the 3′-sialyllactose concentration caused by hydrolysis when the reaction time exceeded three hours, hence a true trans-sialidase like the T. cruzi trans-sialidase is superior due to the possibility of extended reaction time and a consequently higher product formation.

The results showed that CGMP efficiently can be used as sialic acid donor for production of gram quantities of 3′-sialyllactose. Since CGMP is the naturally sialyl-glycosylated peptide residue which constitutes the soluble part of the bovine milk κ-casein molecule after chymosin-cleavage of the casein producing the clotted cheese mass, the CGMP is an abundantly available food-grade co-processing product from cheese production. Optimisation of the reaction conditions on this substrate is important in order for the process to be relevant for the prospective industrial use of enzymatically produced 3′-sialyllactose and other sialylated HMOs as functional food ingredients. The final concentration of 2.4 mM 3′-sialyllactose compared to the initial CGMP-bound sialic acid of 7.5 mM resulted in a 32% conversion yield. Taking the theoretical distribution of sialic acid linkages in CGMP into consideration (50% 3′ and 50% 6′ [24]) the conversion yield was more likely 64%, based on available 3′ linkages. The excess of acceptor in the reaction mixture was apparently not enough to drive the reaction towards full conversion. This might be due to the relatively large molecular structure of CGMP, where it can be assumed that not every 3′ linked sialic acid is physically available to the trans-sialidase. Clarification demands further investigation of the glycoside structure of CGMP.

Product characterisation

The structure and purity were confirmed by high performance anionic exchange chromatography with pulsed amperometric detection (HPAE-PAD), liquid chromatography with mass spectrometry (LC-MS) and NMR spectroscopy. Initial HPAE-PAD analysis of the final product from ion exchange chromatography showed identical retention time compared to commercial 3′-sialyllactose. Furthermore, the product had no traces of residual lactose (data not shown). The structural configuration of the produced 3′-sialyllactose was identified by LC-MS in comparison with commercial 3′- and 6′-sialyllactose. The LC-MS was performed using a porous graphitic carbon column, which has been shown efficient in separating isomeric structures of oligosaccharides [31,32]. 3′- and 6′-sialyllactose isomers were eluted at different retention times (dual peaks for each standard was due to the mutarotation of sialyllactose in solution [32]) and fully resolved (Fig. 4a). The sialyllactose product was identified as to be 3′-sialyllactose based on retention time and peak pattern (Fig. 4b). The structure was further confirmed by 1H and 13C NMR assignments using standard 1D and 2D NMR experiments. The spectral assignment was in agreement with literature values for 3′-sialyllactose (Fig. 5) [33–35]. The spectral region between 3.5 and 4 ppm holds the majority of the proton resonances for 3′-sialyllactose, which are elaborated in the NMR assignment table in the supplementary material Figs. S3 and S4. The resonance from residual solvent HDO at 4.78 ppm was suppressed by pre-saturation. As the only major impurity the sample contained some ammonium formate, present at 8.44 ppm in the 1H NMR spectrum (outside spectral range (Fig. 5), see supplementary material Fig. S3 and at 171.8 ppm in the 13C NMR spectrum (supplementary material Fig. S3). Apart from presence of ammonium formate in the sample the purity was very high, estimated from 1D and 2D NMR. For full 1H and 13C NMR spectra and assignments of 3′-sialyllactose see the supplementary material Figs. S3 and S4.

In addition to the product being 3′-sialyllactose, it was confirmed that this T. cruzi trans-sialidase only possesses the transferase activity and not the hydrolytic activity reported both for T. cruzi [29] and trans-sialidase enzymes from other organisms like T. rangeli [13], since no free sialic acid was observed throughout the analysis by neither HPEC-PAD, LC–MS nor NMR spectroscopy.

Alternative acceptors

Besides lactose as acceptor, LNT and GOS were investigated. The most abundant complex oligosaccharide in human milk (besides lactose), is the tetramer LNT (Galβ1-3/4GlcNacβ1-3Galβ1-4Glc), and mono- and disialylated isomers of this molecule are the second most predominant sialylated compounds next to sialyllactose [36]. Even though GOS are not HMOs, GOS were chosen as an easily accessible source of substrate that provides various chain lengths and multiple C3 hydroxyl groups that might serve as trans-sialidation (acceptor) sites. The acceptor specificity of the T. cruzi trans-sialidase has been studied previously mainly to understand the mechanism in relation to: (a) T. cruzi induced infection and pathogenesis and how the organism evades host immune response by trans-sialidase-catalysed sialylation of parasite cell surface mucins [37], and (b) enzyme inhibitor development [38,39]. It is thus well established that the T. cruzi trans-sialidase, including the non-virulent recombinant T. cruzi trans-sialidase used in synthesis [16,17,19] and as the basis for the present work, selectively catalyses the transfer of α(2,3)-linked sialyl residues from donor molecules to non-reducing terminal β-galactopyranosyl units of acceptors. It is also known that the T. cruzi trans-sialidase has relatively relaxed acceptor specificity, and that, for example, the open-chain lactose derivatives lactitol and lactobionic acid are good competitive inhibitors and thus good acceptors [38] and that the enzyme also becomes partly inhibited by isomeric di-β-galactosides and β-thiogalactosides (as judged by % enzyme activity remaining in a radiochemical assay) [39]. The T. cruzi trans-sialidase has also been shown to catalyse the selective mono-sialylation of terminal β(1,3)-linked galactose-units when 3′-sialyllactose is used as donor as well as the synthesis of 3′sialyl-N-acetyllactosamine using the non-natural methylumbelliferyl α-sialoside as
donor [40]. However, to our knowledge neither LNT nor GOS have been evaluated as acceptors for T. cruzi trans-sialidase catalysed sialylation in relation to synthesis of HMOs or putative HMO-like prebiotics. In the present study, by running the reactions under similar conditions as for the production of 3′-sialyllactose, it was revealed that both LNT and GOS could serve as acceptors for the T. cruzi trans-sialidase. The presence of sialyl-LNT and sialyl-GOS was confirmed by LC–MS. Only qualitative verifications were performed due to lack of standards. Only single sialylated LNT was found, which is in compliance with the previously described specificity of T. cruzi trans-sialidase towards terminal 3′ linkages [41]. Contrary, the MS/MS data (Fig. 6) on the sialyl-GOS fraction purified by anion exchange chromatography, as described for preparative production, revealed the presence of doubly sialylated products with a GOS backbone of degree of polymerization (DP) 3–6, although only in minor amounts based on relative abundances. The relative abundances of the different degrees of polymerizations in neutral oligosaccharides in the GOS mixture and the resulting sialylated oligosaccharides were compared (Fig. 7). A shift in the distribution towards shorter compounds was observed, based on relatively more of sialylated DP2 and less of sialylated DP4 and DP5 compared to the distribution in GOS, indicating that the T. cruzi trans-sialidase has a higher affinity for shorter acceptors.

**Functionality**

Single culture fermentations were performed as an initial test in order to estimate the capability of 3′-sialyllactose to induce growth of seven common probiotic bacteria B. longum 232, B. longum BI-05, B. infantis 233, B. infantis 1497, B. infantis 2238, B. lactis HN019, and L. acidophilus NCFM and one potential pathogen C. perfringens. PCR screening revealed that the three B. infantis strains all contained the constitutive sialidase Blon_2348 present in the HMO utilisation cluster described by Sela et al. [42], whereas the two B. longum strains did not (data not shown). The growth of the eight strains were measured as OD × min for 12 hours using either minimal MRS medium without glucose, the produced 3′-sialyllactose or galactan as substrate. The produced 3′-sialyllactose was able to induce statistically significant growth of B. longum 232, B. infantis 233, B. infantis 1497 and B. lactis HN019 strains compared to the pure media (Fig. 8). It was expected that all three B. infantis strains would be able to utilise 3′-sialyllactose, due to the presence of the Blon_2348 sialidase encoding gene. Exo-α-sialidase activity (EC 3.2.1.18) has previously been positively demonstrated in B. longum subsp. infantis (ATCC15697 (=JCM1222)) and in two strains of *Bifidobacterium bifidum* (JCM1254 and JCM7004) [43], but it has previously been shown that some B. longum strains fail to grow on neutral HMOs [2] and sialylated HMOs [44] as a sole carbon source, which might explain the lack of induced growth of B. longum BI-05. Contrary, 3′-sialyllactose induced the growth of B. longum 232 despite the lack of the Blon_2348 sialidase. Also B. lactis HN019 was able to grow on 3′-sialyllactose. Potato galactan, which
has previously shown bifidogenic effect [45] was only able to induce growth of B. infantis 233 and B. infantis 1497, along with L. acidophilus NCFM (Fig. 8). The lack of growth for the other strains might be due to the difference of the galactan obtained from Megazyme used in this study and the galactan extracted by Thomassen et al. [45] and the fact that the bifidogenic effect was obtained at species level in a mixed fermentation rather than at the specific strain level.

The case that 3’-sialyllactose was able to stimulate the growth of C. perfringens, was expected based on the previously reported presence of genes responsible for releasing and degradation of sialic acid [46], but the extent of the induction was surprising based on the limited growth by C. perfringens ATCC13124 on neutral HMOs [3].

A common incident for the three strains with no growth induction, along with C. perfringens, was the relatively high background growth on the MRS media without any added carbon source (Fig. 8), which makes it more difficult to determine a significant growth induction by the 3’-sialyllactose and galactan. Using a more suitable minimal medium, as described by Marcobal et al. [3], would be beneficial for detecting small changes and might give a more comprehensive conclusion.

Another aspect that might influence the results is the fact that the individual cell suspensions were inoculated at a fixed inoculation time, and not with respect to reaching the same OD. This could mean that the strains were not in the same growth phase, and hence responded differently to 3’-sialyllactose.

Conclusions
A trans-sialidase from T. cruzi was successfully cloned and expressed in P. pastoris. The pH and temperature optimum was determined along with the optimal donor:acceptor ratio. A process for production of gram quantities of 3’-sialyllactose utilising a dairy side stream as sialic acid donor was established, and the enzyme was also shown to be able to catalyse sialyl-transfer from CGMP to other acceptors, that is, GOS and LNT. The 3’-sialyllactose produced was able to induce the growth of selected Bifidobacterium strains, including some infant gut-related B. longum subsp. infantis strains. The enzymatic method described here provides a proof of concept and can serve as a benchmark for production of gram quantities of 3’-sialyllactose and other sialylated oligosaccharides.
Acknowledgement

We thank the Danish Instrument Center for NMR Spectroscopy of Biological Macromolecules for NMR time on the 800 MHz spectrometer.

References


Appendix A. Supplementary data

Supplementary data related to this article can be found in the online version, at http://dx.doi.org/10.1016/j.nbt.2013.11.006.
Supplementary material

This supplement contains:

Figure S1. Protein sequence of *T. cruzi* transsialidase.

Figure S2. Gene sequence of *T. cruzi* transsialidase (with tags).

Figure S3. $^1$H NMR spectrum of 3’-sialyllactose.

Figure S4. $^{13}$C NMR spectrum of 3’-sialyllactose.
Figure S1. Protein sequence of *T. cruzi* transsialidase (without signal peptide, tags, and stop-codon).

**ATGAGATTTCTCTTCAATTTTTACTGCTGTTTTATTCGCAGCATCCTCCGCAATTAGCTGCTCCAGTCAACACTACAACAGAAGAT**  
**GAAACGGCACAAATTCCGGCTGAAGCTGTCATCGGTTACTCAGATTTAGAAGGGGATTTCGATGTTGCTGTTTTGCCATTTTC**  
**CAACAGCACAAATAACGGGTTATTGTTTATAAATACTACTATTGCCAGCATTGCTGCTAAAGAAGAAGGGGTATCTCTCGAG**  
**AGAGAGAGGCTGAA**

Figure S2. Gene sequence of *T. cruzi* transsialidase. **Signal peptide** (alpha factor signal sequence), *T. cruzi* transsialidase, [Alanine (**GCT**) and Leucine (**CTA**) insert before the c-myct epitope], **c-myct tag** and **histidine tag** and **stop codon**.
Figure S3. $^1$H NMR spectrum of 3'-sialyllactose. $^1$H NMR in D$_2$O: 5.21 (d, $J = 3.7$ Hz, H1α), 4.65 (d, $J = 8.0$ Hz, H1β), 4.52 (d, $J = 7.9$ Hz, H1''), 4.10 (ddd, $J = 3.0$ Hz, 3.3 Hz, 9.9 Hz, H3''), 3.95 (m, H5α), 3.95 (m, H4''), 3.97 (m, H6β), 3.92 (m, H6α), 3.88 (m, H8''), 3.87 (m, H6α), 3.86 (m, H9''), 3.84 (m, H5''), 3.82 (m, H3α), 3.82 (m, H6β), 3.76 (m, H6''), 3.71 (m, H6''), 3.70 (m, H5''), 3.67 (m, H4α), 3.68 (m, H4''), 3.65 (m, H4β), 3.64 (m, H3β), 3.64 (m, H9''), 3.62 (m, H6''), 3.59 (m, H5β), 3.59 (m, H7''), 3.58 (m, H2α), 3.57 (m, H2''), 3.27 (dd, $J = 8.1$ Hz, 9.3 Hz, H2β), 2.75 (dd, $J = 4.7$ Hz, 12.1 Hz, H3''$_{eq}$), 2.02 (s, Ac), 1.79 (dd, $J = 11.7$ Hz, 12.5 Hz, H3''$_{ax}$).
Figure S4. $^{13}$C NMR spectrum of 3'-sialyllactose. $^{13}$C NMR in D$_2$O: 175.8 (Ac), 174.7 (C1”), 103.5 (C1’), 100.6 (C2”), 96.6 (C1β), 92.7 (C1α), 79.1 (C4α), 79.0 (C4β), 76.3 (C3’), 76.0 (C5’), 75.6 (C5β), 75.2 (C3β), 74.6 (C2β), 73.7 (C6”), 72.6 (C8”), 72.2 (C3α), 72.0 (C2α), 70.9 (C5α), 70.2 (C2’), 69.2 (C4’), 68.9 (C7”), 68.3 (C4’), 63.4 (C9”), 61.9 (C6’), 60.9 (C6β), 60.8 (C6α), 52.5 (C5’’), 40.5 (C3’’), 22.9 (Ac).
Paper IV


Biocatalytic production of 3’-sialyllactose by use of a modified sialidase with superior trans-sialidase activity.

Biocatalytic production of 3′-sialyllactose by use of a modified sialidase with superior trans-sialidase activity

Malwina Michalak a, Dorte M. Larsen a, Carsten Jers a, João R.M. Almeida a,1, Martin Willer a,2, Haiying Li b, Finn Kirpekar b, Louise Kjærrulf c, Charlotte H. Gotfredsen e, Rune T. Nordvang a, Anne S. Meyer a,*, Jørn D. Mikkelsen a

a Center for BioProcess Engineering, Department of Chemical and Biochemical Engineering, Building 229, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
b Department of Biochemistry and Molecular Biology, Southern University of Denmark, 5230 Odense, Denmark
c Department of Chemistry, Building 201, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

A R T I C L E   I N F O

Article history:
Received 6 June 2013
Received in revised form 6 October 2013
Accepted 31 October 2013
Available online 9 November 2013

Keywords:
trans-Sialylation
Human milk oligosaccharides
Trypanosoma rangeli
3′-Sialylactose

A B S T R A C T

Casein glycomacropeptide (cGMP) and lactose, which are purified (or semi-purified) components obtained from side streams from dairy industry operations, were used as substrates for enzyme catalyzed production of 3′-sialyllactose, a model case compound for human milk oligosaccharides (HMOs). The enzyme employed was a mutated sialidase, Tr6, derived from Trypanosoma rangeli, and expressed in Pichia pastoris after codon-optimization. The Tr6 contained 6 point mutations and exhibited trans-sialidase activity. The Tr6 trans-sialidase reaction conditions were tuned for maximizing Tr6 catalyzed 3′-sialyllactose production by optimizing pH, temperature, acceptor, and donor concentrations using response surface designs. At the optimum reaction conditions, the Tr6 catalyzed the transfer of sialic acid from cGMP to lactose at high efficiency without substantial hydrolysis of the 3′-sialyllactose product. The robustness of the Tr6 catalyzed reaction was verified at 5 L-scale providing a yield of 3.6 g 3′-sialyllactose at an estimated molar trans-sialylation yield of 50% on the 3′-sialyl in cGMP. Lacto-N-tetraose and lacto-N-fucopentaoses also functioned as acceptor molecules demonstrating the versatility of the Tr6 trans-sialidase for catalyzing sialyl-transfer for generating different HMOs. The data signify the applicability of enzymatic trans-sialylation on dairy side-stream components for production of human milk oligosaccharides.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Human milk oligosaccharides (HMOs) appear to constitute an innate immunologic mechanism by which human milk confers breast-fed infants some level of protection against infections [1]. HMOs constitute the third most abundant component of human milk after lactose and lipids, and levels of up to 8 g/L have been reported [2], mainly during the first days of lactation. The HMOs function as soluble receptors that inhibit pathogens such as Campylobacter jejuni and Escherichia coli K1 from adhering to their target receptors on the mucosal surface of the host gastrointestinal tract, and thereby help reduce the incidence of diarrhea and other diseases in breast-fed infants [1,3]. HMO molecules are also substrates for probiotic gut bacteria, including Bifidobacterium longum subsp. infantis [4] (previously known as Bifidobacterium infantis) and thus seem to support the development of a healthy gut microflora in breast-fed infants by acting as bifidogenic prebiotics [4].

All HMO molecules contain lactose (Galβ1–4Glc) at their reducing end and are built from sequential elongations via linkage to one or more units of N-acetylgalactosamine and galactose (N-acetyllactosamine) and can be decorated with several sialic acid (N-acetyl-neuraminic acid) and fucose residues [1,2,5]. Because of the different backbone lengths, bonds, decorations via various bond types, and different combinations of the basic building blocks many different HMO structures can exist. About 180 different HMO species have thus been identified in a pooled human milk sample from five individuals, and nearly 16% of the total oligosaccharide abundances were found to correspond to siaIylated oligosaccharides [5]. Only a few HMOs are present in bovine milk and only at low concentrations.
Enzymatic synthesis (analytical scale) of sialylated oligosaccharides by use of the trans-sialidase from the pathogenic parasite Trypanosoma cruzi was reported already in 2000 [8]. The T. cruzi trans-sialidase has been studied intensively due to its suggested involvement in the mammalian host cell invasion process and pathogenesis of T. cruzi leading to Chagas disease [9]. Hence, recombinant, catalytic domain forms of this enzyme have been expressed in E. coli [10,11] and it has also been reported that the recombinant enzyme can catalyze the sialylation of lactose derivatives such as lactitol and lactobionic acid using 3′-sialyllactose (α-Neu5Ac-(2-3)-gal-β-(1-4)-glc) as donor [12]. The recombinant version of the T. cruzi enzyme expressed in E. coli has also been used for preparative sialylation of a synthetic pentasaccharide from the mucins of T. cruzi [13] and methods for enriching bovine milk with 3′-sialylactose employing the T. cruzi trans-sialidase and claims for using a sialidase from other organisms have also been patented [14,15]. The full T. cruzi dual-domain trans-sialidase enzyme is considered a virulence factor for the invasive T. cruzi phenotype [16], but the recombinant T. cruzi trans-sialidase used in synthesis would not be virulent, and would moreover be inactivated when heating and separation steps are included in the trans-sialylation process. Nevertheless, due to its origin from a pathogenic organism it may be difficult to obtain approval and consumer acceptance for using the T. cruzi trans-sialidase as an aid for producing functional food ingredients for infant consumption. Regarding catalysis facilitated by sialidases, i.e. sialidase glucosyl hydrolases (neuraminidases) (EC 3.2.1.18), the problem is that the reaction will invariably result in hydrolysis of the target 3′-sialyllactose product.

However, mutants of the sialidase from the non-pathogenic Trypanosoma rangeli have been reported to exhibit low trans-sialidase activity (maximum 11% of the trans-sialidase activity of the T. cruzi trans-sialidase) [17]. We hypothesized that the use of such a sialidase mutant from T. rangeli exhibiting trans-sialidase activity could provide an alternative route for enzymatic, biomimetic in vitro synthesis of HMOs. The present study was undertaken to assess the catalytic ability of such a T. rangeli derived trans-sialidase mutant for (i) enzymatic in vitro production of gram-levels of 3′-sialyllactose, and (ii) enzymatic sialyl-transfer to other HMO structures such as lacto-N-tetraose and lacto-N-fucopentaoses. An additional objective was to utilize abundantly available food-grade co-processing products from the dairy industry, i.e. lactose and casein glycomacropeptide (cGMP), as substrates for the enzymatic reaction to make the process relevant in relation to the prospective use of sialylated HMOs as functional food ingredients.

2. Materials and methods

2.1. Substrates

β-D-lactose and standard of N-acetylneuraminic acid (sialic acid) were purchased from Sigma-Aldrich (Steinheim, Germany). Standards of 3′-sialyllactose and 6′-sialyllactose were purchased from Carbosynth (Compton, United Kingdom). Lacto-N-tetraose (LNT), Lacto-N-neotetraose (LNNt), Lacto-N-fucopentaose I (LNFp I), and Lacto-N-fucopentaose V (LNPf V) were purchased from Elicityl SA (Créteil, France). Casein glycomacropeptide (cGMP) in the form of the commercially available product Lacprodan® cGMP-20 (intended for use in infant formulas and PRU-products), containing a total of 5.7% (w/w), equivalent to 0.2 mmol/g dry matter, of colloidal linked sialic acid was a gift from Ingredients from nature (Aarhus, Denmark). Before use, as a technical precaution in relation to HPAEC-analysis, low molecular weight impurities in the cGMP solution were removed by filtration on a 5 kDa membrane (Sartorius AG, Goettingen, Germany).

2.2. Vector and strain construction

A Trypanosoma rangeli sialidase gene (Genbank accession no. U83180.1) with the following mutations, M96V, A89P, S120Y, C249Y, Q284P and I37L [17] was codon-optimized and synthesized by DNA 2.0 (Menlo Park, CA, USA). The gene was inserted into pPhizocC (Invitrogen, Life Technologies Corp. Carlsbad, CA, USA) between the XhoI and Xhol restriction sites generating a translational fusion to the α-factor signal sequence and C-terminal c-myc and 6His tag. The vector was named pHizocC-Tr6 (the gene sequence is available in supplemental material Figure S1). Transformation and selection of Pichia pastoris X-33 expressing the Tr6 gene were carried out according to the manufacturer’s instructions. Zeocin resistant transformants were analyzed for protein expression by Western blot using a c-myc antibody [18].

2.3. Production of the Tr6 enzyme in P. pastoris, purification and characterization

The enzyme, Tr6, was produced in a 5 L Sartorius Biostat Aplus fermentation of P. pastoris, in a fermentation run principally as described previously [19], with the exception that the methanol fed-batch phase was done at 25 °C. The Tr6 enzyme was purified by Cu²⁺ affinity column chromatography using a CIM® IDA-8mL Tube Monolithic Column (BIA Separations GmbH, Villach, Austria) [18]. The protein concentrations were determined at 280 nm using Genes™-TAKES™ module version 1.09 according to the manufacturer’s instructions (Biotek™ Instruments, Inc.).

The purity of the Tr6 was evaluated by SDS-PAGE (BioRad, CA, US) using Coomassie brilliant blue staining as described previously [18]. To test for N-glycosylation the purified Tr6 was treated with endoglycosidase H (EndoH, Medinova) for 1 h at 37 °C. Proteins treated with and without endoglycosidase H were analyzed by Western blot, using monoclonal anti-poly-histidine–peroxidase antibodies, principally as described previously [18].

2.4. Investigation of conditions for trans-sialylation by statistically designed experiments

M Matrix Edition 7.0 (Umetrics AB, Umeå, Sweden) was used as a tool to design the experimental frame, a quadratic central composite design, and to fit and analyze the data by multiple linear regression analysis. Lactose was used as an acceptor and cGMP as donor of sialic acid. The buffer used in the reactions was 15 mM phosphate-citrate buffer with pH values indicated below. Tr6 was incubated at a concentration of 5.3 mg/mL. Three pH regimes 5.5, 6.5 and 7.5 were tested. Incubation temperatures were 15, 20 and 25 °C. Lactose concentration varied from 0.34 to 351 mM and concentration of sialic acid residues bound in cGMP varied from 0.6 to 4.6 mM in the experimental design frame: lactose and cGMP were solubilized in buffer and pre-incubated at specific temperatures, before the reactions were initiated by addition of enzyme. The biocatalysis process was allowed to proceed for 20 min, and then stopped by heating for 10 min at 90 °C. Subsequently each sample was filtered on a 5 kDa polyethersulfone membrane to remove cGMP (Vivaspin, Sartorius AG, Goettingen, Germany). Concentrations of 3′-sialyllactose and sialic acid were determined by HPAEC and/or LC/MS as described below.

2.5. Time study of trans-sialylation catalyzed by Tr6 expressed in P. pastoris

Tr6 was incubated as described above at two different conditions: 25 °C, pH 5.5 with 36.8 mM lactose and 2.6 mM sialic acid bound in cGMP or with 117 mM lactose and 4.6 mM sialic acid bound in cGMP. The reaction was followed for 100 min.

2.6. Enzymatic production of 3′-sialyllactose

The enzymatic trans-sialylation reaction was accomplished in a stirred reactor vessel in a total reaction volume of 5 L. 117 mM lactose and 4.6 mM sialic acid linked to cGMP was dissolved in 15 mM phosphate–citrate buffer (pH 5.5), and after pre-incubation in the buffer at 25 °C for 10 min, 5.3 mg/L Tr6 was added. The enzymatic catalysis lasted 20 min, then the reaction was terminated by inactivation of the enzyme at 90 °C for 10 min. cGMP was removed from the reaction mixtures by a cross-flow filtration using a 5 kDa cutoff membrane (Sartorius AG, Goettingen, Germany). The permeate containing 3′-sialyllactose, sialic acid, lactose and buffer was collected and freeze-dried. After re-solubilization in water (775 mL) the reaction products were separated by anion exchange chromatography using a HiScale 50/20 column (GE Healthcare) with a packed bed of Sepharose Q FF on an AKTA purifier 100 work station. The elution was monitored at 210 nm. Before injection the column was equilibrated with 2 L of water. After injection of the sample, the column was washed with 1.2 L water and the products 3′-sialyl lactose and sialic acid were then eluted with 1.4L of 40 mM ammonium formate. The products were hypholysed and residual ammonium formate was removed by repeated solubilization and lyophilization.

2.7. Production and purification of sialylated lacto-N-tetraoses and lacto-N-fucopentaoses

The reactions were carried out at 5–100 mL scale employing the same reaction conditions as described for 3′-sialyllactose production except that LNT, LNSt, LNFp
I and LNFP V were used as acceptors instead of lactose. LNT was used at 117 mM, whereas LNnT, LNFP I and LNFP V were used at 25 mM.

2.8. High-performance anion exchange chromatography–pulsed amperometric detection (HPAEC–PAD)

Separation and quantification of 3'-sialyllactose and sialic acid were carried out by HPAEC–PAD analysis using a CarboPac™ PA1 (4 mm × 250 mm) column and a Dionex BioLC system. The eluent system comprised deionised water (A), 0.5 M NaOH (B) and 1 M NaOAc (C) and the elution program was modified from the method described in [20] as follows: For the first 3 min a ratio of 80: 20 (% A:B) was applied, then a linear gradient from 80: 20 % A:B to 60: 20:20 % A:B:C was used during 3–27 min. Strongly retained anions were removed from the column by isocratic elution at 40:20:40 % A:B:C from 27 to 31 min. Subsequently the column was re-equilibrated for 7 min with 80: 20 % A:B.

2.9. Capillary liquid chromatography/mass spectrometry

For liquid chromatography/mass spectrometry (LC/MS) analyses, an Agilent 1100 LC/Agilent 6340 ion trap MS system was used. Oligosaccharides were separated using a Hypercarb porous graphitic carbon (PGC) column (0.32 × 150 mm, 5 µm, Thermo Scientific) at 30 °C. Samples (0.5 µL) were loaded onto the column in 10 mM ammonium bicarbonate. Gradient elution was achieved using a binary solvent system consisting of (A) 10 mM ammonium bicarbonate, adjusted to pH 8.5 with ammonium hydroxide, and (B) 100% acetonitrile at a flow rate of 5 µL/min. The gradient was initially at 98:2 (% A:B) for 5 min, followed by a linear increase to 42:58 (% A:B) at 33 min. This concentration of B was held for 3 min. Subsequently the eluent was returned to 98:2 (% A:B) at 40 min and the system was allowed to equilibrate for 10 min prior to the next injection. All solvents used were of the highest HPLC grade. The mass spectrometry was performed in negative ion mode, and was scanned in the range m/z 150–2200 (2 microns, maximum accumulation time of 150 ms, an ion current count of 200,000) followed by data-dependent MS2 scans of the four most abundant ions in each MS1 scan.

2.10. Nuclear magnetic resonance (NMR) spectroscopy

NMR spectra were recorded in D2O (99.9% D, Sigma–Aldrich) at 25 °C on a Varian Unity Inova 500 MHz spectrometer equipped with a 5 mm probe, using standard 1D and 2D pulse sequences. 1,4-dioxane was used as an external chemical shift reference (δc 3.75 ppm and δc 67.4 ppm).

3. Results

3.1. Expression in P. pastoris and characterization of the trans-sialidase Tr6

Since it is generally accepted that codon usage can affect (i.e. decrease) recombinant protein yields [21] a synthetic gene with a codon distribution optimized for P. pastoris expression [22] was synthesized for the production of the T. rangeli sialidase Tr6. The secreted Tr6 polypeptide comprised 663 amino acids including a 23 amino acid C-terminal c-myc and 6xHis tag (Supplementary material, Figure S1) with a theoretical molecular weight of 73 kDa. A positive transformant expressing the Tr6 enzyme was selected and used for enzyme production at a 5 L-scale. This production gave a concentration of Tr6 of 1 g/L. After purification, the enzyme had an apparent molecular weight of about 80 kDa by SDS-PAGE (Fig. 1a), but treatment of the Tr6 enzyme with EndoH followed by Western blot analysis revealed that the expressed protein was glycosylated and the molecular weight of Tr6 was 75 kDa after deglycosylation (Fig. 1b).

3.2. Tuning of the Tr6 reaction conditions for maximizing trans-sialidase activity

The wild type sialidase from T. rangeli and the derived Tr6 mutant enzyme have an inherent high level of sialidase activity, but a relatively low level of trans-sialidase activity [17]. In the present study we wanted to evaluate the relationship between Tr6 sialidase and trans-sialidase activities and define the reaction conditions at which the trans-sialidase activity would be maximal. For production of 3'-sialyllactose, a systematic evaluation of the influence of the reaction factors, notably the reaction temperature, pH, concentration of the sialic acid donor (cGMP) and concentration of the
acceptor (lactose), revealed that increased temperature (from 15 to 25 °C) and decreased pH (from 7.5 to 5.5) gave rise to enhanced yields of 3′-sialyllactose (Supplementary material, Figure S2), but at the reaction optimum at pH 5.5 and 25 °C, at low concentration of the acceptor, lactose (4 mM), only low levels of 3′-sialyllactose were produced, and free sialic acid was released simultaneously (Fig. 2a).

These data signified that the reaction conditions did not sufficiently suppress the sialidase activity of the enzyme. The retention of some sialidase activity was, however, in agreement with previous results reported for T. rangeli mutants expressed in E. coli[17]. Nevertheless, when higher lactose concentrations (117–351 mM) were used, i.e. in effect a higher substrate acceptor:donor ratio, the Tr6 enzyme was able to catalyze the production of high amounts of 3′-sialyllactose by trans-sialidase action without simultaneously catalyzing release of free sialic acid (Fig. 2b). At this high acceptor concentration, and higher acceptor:donor ratio, the specific activity of the Tr6 trans-sialidase activity was determined to be 16.2 μmol min⁻¹ mg⁻¹.

3.3. Enzymatic production, purification and analysis of 3′-sialyllactose at 5 liter-scale

For the enzymatic production of 3′-sialyllactose 5.3 mg/L of Tr6 was incubated for 20 min at the optimal reaction conditions of pH 5.5, 25 °C, using 117 mM lactose (equivalent to 40 g/L) and 4.6 mM sialic acid residues bound in cGMP (equivalent to 26 g cGMP/L). After deactivation of the enzyme, the 3′-sialyllactose product was purified via a two-step down-stream process: the first step was ultrafiltration to remove cGMP and Tr6. The permeate was then collected and after freeze-drying 180 g of dry product was obtained containing 3′-sialyllactose, unreacted lactose, sialic acid and buffer used in the enzymatic process. In the second down-stream processing step, the different compounds were then separated by anion exchange chromatography using ammonium formate as the eluent (Figure S3). The quality of the separation was confirmed by capillary LC/MS and NMR spectroscopy (as discussed below). In order to remove the ammonium formate, the samples were lyophilized six times. The final yield of 3′-sialyllactose was 3.6 g, and the amount of free sialic acid was 133 mg.

Only approximately 50% of the sialic acid residues in the cGMP are bound at the 3′-position, whereas the rest is bound at the 6′-position [23]. Since the Tr6 catalyzed the production of 3′-sialyllactose, it can be assumed that only 3′-bound sialyl, i.e. ~50% of the total sialic acid in cGMP, was available for this enzyme. Therefore, the molar yield of 3′-sialyllactose was ~50%[47.74%] based on the available 3′-sialic acid in cGMP.

The corresponding molar yield of the sialyllactose was 1% on the lactose, and for the free sialic acid the yield was ~2% based on the total sialyl-residues in the cGMP.

The identity of the 3′-sialyllactose product was confirmed by capillary LC/MS and NMR (Fig. 3). 3′-Sialyllactose obtained from the 5 L scale reaction eluted at the same time as the standard 3′-sialyllactose (data not shown), indicating that the structure of the compound produced was 3′-sialyllactose. Both 6′-sialyllactose and 3′-sialyllactose standards were fully resolved confirming the ability of the PGC column to provide good separation of different oligosaccharides and stereoisomers [24]. No 6′-sialyllactose was detected in the product profile, only 3′-sialyllactose.

The 1H and 13C NMR assignments obtained using 1D and 2D NMR (Fig. 3 and supplementary material Figure S4) agreed with literature values for 3′-sialyllactose [25,26]. Besides resonances of 3′-sialyllactose the 1H NMR spectrum also showed a singlet resonance at 8.44 ppm (data not shown), which resulted from residual ammonium formate from the purification buffer.

3.4. Enzymatic production of sialylated lacto-N-tetraose and fuco-N-pentaoses

It was also demonstrated that the Tr6 was able to catalyze the sialylation of lacto-N-tetraose and fuco-N-pentaoses, albeit the
yields of the products were quite low; hence 5–12 mg of sialylated LNT, LNNT, LNFP I and LNFP V were obtained employing the optimal reactions conditions for the Tr6 (pH 5.5, 25 °C), and as high as possible acceptor/donor ratio (see Section 2.7) (Figure S5). LC/MS analysis showed the molecular species of LNT and sialylated-LNT with the mass of sialylated-LNT at 997.3 (Fig. 4), corresponding to the sum of the LNT ion (706.3) and dehydrated sialic acid (291) [27]. The mass of sialylated LNNT was also 997.3, whereas the masses for sialylated LNFP I and LNFP V, respectively, were both determined to be 1143 in agreement with the mass increment of 146 produced by addition of fucose (data not shown). From the LC-MS results it was evident that the sialylated products of LNT, LNNT, LNFP I and LNFP V were predominantly singly sialylated (Figure S5).

4. Discussion

In the present study, the mutated Tr6 sialidase enzyme activity, and expressed at a relatively high yield in P. pastoris (1 g/L). In comparison, previously, only 5 mg/L of the trans-sialidase from T. cruzi was produced in the same expression host, P. pastoris [28]. This gene shared 67% homology to that of the wild type gene from T. rangeli. The relatively high yield of Tr6 in the present study could be due to theodon optimization or the more controlled conditions and higher cell density attained in a fermentor as compared to shake flasks. A high amount of Tr6 was furthermore a prerequisite for a careful examination of the trans-sialidase catalytic activity of Tr6 at a range of process conditions examined in the present study. At the optimal process conditions it was observed, that the Tr6 sialidase from T. rangeli, could be tuned into an efficient trans-sialidase with more than 93% trans-sialidase activity and only 7% residual sialidase activity. These results differed from the trans-sialidase and sialidase activities recorded for a Tr6 enzyme produced in E. coli, which showed relatively lower trans-sialidase activity [17].

whereas high substrate concentrations in general promote high catalytic rates according to Michaelis–Menten kinetics, high acceptor/donor ratios are advantageous for promoting trans-sialidase reactions, and generally used in T. cruzi trans-sialidase catalysis studies [e.g. 8, 29] as well as in β-galactosidase catalyzed galacto-oligosaccharide synthesis reactions [30,31]. When lactose is the acceptor for the trans-sialidase catalysis a high excess of acceptor is not an economic problem, but once more expensive tetra- and penta-saccharides are acceptors, measures such as recycling of the unused acceptor may have to be considered when comparing the application potential. Another issue is enzyme re-use to improve biocatalytic productivity. A time study performed at high lactose acceptor concentration was used in the present work to determine the maximal, specific trans-sialidase activity of Tr6. The specific Tr6 trans-sialidase activity was determined to 16.2 nmol min⁻¹ per µg of pure protein (i.e. 16.2 µmol min⁻¹ mg⁻¹). This specific activity was in essence equivalent to the apparent V_max of 14.7 nmol min⁻¹ µg⁻¹ reported for the similarly mutated T. rangeli sialidase expressed in E. coli with 3'-sialylactose as donor and lactose as acceptor substrate [17], indicating that the maximal conversion rate found, was in fact equal to the maximum catalytic rate of the Tr6 enzyme. It is also interesting to note, that the specific Tr6 trans-sialidase activity (of 16.2 µmol min⁻¹ mg⁻¹) was ~4 times higher than the reported V_max of the T. cruzi trans-sialidase (V_max 3.8 nmol min⁻¹ µg⁻¹) measured using 3'-sialylactose as the donor and lactose as the acceptor [17], and ~40,000 times higher than the maximal specific activity (V_max 0.41 nmol min⁻¹ mg⁻¹) reported by Agusti et al. [32] for T. cruzi trans-sialidase catalyzed transfer of sialyl from 3'-sialyllactose to terminal galactose in (synthetic) Galβ(1 → 6)GlcNAc mucins of the type involved in infection and pathogenesis of T. cruzi. With cGMP as the donor (at 5 g/L) and lactose as the acceptor molecule (at 20 g/L) the rate of sialylactose synthesis using the trans-sialidase from T. cruzi occurs at a rate which is ~450 times less than that obtained in the present study with Tr6 [14]. As well, Bifidobacterium infantis sialidase catalyzed trans-sialylation, employing high levels of both cGMP and lactose [15], producing only a low level of sialyllactose, took place at a rate which was 87-fold lower than that obtained for Tr6 in the present study.

However, the specific V_max rates reported recently for optimal galactose transfer reactions for galacto-oligosaccharide synthesis using sursucrose as donor and lactose as acceptor, employing the widely studied Bacillus circulans β-galactosidase (commercially known as Biolacta FNS®) may reach ~2 mM min⁻¹ [31]. This specific activity is ~12 times higher than the specific trans-sialidase activity of Tr6 of 16.2 nmol min⁻¹ µg⁻¹ assuming an enzyme dosing of the Biolacta FNS® based on a para-nitrophenol galactoside activity of the B. circulans β-galactosidase of ~1 U/mg [31,33].

In general it has been shown that a few hydrolyses can carry out a reverse reaction to synthesize organic molecules when the enzymatic process conditions have been optimized. This optimization can be performed by changing the reaction conditions such as temperature in the case of β-galactosidase for galactooligosaccharide production [34], glucose content in the case of the acyltransferase for production of glucose emulsifiers [35], and maltose level in the case of generating isomaltooligosaccharides (IMO) by glucoamylase [36]. These studies also showed that optimized process conditions can have a profound, stimulating effect on the rate of production and final yield of the products. The latter is in agreement with the present work whereby a sialidase with low trans-sialidase has been tuned by favorable process conditions to exert a very high trans-sialidase activity. The level of Tr6 produced in P. pastoris allowed a scaling of the enzymatic process to 5 L to establish proof-of-concept for the Tr6 trans-sialidase productivity at scale, providing 3.6 g of 3'-sialyllactose. To further validate the in vitro concept with the new biocatalyst, LNT, LNNT, LNFP I and LNFP V were also successfully decorated with sialic acid derived from cGMP using the same procedure as employed for 3'-sialyllactose production. The availability of optimized enzymes able to act on
components from large scale side-streams from the agricultural and dairy industries is a very important first part of the decision base for the food industry for developing competitive processes for new functional food ingredients. The further interplay between molecular evolution of the biocatalyst, i.e. Tr6, and biochemical engineering optimization could further increase the yield of the process developed here.

Acknowledgements

We express our gratitude to “The Strategic Research Council” (DF) for the financial support to the strategic project “Enzymatic production of human milk oligosaccharides”. We also acknowledge the gift of the cGMP from Arla Foods (A/S), and we thank Jesper Holck and Stine Jørgensen, Center for Bio-Process Engineering, DTU Chemical Engineering, for analysing the sialic acid content in cGMP.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.procbio.2013.10.023.

References

Paper V


Novel α-L-fucosidases from a soil metagenome – identification and characterization of their hydrolytic and transglycosylation abilities.

Applied and Environmental Microbiology, 2014, in preparation
Novel α-L-fucosidases from a soil metagenome - identification and characterization of their hydrolytic and transglycosylation abilities

(Draft to be submitted for publication in Applied and Environmental Microbiology)

Authors
Mateusz Lezyk\textsuperscript{a}, Carsten Jers\textsuperscript{a}, Louise Kjaerulf\textsuperscript{b}, Charlotte H. Gotfredsen\textsuperscript{b}, Maria D. Mikkelsen\textsuperscript{c}, Jørn D. Mikkelsen\textsuperscript{a,*}

Address
\textsuperscript{a}Center for BioProcess Engineering, Department of Chemical and Biochemical Engineering, Building 229, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

\textsuperscript{b}Department of Chemistry, Building 201, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

\textsuperscript{c}Department of Plant and Environmental Sciences, University of Copenhagen, 1871 Copenhagen, Denmark

*To whom correspondence should be addressed:

Jørn D. Mikkelsen

jdm@kt.dtu.dk

Tel: +4545252938
Abstract

This paper describes discovery of novel α-L-fucosidases for synthesis of various fucosylated oligosaccharides but also for hydrolysis of variably-linked α-L-fucose in selected plant materials. Several novel α-L-fucosidase-encoding genes were identified by functional screening of a soil-derived metagenomic library. In total, seven genes were expressed in E. coli as recombinant 6xHis-tagged proteins and were successfully produced and purified. All discovered enzymes can be classified in glycosyl hydrolase family 29 (GH-29). With exception of one, all enzymes were substrate-inhibited with Ki values ranging from 2.6 to 30.1 mM and were most hydrolytically active in pH range 6-7, when tested with p-nitrophenyl-α-L-fucose as the substrate. A difference in capability for hydrolysis of 2′-fucosyllactose, 3-fucosyllactose and various plant materials was observed and appeared to correlate well with sequence similarity. Several enzymes were found to catalyse transglycosylation either by self-condensation or using lactose as acceptor and one exhibited an unusually high transglycosylation/hydrolysis ratio.

Keywords

α-L-Fucosidase; Trans-fucosylation; Fucosyllactose; HMO; Metagenomics; Enzyme discovery;
1. Introduction

L-Fucose (6-deoxy-L-galactose) is a monosaccharide with numerous biological roles and is observed in natural materials from various sources. In plant xyloglucans, which are often highly branched, L-fucose exists usually as terminal residue and is almost exclusively found in its α-anomeric configuration (1). Fucose can be O-methylated to various degrees in these materials (1, 2). Substantial amounts of sulfated fucose are found in fucoidan, a polysaccharide extracted from different species of brown algae (3). In addition, extracellular polysaccharides from various bacteria, fungi and micro-algae also contain L-fucose (1).

Fucosylated glycoconjugates play a significant role in a wide range of biological processes such as defense in plants (4), receptor signaling, inflammation, metastasis and disease (fucosidosis) in mammals (5). The occurrence of fucose-containing compounds is well established in human milk oligosaccharides (2’-FL, 3-FL or lacto-N-fucopentaoses) (6, 7) blood group antigens (H, Le^a, Le^b determinants) (8) as well as at the non-reducing end of many glycan structures present on mammalian cells surfaces (1). The α-L-fucosyl residue is found α(1-2)-linked to D-galactose residues or α(1-3)-, α(1-4)-, or α(1-6)-linked to N-acetyl-D-glucosamine (9). Some of the more abundant fucosylated oligosaccharides in human milk, 2’-FL and 3-FL, not only stimulate the growth of bifidobacteria but also serve as antiadhesive antimicrobials (9).

A widespread scientific interest in assessing biomedical properties of these fucosylated glycoconjugates, in particular oligosaccharides, stimulates efforts for their efficient and specific synthesis. Approaches include either chemical or enzymatic synthesis or combination of the two. As compared with proteins, there is no universal mechanism for carbohydrate synthesis in living organisms. Moreover, enormous diversity of their stereochemistry makes it very difficult to automate
their synthesis. Although chemical methods have been developed, they are hampered by complexity of regio- and stereochemical control necessitating multiple protection and deprotection steps (10). The enzymatic synthesis can be accomplished by glycosyltransferases (GT) or glycoside hydrolases (GH). Mild conditions, higher specificity and simplicity of the enzymatic process are the advantage of this approach.

GTs are responsible for the synthesis of glycoconjugates and oligosaccharides in all living organisms. They do not hydrolyze the product and are extremely efficient but their use is limited by expensive fucosyl nucleotide (GDP-Fuc) substrate, high specificity for natural donor and acceptor molecules and difficulty of enzyme purification (11). Attention has been devoted also to development of GT-based microbial systems with endogenous production of the activated sugar substrate intermediate (12). A maximum 2’-FL concentration of 1.23 g/l obtained from a batch fermentation with 14.5 g/l lactose was reported by Lee et al. (13).

The other alternative for the enzymatic synthesis of fucosyl glyconjugates is the use of GHs in transglycosylation reactions. When the glycosyl-enzyme intermediate interacts with acceptor other than water, transglycosylation reaction occurs (14). GHs that catalyze the hydrolysis of L-fucose are found in all three domains of life. Based on amino acid sequence similarity, all \( \alpha \)-L-fucosidases hydrolyzing fucosyl derivatives with retention of the anomeric configuration are classified in CAZY (15) GH-29 family, whereas inverting \( \alpha \)-L-fucosidases are classified in family GH-95. The retaining GH-29 \( \alpha \)-L-fucosidases utilize a classical Koshland double displacement mechanism in which an aspartate residue is used as the catalytic nucleophile (16, 17). Under normal conditions these enzymes cleave glycosidic bonds in oligosaccharides and glycoconjugates. Glycosidases are attractive for oligosaccharide synthesis because glycoside substrates are relatively inexpensive, stable and some readily available.
The usual promiscuity of GH binding sites enables glycosylation of a wide variety of acceptor molecules.

Several α-L-fucosidases have been shown to exhibit transglycosylation properties. Examples include α-L-fucosidases from *Aspergillus niger*, *Corynebacterium* sp., *Penicillium multicolor* (18), *Pecten maximus* (19), *Alcaligenes* sp. (20), *Thermus* sp. Y5 (21), and canines (19). Depending on enzyme source and reaction conditions, the final reaction yields remain moderate.

Substantial advances in the discovery and engineering of these enzyme classes were thoroughly reviewed (22-24). α-fucosidase from *Thermotoga maritima* was successfully transformed into an α-transfucosidase by directed evolution as reported by Osanjo *et al* (25). Moreover, much interest was devoted to glycosynthases - a new class of mutant glycosidases in which the active site nucleophile is replaced with a non-nucleophilic residue. (26) The mutation completely inactivates the enzyme but the activity can be restored in the presence of substrates with good leaving groups. This approach was successfully applied to *Sulfolobus solfataricus* α-L-fucosidase whose engineering led to yields as high as 91 % for selected acceptors and β-L-fucosyl azide as donor (27).

Main challenges in applicability of GHs include enzyme mediated product degradation (hydrolysis) and low regiospecificity. Moreover, most studies focus on the use of relatively expensive fucose donors that limits the possibility of designing economically feasible processes. Selected high-fucose containing materials such as tragacanth, fucoidan or xyloglucan can offer interesting alternatives, if enzymes combining capability to synthesize various fucoooligosaccharides and utilize such fucose donors are discovered.

Functional screening of metagenomic libraries for discovery of novel enzymes has gained a lot of attention over the last years (28). It is an interesting alternative both to traditional culture-based methods and enzyme engineering for identification of novel variants of well-established enzyme
classes with new abilities. Despite challenges related to designing efficient and specific screening campaigns, several enzymes – including glycoside hydrolases – originating from various metagenomes have been identified and characterized (29).

In this study we screened a soil metagenomic library for novel α-L-fucosidases. Relevant genes were identified and recombinant enzymes produced in *E. coli*. Subsequently, their hydrolytic activity and capability to synthesize various fucosylated oligosaccharides were investigated. The *T. maritima* α-L-fucosidase and its L322P mutant which are relatively well characterized in scientific literature were used as benchmarks in this study (25). The use of α-L-fucosidases to synthesize the human milk fuco-oligosaccharides 2’-FL and 3-FL was demonstrated.
2. Materials and methods

2.1 Substrates

*p*-Nitrophenyl-\(\alpha\)-L-fucose (pNP-Fuc) and 5-bromo-4-chloro-3-indolyl-\(\alpha\)-L-fucose (X-Fuc) were purchased from Carbosynth (Compton, United Kingdom). 2’-FL and 3-FL were obtained from Elicityl (Crolles, France). \(\beta\)-lactose was purchased from Sigma-Aldrich (Steinheim, Germany). Fucose was supplied by DuPont Danisco (Denmark). Samples of fucoidan extracted from *Fucus veniculosus*, xyloglucan from citrus peel, and tragacanth gum from *Astragalus gossypinus* were kindly provided by Marcel Tutor Ale and Hassan Ahmadi Gavlighi from DTU Chemical Engineering (Kgs. Lyngby, Denmark).

2.2 Strains and plasmids

The *E. coli* ML297 strain is a \(\Delta lacZYA\) derivative of EPI300\(^{TM}\)-T1\(^R\) (Epicentre, USA) that was transformed with the heat-inducible lysis vector pEAS-1a (DualSystems Biotech, Switzerland). This strain was used for the maintenance and propagation of constructed metagenomic library. *E. coli* DH5\(\alpha\) was used for subcloning and plasmid propagation. *E. coli* BL21(DE3) and OverExpress C41 (DE3) strains were used for production of recombinant \(\alpha\)-L-fucosidases (Novagen and Lucigen, USA).

2.3 Construction of metagenomic library

Metagenomic DNA was extracted from 10 g of soil sample, using the PowerMax Soil DNA Isolation kit (Mobio Laboratories Inc.), using the manufacturer’s recommendations. Libraries of 40-50 kb metagenomic DNA in *E. coli* were created using the CopyControl Fosmid Library Production kit (Epicentre Biotechnologies). In brief, DNA was end-repaired and high molecular weight (40-50 kb) fragments were size selected and purified using 1 % low melting point-agarose gel electrophoresis.
(overnight). 0.25 μg of metagenomic DNA was ligated with 0.5 μg of the linearized fosmid pCC1FOS vector and packaged using replication-deficient phage extract. *E. coli* strain ML297 was transfected with the resulting phages and library size was determined by plating serial dilutions on LB agar plates containing 12.5 μg/ml chloramphenicol and 100 μg/ml ampicillin. *E. coli* metagenomic DNA library was grown to mid-log phase for 8 h shaking at 30 °C in 50 ml LB supplemented with relevant antibiotics, and 2 ml aliquots with 15 % glycerol was stored frozen at -80 °C. Each frozen stock was subsequently confirmed to have about 50,000 colony forming units per ml. On the basis of the estimated library sizes, an average redundancy of about 1-2 clones per library aliquot could be inferred.

2.4 Screening for α-L-fucosidase-bearing metagenomic clones

To screen for novel α-L-fucosidase genes, a 1 Gbp soil metagenomic library was plated on standard size LB agar plates (10 cm in diameter) supplemented with 34 μg/ml chloramphenicol, 100 μg/ml ampicillin and 100 μg/ml of X-Fuc to a density of approximately 1,000 CFU. In total about 100,000 colonies were cultivated for 3 days at 30 °C (to avoid heat-induced lysis) and colonies turning blue were selected. Fosmid material was isolated from liquid cultures of selected clones using FosmidMAX™ DNA Purification Kit (Epicentre, USA). Restriction analysis of fosmids with *Bam*HI and *Pst*I restriction was performed to find out the redundancy of all clones identified as positive. To facilitate further sequencing and gene identification, the fosmid DNA was digested with *Bam*HI and further subcloned into *Bam*HI-linearized and dephosphorylated pUC18 vector.

Ligated DNA was used to transform electrocompetent cells of *E. coli* DH5α strain and resulting transformants were plated on LB agar plates supplemented with 100 μg/ml ampicillin and 100 μg/ml X-Fuc and incubated overnight at 37 °C. Transformants degrading X-fuc contained fragments of the original fosmid with a functional α-L-fucosidase gene and were selected for further analysis.
2.5 Sequencing, identification and cloning of α-L-fucosidase genes

pUC18-derivatives containing α-L-fucosidase-encoding genes were linearized using BamHI or PstI and gel-purified. These fragments were sequenced by Ion Torrent PGM sequencing using 316 chip (DMAC, Lyngby, Denmark). Sequence reads were trimmed using clc_quality_trim script (CLC BIO, Aarhus, Denmark) and assembled using Ray Meta (30). BlastX (31) was used to search for closest homologs on protein level and to annotate all contigs. Primers were designed to close gaps between contigs containing truncated α-L-fucosidase-annotated genes at their 5’ or 3’ ends. For expression in E. coli, genes were PCR-amplified using specific primers (Table 1) and PCR products were restricted and inserted in the plasmid pETM10 between NcoI and KpnI. A gene encoding the α-L-fucosidase of Thermotoga maritima (Thma) was codon-optimized for E. coli and synthesized by GeneArt AG (Figure S3). The gene was PCR-amplified using relevant primers (Table 1), restricted and inserted in pETM10 between NcoI and KpnI. The mutation L322P was introduced by PCR using overlapping primers. The resulting plasmids were used to transform E. coli BL21(DE3) or C41(DE3) strains.

2.6 Expression and purification of α-L-fucosidases

E. coli BL21(DE3) and C41(DE3) harboring recombinant plasmids were cultured in LB medium shaking at 30 °C prior to induction at OD₆₀₀ 0.6 with 0.2 mM or 1 mM IPTG. Expression was continued overnight at 25 °C with 180 rpm. The cell pellets were harvested by centrifugation and resuspended in binding buffer (20 mM phosphate-citrate buffer, 100 mM NaCl, 20 mM imidazole, pH 7.4). Cells were lysed by sonication and centrifuged at 5,000 g for 20 min. The supernatant was subjected to sterile filtration through a 0.22 μm filter and subsequently loaded onto a Ni²⁺-sepharose HisTrap HP column (5 ml, GE healthcare). The purification was carried out using an ÄKTA purifier (GE healthcare). The column was equilibrated with binding buffer until UV and conductivity baselines
stabilized. After loading of the protein sample, unbound proteins were removed by washing with 10 CV of binding buffer, and the target recombinant enzyme was eluted with a linear gradient of 10 CV containing 20 to 500 mM of imidazole. The fractions containing 6xHis-tagged α-L-fucosidases were collected and desalted using PD-10 desalting columns (GE Healthcare) to remove imidazole. Protein concentrations were determined using the BCA protein assay (Thermo scientific) with BSA as the standard.

2.7 Characterization of hydrolytic activity of novel α-L-fucosidases

pH-dependence for hydrolysis was tested using 0.1 and 1.0 mM pNP-Fuc as substrate at 30 °C. The reactions were done in triplicates with the following buffers: 50 mM phosphate-citrate buffer (pH 3-8), 50 mM tricine buffer (pH 8-9), 50 mM glycylglycine buffer (pH 8-9), 50 mM glycine-NaOH buffer (pH 9-10). Reactions with 0.1 mM pNP-Fuc were initiated by addition of enzyme at a final concentration of 0.2 μg/ml Mfuc1, 0.2 μg/ml Mfuc2, 0.6 μg/ml Mfuc3, 0.3 μg/ml Mfuc4, 0.2 μg/ml Mfuc5, 1.0 μg/ml Mfuc6, 0.2 μg/ml Mfuc7, 0.3 μg/ml Thma wild type, or 0.5 μg/ml Thma L322P. For reactions with 1.0 mM pNP-Fuc, with the exception of Mfuc2, the double concentration of enzyme was used. The release of p-nitrophenol (pNP) was followed by withdrawing samples at different time points (2, 4 and 8 min for 0.1 mM pNP-Fuc and 5, 10 and 20 min for 1.0 mM pNP-Fuc) and measurement of absorbance at 405nm after dilution with 0.5 M sodium carbonate.

For a substrate saturation experiment, reactions were set up as described for the pH-dependence experiment (0.1 mM pNP-Fuc) except that reactions were done at respective optimum pH and with 0.1, 0.25, 0.5, 1.0, 2.5, and 5 mM pNP-Fuc, and in case of Mfuc2 0.3 μg/ml of enzyme was used. Samples were withdrawn at 2.5, 5 and 10 min. Reactions were done in triplicate and kinetic data for Mfuc6 were fitted using allosteric model and the data for remaining enzymes with substrate inhibiton model. Non-
linear regression analysis was performed with the use of GraphPad Prism (GraphPad Software, USA).

For estimation of the temperature stability, the enzymes were incubated in 5 mM buffer at optimal pH at a concentration of 0.8 μg/ml Mfuc1, 0.4 μg/ml Mfuc2, 1.2 μg/ml Mfuc3, 0.9 μg/ml Mfuc4, 0.7 μg/ml Mfuc5, 3.9 μg/ml Mfuc6, 0.5 μg/ml Mfuc7, 0.9 μg/ml Thma wild type, or 1.5 μg/ml Thma L322P for 10, 20, 30 and 40 min at various temperatures as indicated (30-100 °C). Residual enzyme activity was measured in a continuous assay using 1 mM pNP-Fuc with 50 mM buffer (pH 7) and 10 % of the enzyme concentration stated above and following release of pNP spectrophotometrically at 405 nm. The substrate specificity of purified enzymes was analyzed at 30 °C using 1 mM of the following chromogenic substrates in 50 mM phosphate-citrate buffer (pH 7.0): pNP-β-D-Gal, pNP-β-D-Glc, pNP-β-D-Lac using enzyme concentrations stated above. The release of pNP was followed spectrophotometrically at 405 nm in continuous assays.

The hydrolysis of 2’-FL, 3-FL and fucose containing polymers, citrus peel xyloglucan, tracacanth and fucoidan was performed in duplicates in 96-deep well plates in total volume of 500 μl, at 30 °C and enzyme loading of 10 μg (except 20 μg for Mfuc7). Initial concentration of 1 mM and 150 min-reaction time for FLs were used. For polymer substrates the initial concentration of substrates was chosen to correspond to a final fucose content of 0.25 mM. The reactions were carried out for 240 min and post-reaction mixtures were filtered using AcroPrep™ Advance 96-Well Omega Filter and ultrafiltered using Ultrafilter (10 kDa MWCO) plates, respectively (Pall, USA). Diluted permeates were analyzed by HPAEC-PAD.

2.8 Immuno-glycan microarray analysis of fucosylated xyloglucan cleavage

The extraction of cell wall components and subsequent analysis was performed essentially as described before (32). 4 M NaOH (containing 0.1 % NaBH₄) extracts from the different plant species were
spotted using a microarray robot (Sprint, Arrayjet, Roslin, UK). Each extracted sample was printed as a two-fold dilution followed by 3 five-fold dilutions. All sample dilutions were performed in printing buffer (55.2 % glycerol, 44 % water, 0.8 % Triton X-100). The samples were printed onto nitrocellulose membrane with a pore size of 0.45 μm (Whatman, Maidstone, UK). Once printed, arrays were blocked with PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.7 mM KH₂PO₄, pH 7.5) containing 5 % w/v low fat milk powder (M-PBS) for 1 hour. Hereafter, they were washed in PBS and incubated with or without 10 μg of tested α-L-fucosidases at 30 °C at optimal pH. After incubation, arrays were washed thoroughly with PBS, followed by 2 hours incubation with xyloglucan-specific monoclonal antibodies LM15 (PlantProbes, Leeds University, UK) diluted 10-fold or the antibody CCRC-M1 specific for fucosylated xyloglucan (Complex Carbohydrate Research Center) diluted 250-fold in M-PBS. Subsequently, arrays were washed in PBS and incubated with anti-rat or anti-mouse secondary antibody respectively conjugated to alkaline phosphatase (Sigma, Poole, UK) diluted 1/5000 in M-PBS. After washing with PBS and dH₂O, arrays were developed in a solution containing 5-bromo-4-chloro-3-indolyolphosphate and nitro blue tetrazolium in alkaline phosphatase buffer (100 mM NaCl, 5 mM MgCl₂, 100 mM diethanolamine, pH 9.5). Developed microarrays were scanned at 2400 dpi (CanoScan 8800F, Canon, Søborg, Denmark) and converted to TIFFs. Antibody signals were measured using appropriate software (Array-Pro Analyzer 6.3, Media Cybernetics, Rockville, USA). Data were presented as two datasets where maximal spot signal was set to 100 and all other values normalized accordingly. Color intensity is correlated to mean spot signal value and a cut-off of 5 was applied.

2.9 Identification of selected reaction products by nuclear magnetic resonance (NMR).

For initial identification of trans-fucosylation products, non-buffered reactions were performed using 20 mM pNP-Fuc as the donor and 25 mM lactose as the acceptor at 30 °C for 50 min. 500 μl reactions
were started by the addition of enzymes to final concentration of 42 μg/ml Mfuc1, 12 μg/ml Mfuc2, 41 μg/ml Mfuc3, 34 μg/ml Mfuc4, 18 μg/ml Mfuc5, 65 μg/ml Mfuc7, 17 μg/ml Mfuc6, 51 μg/ml Thma wild type, and 58 μg/ml Thma L322P. The enzymes were removed using Vivaspin 500 filters with a 5 kDa cut-off (Sartorius, Germany) at 4 °C. The permeate was freeze-dried and NMR spectra were acquired in 500 μL D2O (99.9 % D, Sigma-Aldrich) at 25 °C on a Varian Unity Inova 500 MHz spectrometer equipped with a 5 mm probe or a Bruker Ascend 400 MHz spectrometer with a 5 mm Prodigy cryoprobe using standard pulse sequences. External standard of 1,4-dioxane was used as chemical shift reference (δH 3.75 ppm and δC 67.4 ppm). Structure elucidation was achieved by full 1H and 13C NMR assignments using standard 1D and 2D homo- and heteronuclear NMR experiments (33, 34). Chemical shifts of compounds annotated in this study are presented in the Supplementary material (Figure S7).

2.10 High-performance anion exchange chromatography (HPAEC-PAD) and reverse phase chromatography (RP-HPLC)

A Dionex ICS-5000 system consisting of DP-5000 gradient pump, ED-5000 electrochemical (pulsed amperometric) detector coupled to an AS-AP autosampler (Dionex Corp., CA) was used for HPAEC-PAD analysis. The separation was accomplished using a CarboPac PA1 analytical column (4 mm × 250 mm) with CarboPac PA1 guard column (4 mm × 50 mm) (Dionex Corp.) at 30 °C with a flow rate of 1 ml/min. Hydrolysis products of 2’-FL, 3-FL and high-molecular substrates were analyzed with the following programme: 0-16 min at 75 mM NaOH, 17-21 min at 250 mM NaOH and regeneration 22-35 min at 75 mM NaOH. For analysis of products in transfucosylation reactions, a 40 min isocratic separation using 75 mM NaOH as eluent was performed.
To analyse peak identities of individual \( p \)-nitrophenyl \( \alpha \)-L-fucosyl-\( \alpha \)-L-fucose products (pNP-Fuc-Fuc), transfucosylation reactions were performed as described in section 2.11 with the use of Mfuc6 and Mfuc7. After enzyme removal, compounds in the 20 \( \mu \)l permeate samples were separated by RP-HPLC using a ODS-L Optimal column (250 mm x 4.6 mm) at 30 °C and with a flow rate of 1 ml/min. A two-
solvent system consisting of acetonitrile (A) and water (B) was used with a linear gradient from 3 to 30 % A for 40 min and regeneration at 3 % A for 15 min. The individual pNP-containing compounds were detected by UV spectrophotometry at 225 nm and collected into separate fractions. The procedure was repeated three times using each of the post-reaction mixtures, fractions corresponding to individual peaks were pooled and freeze-dried. The fractions were analysed by \(^1\)H NMR and HPAEC-PAD as detailed above.

2.11 Time course of transglycosylation with selected enzymes

Transfucosylation with selected enzymes was evaluated in a time-course experiments using 25 mM pNP-Fuc as donor and 100 mM lactose as acceptor, at 30 °C and pH optimum (as determined for hydrolysis of pNP-Fuc). The final enzyme concentrations were as for the transfucosylation experiment mentioned above (methods 2.9). The reactions were carried out in duplicates in 96-well deep well microtiter plates. 30 \( \mu \)l samples were withdrawn at indicated time points and diluted 10-fold with ice-
cold milliQ water. The post-reaction mixtures were filtered using AcroPrep™ Advance 96-Well Omega Filter and ultrafiltered using Ultrafilter (10 kDa MWCO) plates, respectively (Pall, USA). The permeates were further diluted for determination of pNP content by spectrophotometry at 410 nm and HPAEC-PAD analysis.
2.12 Accession numbers

The complete nucleotide sequences of novel α-L-fucosidase encoding genes are available in GenBank under the following accession numbers: KJ626336 (mfuc1), KJ626337 (mfuc2), KJ626338 (mfuc3), KJ626339 (mfuc4), KJ626340 (mfuc5), KJ626341 (mfuc6), and KJ626342 (mfuc7).
3 Results

3.1 Screening for α-L-fucosidases, subcloning and identification of gene sequences

In this study a soil metagenomic library was used to screen for novel α-L-fucosidases. Of approximately 100,000 colonies screened using the chromogenic substrate 5-bromo-4-chloro-3-indolyl-α-L-fucose (X-Fuc), α-L-fucosidase activity was detected in 40 clones. From these clones, fosmids were isolated and subjected to restriction analysis resulting in identification of 16 non-redundant clones. Fosmid inserts of about 50 kb were digested with BamHI or PstI and the smaller restriction products were inserted in the vector pUC18 and screened on X-Fuc yielding 7 positive clones. After sequencing by Ion Torrent, subsequent annotation and ORF analysis revealed 9 contigs containing full-length (5 contigs) or partial sequences of α-L-fucosidase genes. Analysis of the putative promoters and ribosomal binding sites allowed identification of translation starts and stops and missing sequences were determined by PCR using primers annealing up- and downstream of each α-L-fucosidase gene. The 7 identified α-L-fucosidase-encoding genes were denoted mfucl, 2, 3, 4, 5, 6 and 7.

3.2 Sequence analysis

An overview of the genes identified in this study is presented in Table 2. Based on BLAST analysis all enzymes, with exception of Mfuc3 showed sequence similarity with α-L-fucosidases with sequence identity to best hit ranging from 56-76 % (Table 2). Mfuc3 shares 41 % with the closest homolog identified by BlastX – β-galactosidase trimerisation domain-containing protein. It is also the least similar (32 % identity) when α-L-fucosidas reported in this work are compared (not shown). The genes had a GC-content ranging from 58.5 to 65.4 and most of the α-L-fucosidases had a theoretical molecular weight between 47.3 and 52.2 kDa while Mfuc3 was larger with a molecular weight of 75.6 kDa. The sequences were analysed using the NCBI Conserved Domain software (35) and all sequences
except Mfuc3 were found to contain an α-L-fucosidase domain (pfam01120). Significant hits for Mfuc3 were to GHL6 (pfam14871) a family of hypothetical glycosyl hydrolases and to A4 beta-galactosidase middle domain. Using the sequence-based annotation tool in the CAZYmes Analysis Toolkit (36), all identified enzymes, were classified in GH 29. Mfuc1-7 and some of the most established representatives of GH-29 family are presented in a phylogram (Figure 1). It has been shown that the catalytic nucleophile in GH-29 enzymes is a conserved aspartate residue (16). This residue was also identified in each of the α-L-fucosidases reported here by multiple alignment.

3.3 Production of recombinant α-L-fucosidases in E. coli

The seven metagenomic α-L-fucosidase genes were inserted in the pETM10 vector and were expressed in E. coli BL21(DE3) and C41(DE3). The highest yields of recombinant proteins as judged by SDS-PAGE, were obtained for Mfuc1, 2 and 4 when expressed in E. coli BL21(DE3) and Mfuc3, 5, 6 and 7 expressed in E. coli C41(DE3) (data not shown). SDS-PAGE of purified 6xHis-tagged α-L-fucosidases verified the expected size of enzymes (Figure 2).

3.4 Termostability, pH optimum, and apparent kinetics of pNP-Fuc hydrolysis.

Substantially, all the seven enzymes were able to hydrolyze the pNP-Fuc substrate. The specificity for hydrolysis of other non-fucose pNP-glycosides was also evaluated. However, none of the enzymes tested exhibited detectable activity. The physicochemical properties of the enzymes were characterized using pNP-Fuc as substrate. Most enzymes were active in a relatively broad pH ranges and exhibited maximum activity at pH 6-7 (Table 3). Mfuc6 hydrolyzed pNP-Fuc optimally at pH 9. Also the thermostability of the enzymes was tested and although some differences were observed, all enzymes had a half-life of less than 5 min at 60 °C and can be categorized as mesophilic. In contrast, the
benchmark enzyme from *T. maritima* had a half-life of about 48 min at 100 °C. The complete data sets recorded for each enzyme are shown in supplementary Figures S1 and S2.

Next, we determined the dependence of initial reaction rate on the substrate concentration in the range from 0.1 to 5 mM of pNP-Fuc and found that the $K_m$ values ranged from 68 μM to 308 μM (Table 4). All enzymes except Mfuc6 exhibited some level of substrate inhibition with Ki values ranging between 2.6 and 30.1 mM (Table 4; Figure S5). Interestingly, for Mfuc6 the substrate-saturation data was best modeled using an allosteric sigmoidal model, and the Hill coefficient of 0.40 indicates a negative cooperativity in substrate binding.

3.5 Substrate specificity

After characterizing the enzymes using the artificial substrate pNP-Fuc, we wanted to analyse the specificity towards various fucose-containing compounds of natural origin. First, we tested the activity of the enzymes towards two human milk fucose-oligosaccharides – 2’-FL and 3-FL in which fucose is $\alpha$(1-2)-linked to galactose and $\alpha$(1-3)-linked to glucose, respectively (Table 4). The enzymes Mfuc3 and 6 expressed no activity on these substrates. The other enzymes degraded 2’-FL, but only Mfuc5 and to a lesser extent Mfuc1 and 7 were able to hydrolyze 3-FL. Xyloglucan extracted from citrus peel - containing 27 mg fucose/g was tested but only Mfuc5 exhibited substantial activity on this polymer and released 39 ± 2 % of total fucose (data not shown). To further examine this observation, we used glycan arrays to investigate the activity on fucosylated xyloglucan originating from two flowering plants *Arabidopsis thaliana* and *Sambucus nigra*. The presence of fucose in these plant cell wall materials was monitored by binding of monoclonal antibody CCRC-M1 that is specific for fucose-containing xyloglucan structures (37). Four enzymes, Mfuc1, 2, 4, and 5 were able to catalyse hydrolysis of fucosylated xyloglucans from the plants as shown by complete loss of CCRC-M1.
antibody labelling (Figure 3). Less pronounced activity was observed for Mfuc7. As a control, binding of monoclonal antibody LM15 was shown to not be influenced by the action of the tested α-L-fucosidases. This antibody is specific for xyloglucan epitopes XXX(G) and XLX(G)/XXL(G) that are not attacked by α-L-fucosidases (38, 39).

Finally, we tested whether the enzymes were able to act on fucoidan from Fucus veniculosus, a polymer containing α(1-2)-linked sulphated fucose (40) and on tragacanth gum from Astragalus gossypinus a fucose-rich pectin-like structure where fucose is thought to be α(1-2)-linked to xylose side chains (41). However none of the enzymes were able to release fucose from these polymers.

3.6 Transfucosylation activity of metagenomic α-L-fucosidases

For initial evaluation of transglycosylation properties we set up reactions using 20 mM pNP-Fuc as donor and 25 mM lactose as acceptor. The formation of transglycosylation products was evaluated by HPAEC-PAD and NMR. When evaluated using HPAEC-PAD Thma wild type and its L322P mutant produced 0.1 mM of 2’-FL that corresponds to 1 % yield based on donor substrate (Table 6).

Production of 0.1 mM of 3-FL was catalyzed by Mfuc5, corresponding to 1 % yield. The remaining enzymes produced no or lower amounts of FLs. Interestingly, other products were detected. Three different self-condensation products were identified by NMR, corresponding to p-nitrophenyl α-L-fucosyl-(1-2)-α-L-fucose as well as the (1-3)- and (1-4)-linked pNP-Fuc-Fuc glycosides (Figure 6).

Linkage positions were determined by strong and unambiguous heteronuclear multiple bond correlations (HMBC) across the glycosidic bonds between the respective fucose residues, as well as from Fuc H1 to C1 of pNP. Nuclear Overhauser effect spectroscopy (NOESY) connectivities were used to confirm the structures. Selected NMR spectra and chemical shifts of compounds characterized in this study are collected in supplementary Figures S6 and S7. 2’-FL and 3-FL were present in insufficient
amounts for definitive confirmation by NMR. Due to very low yield we were not able to identify all of
the products formed in the reaction that were detected by HPAEC-PAD (Figure 5). RP-HPLC
separation of individual α-linked pNP-Fuc-Fuc compounds from reaction mixtures of Mfuc6 and
Mfuc7 enabled the identification of the respective HPAEC-PAD peaks (Figure 5) by \(^1\)H NMR
spectroscopy. Higher DP oligosaccharides were not detected when analyzing \(^1\)H NMR spectra. The
concentrations of detected condensation products were estimated by comparison of characteristic non-
overlapping \(^1\)H NMR signals and these of fucose in each sample and are presented in Table 6.
Predominantly, pNP-Fuc-(1-2)-Fuc was produced in the reaction catalyzed by Mfuc6 at 5.7 mM
corresponding to a yield of 34 % based on the donor. The total concentration of pNP-Fuc-Fuc products
was in this case 6.8 mM. pNP-Fuc-(1-2)-Fuc was also the predominant self-condensation product and
only Mfuc7 produced preferentially pNP-Fuc-(1-3)-Fuc.
It has been demonstrated for other hydrolases that high concentration of acceptor can improve
transglycosylation yields (42). We therefore set up reactions with a higher acceptor:donor ratio (100
mM lactose as acceptor and 25 mM pNP-Fuc as donor) and followed the reactions in time course-
experiments (Figure 4 and S4). The transient maximum yield of 3-FL present in the reaction catalyzed
by Mfuc5 was 0.9 mM whereas 1.6 mM of 2'-FL was produced in the reaction with Thma L322P.
Throughout the course of the reaction catalyzed by Mfuc6, fucose concentration remained very low
compared to concentration of pNP released. Formation of other products was observed, albeit not
quantified due to lack of external standards.
4. Discussion

4.1 Identification of novel α-L-fucosidases from metagenomic libraries

This paper reports the identification and characterisation of seven novel α-L-fucosidases present in the soil metagenome with respect to both hydrolysis and transfucosylation. *E. coli* was used as a heterologous host for both functional screening of the metagenomic library and the expression of discovered genes. This imposes a bias towards prokaryotic genes compatible with *E. coli* transcription and translation machinery (43). Several solutions for enhancing the number of heterologously expressed genes in *E. coli* exist, but no universal way of expressing the full set of enzymatic activities from given metagenome has yet been proposed (44). A key element in a successful screening campaign is the choice of assay. It is a prerequisite that it allows high-throughput screening without compromising the selectivity and sensitivity towards the desired activity (45). In contrast to several high throughput screens developed for glycosidases, assaying the transfer activity is extremely challenging as no obvious change in fluorescence or absorbance is associated with the reaction. Thus, in this study we searched for enzymes with α-L-fucosidase activity and subsequently evaluated the possibility for transfucosylation using lactose as an acceptor. A similar approach has been applied for screening of mutagenesis libraries for enzymes with improved transfucosylation activity (25, 46).

Evidently, α-L-fucosidases that do not act on X-Fuc or whose hydrolytic activity is minimal under the conditions in the *E. coli* cytoplasm may not be detected. Further, enzymes having a strong preference for trans-glycosylation with very limited hydrolase activities such as e.g. the trans-sialidase found in *Trypanosoma cruzi* (47) would likely not be identified by this approach.
4.2 Sequence analysis of novel α-L-fucosidases

α-L-fucosidases are currently classified in the two glycoside hydrolase families GH-29 and GH-95. Based on amino acid sequence, all seven α-L-fucosidases reported in this work could be classified in the GH-29 family. This family uses an Asp as catalytic nucleophile that was also present in the seven new fucosidases. In *T. maritima* α-L-fucosidase, Glu 266 was identified as the general acid/base nucleophile (48). Despite structural conservation, this residue is not fully sequence-conserved in the GH29 family (49) and in agreement with this, only for a subset of the α-L-fucosidases, a Glu was found to align with Thma Glu266 in multiple sequence alignment.

Recently, the GH-29 family was divided into two groups based on substrate specificity and phylogenetic clustering (50). According to this classification the α-L-fucosidases identified in this study belong to subfamily GH-29A. As observed in the phylogram with some of the best characterised representatives of this family (Figure 1), Mfuc1, 2, 4, 5, and 7 formed a new clade indicating relatively close evolutionary relationship. Mfuc3 was more distantly related to the other α-L-fucosidases.

4.3 Hydrolase activity of novel α-L-fucosidases

To characterise the enzymes, we tested their kinetic properties, as well as pH optimum and thermostability. Most of the enzymes were found to be substrate inhibited. In case of Mfuc6 the substrate-saturation data was best modeled using an allosteric sigmoidal model which suggests the presence of allosteric substrate-binding site, and the Hill coefficient of 0.394 indicated a negative cooperativity in substrate binding. Allosteric substrate binding was not reported for α-L-fucosidases before whereas the substrate inhibition behaviour with Michaelis-Menten kinetics is usually observed. When analyzing thermostability, it was observed that most enzymes were mesophilic, and for most enzymes pH optimum was in the range of pH 6-7. A number of characterized α-L-fucosidases also were
reported to have pH optimum in this range (51). In contrast Mfuc6 was most active at pH 9. According to the Brenda Enzyme database (52) containing 252 α-L-fucosidase entries at the time of writing, Mfuc6 have the highest pH optimum reported to date.

To test the substrate specificity, we measured hydrolase activity on a variety of substrates. These included the two human milk oligosaccharides 2’-FL and 3-FL where fucose is α(1-2)-linked to galactose and α(1-3)-linked to glucose respectively. They were also tested on xyloglucan from Arabidopsis thaliana in which fucose is α(1-2)-linked to galactose (53) and xyloglucan from Sambucus niara. Most of the enzymes were able to degrade 2’-FL to various extents, whereas there appeared to be a correlation between sequence conservation and substrate specificity. In the phylogram, a clade consisting of Mfuc1, 2, 4, 5, and 7 was observed (Figure 1). Although with different efficiency these five enzymes were the only enzymes able to release fucose from 3-FL and xyloglucan. The ability of some α-L-fucosidases to catalyze such reactions has been demonstrated previously (54). Finally we tested the enzymes against fucoidan from Fucus veniculosus and tragacanth gum from Astragalus gossypinus but none of the enzymes were active on these substrates. Fucoidan is a fucose polymer with α(1-2)-linkage containing sulphate in the 4th position (3) and tragacanth gum contains fucose residues α(1-2)-linked to xylose side chains (41). So far, no enzymes have been reported to release fucose from tragacanth. In contrast, several α-L-fucosidases were reported to hydrolyze various fucoidan preparations (55).

Although both Mfuc3 and 6 efficiently cleaved pNP-Fuc, they did not act on other substrates tested here. It remains to be determined what their native activity is.
4.4 Trans-α-L-fucosidase activity of novel α-L-fucosidases

Due to the functional properties of fucosylated compounds, their specific synthesis is receiving increasing interest. Enzymatically, this can be done by employing α-L-fucosidases in transfucosylation reactions. 2’-FL and 3-FL are some of the dominant fucosylated/neutral oligosaccharides in human milk (56, 57) and have been shown to have specific biological functions. We therefore made a preliminary assessment of the potential of using the metagenomic α-L-fucosidases for synthesis of these compounds. As a benchmark, we included one of the mutants of *T. maritima* α-L-fucosidase reported to exhibit increased transfucosylation activity (25). Using pNP-Fuc as donor and lactose as acceptor (acceptor:donor ratio 1.25) we analysed the samples by both HPAEC-PAD and $^1$H and $^{13}$C NMR spectroscopy. The transfucosylation products 2’-FL and 3-FL were observed by HPAEC-PAD analysis although they were present in insufficient amounts for definitive identification by NMR.

Under the experimental conditions applied, the benchmark enzyme Thma wild type exhibited the highest capability for formation of 2’-FL and Mfuc5 for formation of 3-FL (Table 6). Since in most cases, the transfucosylation products are also substrates for the enzyme, it is paramount to control reaction time. Further, it has been demonstrated for some hydrolases that increasing the acceptor concentration can result in increased yields (per donor). We therefore used an acceptor:donor ratio of 4 in a time course experiment and observed a maximum 2’-FL concentration of 1.5 mM, corresponding to 7.5 % yield based on substrate, when reaction was catalyzed by Thma L322P (Figure 4). Thma L322P was reported to yield about 4-fold more transfucosylation product as compared to Thma wild type (25). When using higher acceptor concentration, this trend was confirmed also in this study, where Thma wildtype had a maximal transient yield of 4.5 % 2’-FL, that is 0.9 mM. α-L-fucosidases from various sources such as *Alcaligenes* sp. have been reported to catalyse transfucosylation with higher yields than reported here (42). However with the functional importance of e.g. FL it is surprising that in
most reported studies artificial acceptor molecules such as various pNP-glycosides have been used. In fact is has been observed for most glycosidases, including α-L-fucosidases, that the +2 acceptor binding site exhibits a preference for pNP or an aromatic aglycon part likely related to a high density of aromatic residues (25). In one study (42), lactose was used as acceptor in reactions with porcine and *Alcaligenes* sp. α-L-fucosidase. With a yield of 34 %, *Alcaligenes* sp. α-L-fucosidase was able to fucosylate lactose to produce 3’-FL. In contrast, we here demonstrated the use of α-L-fucosidases to synthesize the basic human milk fuco-oligosaccharides 2’-FL and 3-FL.

It was previously shown that many glycosidases are able to produce transglycosylation products using the same molecule as donor and acceptor, leading to formation of self-condensation products (i.e., disaccharides linked to a nitrophenyl group). By NMR analysis it was concluded that all the α-L-fucosidases made self-condensation products to various degrees. Three different pNP-Fuc-Fuc products were identified, corresponding to p-nitrophenyl α-L-fucosyl-(1-2)-α-L-fucose as well as the α(1-3)- and α(1-4)-linked pNP-Fuc-Fuc glycosides. Literature data for the (1-2)- and (1-4)-linked glycosides by Benešová *et al.* (58) were comparable to our observation, however with minor discrepancies on the chemical shift values, which may result from pH effects. Osanjo *et al.* reported NMR data for the (1-3)-linked glycoside in pyridine-\(d_5\) (25). Mfuc6 that was unable to hydrolyze any of the substrates tested in this study, did not produce FL but demonstrated the highest yield of self-condensation products (33.7 %). For Thma the main self-condensation product was reported to be pNP-Fuc-(1-3)-Fuc (25), whereas under our experimental conditions we observed mainly the formation of α(1-2)- linkages. This discrepancy might be due to difference in reaction conditions. This indicates that careful selection of transfucosylation reaction conditions might be a route to modulate regioselectivity. Attention was not paid to the optimization of yields and therefore strategies such as addition of organic solvents, optimization of acceptor:donor ratio, enzyme load or ionic strength may lead to significant
improvement in the yields of the final products during transglycosylation reaction (59, 60).

4.5 Concluding remarks

In conclusion, our results further substantiate that functional mining of metagenome can lead to successful discovery of diverse glycoside hydrolases, specifically α-L-fucosidases. Using recombinant enzymes under experimental conditions applied we were able to perform the synthesis of fucosyllactose and cleave fucose from various plant polymer structures. Results obtained in this study cannot be directly applicable for efficient synthesis of fucosylated oligosaccharides. Biocatalysts discovered in this work, however, constitute a good starting point for further enzyme and reaction engineering towards this goal.

Acknowledgements

This work was supported by the funding from Hempel (Barcelona, Spain) and the Danish Council for Strategic Research (DSF). Generation of the CCRC-M1 monoclonal antibody used in this work was supported by a grant from the NSF Plant Genome Program (DBI-0421683). We are thankful for donation of fucosylated polymer samples, as well as accompanying data regarding fucose content by Marcel Tutor Ale and Hassan Ahmadi Gavlighi.
References


55. Berteau O, Mulloy B. 2003. Sulfated fucans, fresh perspectives: Structures, functions, and


<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Restriction site</th>
</tr>
</thead>
<tbody>
<tr>
<td>mfuc1 fw</td>
<td>TAGCGGTCTCACATGTTGAAACCAAACCAGACCCAC</td>
<td>BsaI</td>
</tr>
<tr>
<td>mfuc1 rv</td>
<td>TAGCGGTATCTTTAAGGTAACCTCAATCACC</td>
<td>BamHI</td>
</tr>
<tr>
<td>mfuc2 fw</td>
<td>TAGCGGTTCACATGCAACACAAGCGGGAAG</td>
<td>BsaI</td>
</tr>
<tr>
<td>mfuc2 rv</td>
<td>TAGCGGTATCTTTAAGGTAACCTCAATCACC</td>
<td>BamHI</td>
</tr>
<tr>
<td>mfuc3 fw</td>
<td>TAGCGGTTCACATGCAACACAAGCGGGAAG</td>
<td>BsaI</td>
</tr>
<tr>
<td>mfuc3 rv</td>
<td>TAGCGGTATCTTTAAGGTAACCTCAATCACC</td>
<td>BamHI</td>
</tr>
<tr>
<td>mfuc4 fw</td>
<td>TAGCGGTTCACATGTAACAACACAAACGACGC</td>
<td>BsaI</td>
</tr>
<tr>
<td>mfuc4 rv</td>
<td>TAGCGGTTCACATGGAACACTCCCTTTAGACAGGTC</td>
<td>EcoRI</td>
</tr>
<tr>
<td>mfuc5 fw</td>
<td>TAGCGGTTCACATGAATAACGATCGAAGGACGTGGG</td>
<td>BsaI</td>
</tr>
<tr>
<td>mfuc5 rv</td>
<td>TAGCGGTTCACATGAATAACGATCGAAGGACGTGGG</td>
<td>EcoRI</td>
</tr>
<tr>
<td>mfuc6 fw</td>
<td>TAGCGGTTCACATGAAACAAACACAAACCTGGTTCG</td>
<td>BsaI</td>
</tr>
<tr>
<td>mfuc6 rv</td>
<td>TAGCGGTTCACATGAAACAAACAAACCTGGTTCG</td>
<td>EcoRI</td>
</tr>
<tr>
<td>mfuc7 fw</td>
<td>TAGCGGAACGACACATGAACAGCTATGACCCGAAAAC</td>
<td>BbsI</td>
</tr>
<tr>
<td>mfuc7 rv</td>
<td>TAGCGGAACGACACATGAACAGCTATGACCCGAAAAC</td>
<td>BbsI</td>
</tr>
<tr>
<td>Thma fw</td>
<td>ATCGGTTCTCTACATGTTTCAAATGAAAAACGTGCTTATAAAC</td>
<td>BsaI</td>
</tr>
<tr>
<td>Thma rv</td>
<td>ATCGGTTCTCTACATGTTTCAAATGAAAAACGTGCTTATAAAC</td>
<td>KpnI</td>
</tr>
<tr>
<td>L322P fw</td>
<td>CGGAAATCTGGCTGAATGTTGGACCGAAAGGC</td>
<td>EcoRI</td>
</tr>
<tr>
<td>L322P rv</td>
<td>CAACATTTCAGAACAGATTTCGGCTTCAAC</td>
<td>EcoRI</td>
</tr>
</tbody>
</table>
Table 2.

<table>
<thead>
<tr>
<th>Gene names</th>
<th>nt .length</th>
<th>GC content</th>
<th>Identity (similarity, %)</th>
<th>Best hit (accession no.)</th>
<th>Organism name (best hit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mfuc1</td>
<td>1281</td>
<td>58.5 %</td>
<td>67 (81)</td>
<td>α-L-fucosidase (YP_001278245)</td>
<td>Roseiflexus sp. RS-1</td>
</tr>
<tr>
<td>mfuc2</td>
<td>1290</td>
<td>65.4 %</td>
<td>73 (83)</td>
<td>α-L-fucosidase (YP_003323631.1)</td>
<td>Thermobaculum terrenum</td>
</tr>
<tr>
<td>mfuc3</td>
<td>2025</td>
<td>60.6 %</td>
<td>41 (57)</td>
<td>beta-galactosidase trimerisation domain-containing protein YP_004464268.1</td>
<td>Mahella australiensis</td>
</tr>
<tr>
<td>mfuc4</td>
<td>1287</td>
<td>61.3 %</td>
<td>70 (81)</td>
<td>α-L-fucosidase (YP_001278245)</td>
<td>Roseiflexus sp. RS-1</td>
</tr>
<tr>
<td>mfuc5</td>
<td>1260</td>
<td>60.4 %</td>
<td>74 (85)</td>
<td>Putative α-L-fucosidase YP_004173881.1</td>
<td>Anaerolinea thermophila</td>
</tr>
<tr>
<td>mfuc6</td>
<td>1251</td>
<td>65.4 %</td>
<td>65 (81)</td>
<td>α-L-fucosidase YP_007147579.1</td>
<td>Cylindrospermum stagnale</td>
</tr>
<tr>
<td>mfuc7</td>
<td>1323</td>
<td>60.7 %</td>
<td>76 (86)</td>
<td>α-L-fucosidase WP_009252575.1</td>
<td>Lachnospiraceae bacterium</td>
</tr>
</tbody>
</table>

Table 3.

<table>
<thead>
<tr>
<th></th>
<th>Mfuc1</th>
<th>Mfuc2</th>
<th>Mfuc3</th>
<th>Mfuc4</th>
<th>Mfuc5</th>
<th>Mfuc6</th>
<th>Mfuc7</th>
<th>Thma WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH opt</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>pH interval</td>
<td>6-9</td>
<td>6-10</td>
<td>6-8</td>
<td>6-9</td>
<td>5-8</td>
<td>6-10</td>
<td>6-7</td>
<td>4-9</td>
</tr>
<tr>
<td>Temperature stability, T_{1/2}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 °C</td>
<td>&gt; 2 h</td>
<td>&gt; 2 h</td>
<td>&gt; 2 h</td>
<td>&gt; 2 h</td>
<td>&gt; 2 h</td>
<td>&gt; 2 h</td>
<td>~98 min</td>
<td>&gt; 2 h</td>
</tr>
<tr>
<td>40 °C</td>
<td>&gt; 2 h</td>
<td>&gt; 2 h</td>
<td>&gt; 2 h</td>
<td>&gt; 2 h</td>
<td>&gt; 2 h</td>
<td>~56 min</td>
<td>~28 min</td>
<td>&gt; 2 h</td>
</tr>
<tr>
<td>50 °C</td>
<td>&gt; 2 h</td>
<td>~9 min</td>
<td>~67 min</td>
<td>&gt; 2 h</td>
<td>&lt; 2 min</td>
<td>~14 min</td>
<td>~8 min</td>
<td>&gt; 2 h</td>
</tr>
<tr>
<td>60 °C</td>
<td>&lt; 2 min</td>
<td>&lt; 2 min</td>
<td>~4 min</td>
<td>&lt; 2 min</td>
<td>&lt; 2 min</td>
<td>&lt; 2 min</td>
<td>&lt; 2 min</td>
<td>&gt; 2 h</td>
</tr>
<tr>
<td>70 °C</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>&gt; 2 h</td>
</tr>
<tr>
<td>90 °C</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>&gt; 2 h</td>
</tr>
<tr>
<td>100 °C</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>~48 min</td>
</tr>
</tbody>
</table>
Table 4.

<table>
<thead>
<tr>
<th></th>
<th>Mfuc1</th>
<th>Mfuc2</th>
<th>Mfuc3</th>
<th>Mfuc4</th>
<th>Mfuc5</th>
<th>Mfuc6</th>
<th>ThmaFuc</th>
<th>L322P</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{max}$</td>
<td>1.6±0.0</td>
<td>2.3 ±0.081</td>
<td>1.6 ±0.090</td>
<td>0.77±0.017</td>
<td>1.9 ±0.04</td>
<td>2.2 ±0.16</td>
<td>0.67 ±0.02</td>
<td>0.47 ±0.01</td>
</tr>
<tr>
<td>$K_m$</td>
<td>0.11±0.01</td>
<td>0.14 ±0.01</td>
<td>0.31 ±0.03</td>
<td>0.071 ±0.01</td>
<td>0.13 ±0.01</td>
<td>0.28 ±0.05</td>
<td>0.06 ±0.01</td>
<td>0.07 ±0.01</td>
</tr>
<tr>
<td>$K_i$</td>
<td>1.7±1.6</td>
<td>7.1 ±0.9</td>
<td>2.6 ±0.3</td>
<td>30 ±6.8</td>
<td>12±1.2</td>
<td>12 ±3.4</td>
<td>22 ±4.9</td>
<td>32±9.5</td>
</tr>
</tbody>
</table>

Table 5.

<table>
<thead>
<tr>
<th></th>
<th>Mfuc1</th>
<th>Mfuc2</th>
<th>Mfuc3</th>
<th>Mfuc4</th>
<th>Mfuc5</th>
<th>Mfuc6</th>
<th>Mfuc7</th>
<th>Thma WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>2'-FL</td>
<td>76 ±1 %</td>
<td>88±1 %</td>
<td>2 ±1 %</td>
<td>47±1 %</td>
<td>87±1 %</td>
<td>2 ±2 %</td>
<td>94±1 %</td>
<td>71±1 %</td>
</tr>
<tr>
<td>3-FL</td>
<td>11 ±2 %</td>
<td>4±2 %</td>
<td>0 ±2 %</td>
<td>3±2 %</td>
<td>34±3 %</td>
<td>0 ±2 %</td>
<td>12±2 %</td>
<td>1 ±2 %</td>
</tr>
</tbody>
</table>

Table 6.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Amount (mM)</th>
<th>pNP-Fuc-(1-2)-α-L-fuc*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pNP-Fuc*</td>
<td>fuc</td>
</tr>
<tr>
<td>Mfuc1</td>
<td>0.1</td>
<td>20.3</td>
</tr>
<tr>
<td>Mfuc2</td>
<td>0.1</td>
<td>17.9</td>
</tr>
<tr>
<td>Mfuc3</td>
<td>0.0</td>
<td>18.3</td>
</tr>
<tr>
<td>Mfuc4</td>
<td>0.4</td>
<td>18.3</td>
</tr>
<tr>
<td>Mfuc5</td>
<td>0.5</td>
<td>12.9</td>
</tr>
<tr>
<td>Mfuc6</td>
<td>1.2</td>
<td>7.3</td>
</tr>
<tr>
<td>Mfuc7</td>
<td>4.6</td>
<td>12.3</td>
</tr>
<tr>
<td>Thma</td>
<td>1.0</td>
<td>13.9</td>
</tr>
<tr>
<td>L322P</td>
<td>1.3</td>
<td>13.0</td>
</tr>
</tbody>
</table>
Figures

Figure 1.

Figure 2
### Figure 3.

<table>
<thead>
<tr>
<th></th>
<th>pH 7</th>
<th></th>
<th>pH 9</th>
<th></th>
<th>pH 6</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>Thma</td>
<td>Mfuc1</td>
<td>Mfuc2</td>
<td>Mfuc4</td>
<td>Mfuc5</td>
</tr>
<tr>
<td><strong>Arabidopsis thaliana</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xyloglucan control (LM15)</td>
<td>23</td>
<td>21</td>
<td>23</td>
<td>22</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>Fucosylated xyloglucan (CCRC-M1)</td>
<td>32</td>
<td>28</td>
<td>0</td>
<td>5</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td><strong>Sambucus nigra</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xyloglucan control (LM15)</td>
<td>89</td>
<td>92</td>
<td>100</td>
<td>95</td>
<td>90</td>
<td>88</td>
</tr>
<tr>
<td>Fucosylated xyloglucan (CCRC-M1)</td>
<td>99</td>
<td>90</td>
<td>0</td>
<td>15</td>
<td>9</td>
<td>0</td>
</tr>
</tbody>
</table>

### Figure 4.

- **Mfuc5**
- **Mfuc6**
- **Thma**
- **L322P**
Figure 5.

Figure 6.
Table 1. List of primers used for amplification of α-L-fucosidase-encoding genes. Restriction sites are underlined, start and stop codons are indicated in bold, and mutagenic codon in bold and underlined.

Table 2. Properties of E. coli functional α-L-fucosidase genes identified from soil metagenomic library

Table 3. Properties of α-L-fucosidases derived from metagenomic library screening as well as α-L-fucosidase from Thermotoga maritima.

1 pH optimum was measured with 0.1 and 1 mM substrate; for Thma two optima (pH 4 and 7) were observed with 1 mM substrate.

2 pH interval where enzymes displayed more than 20% of maximal enzyme activity.

Table 4. Kinetic parameters of pNP-α-Fuc hydrolysis with metagenomic α-L-fucosidases at 30 °C and optimal pH as determined for pNP-Fuc. Kinetic constants were determined from the time-dependent release of pNP at 405 nm. Kinetic data for Mfuc6 were fitted using allosteric model \( Y=V_{\text{max}} * X^h / (K_{\text{prime}} + X^h) \) and the data for remaining enzymes with substrate inhibiton model \( Y=V_{\text{max}} * X / (K_{m} + X * (1+X/K_i)) \). Kprime is related to but not equal to Km. \( h \) is the Hill slope indicating the measure of cooperativity in substrate binding.

Table 5. Demonstration of hydrolytic regiospecificity of metagenome-derived α-L-fucosidases. The numbers represent the percentage of 1 mM initial substrate hydrolysed with 10μg of α-L-fucosidase within 150 min at 30 °C at optimal pH (50 mM phosphate-citrate buffer or glycine-NaOH for Mfuc6).

Table 6. Concentrations (mM) of substrates and products as determined using HPAEC-PAD and NMR.

*Concentrations (mM) estimated by comparison of characteristic non-overlapping \(^1\)H NMR signals and that of L-fucose. L-Fucose concentration in each sample was determined by HPAEC-PAD.

Figure 1. A phylogenetic analysis of discovered α-L-fucosidase sequences and selected representatives of GH-29 family. Neighbor-Joining method (62) was used to infer the phylogeny and evolutionary analyses were
conducted in MEGA6 (63). All positions containing gaps and missing data were eliminated. There were a total of 136 AA positions in the alignment. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (64).

Figure 2. SDS-PAGE of metagenome-derived (lanes 2-8) and benchmark α-L-fucosidases (lanes 9-10), all expressed in *E. coli* and purified by IMAC. Marker – lane 1.

Figure 3. Release of fucose from xyloglucans of selected plant species by α-L-fucosidases. As a control, xyloglucan was measured with the antiobody LM15, while fucosylated xyloglucan was measured using the antibody CCRC-M1.

Figure 4. Time course concentrations of L-fucose (squares), pNP (circles), pNP-Fuc (triangles facing down), 3-FL (diamonds), 2’-FL (triangles) in the transfucosylation reactions catalyzed by Mfuc5, and 6 as well as Thma wild type and L322P mutant. Reactions were done using 25 mM pNP-Fuc as donor and 100 mM lactose as acceptor at 30 °C and optimal pH.

Figure 5. HPAEC analysis of transfucosylation catalyzed with Mfuc5. The identity of pNP-Fuc-Fuc compounds was successfully determined after reverse phase mode separation into individual fractions and annotation by NMR and HPAEC-PAD.

Figure 6. Transglycosylation reactions catalyzed by tested metagenomic α-L-fucosidases and α-L-fucosidase from *T. maritima*. Additionally to hydrolysis products: L-fucose and 4-nitrophenol, the formation of self-condensation products 1: pNP-Fuc-(1-2)-Fuc, 2: pNP-Fuc-(1-3)-Fuc, 3: pNP-Fuc-(1-4)-Fuc and fucosyllactoses was catalyzed by α-L-fucosidases to various degree.
Novel α-L-fucosidases from soil metagenome - identification and characterization of their hydrolytic and tranglycosylation abilities

Supplementary material

Authors
Mateusz Lezyk\textsuperscript{a}, Carsten Jer\textsuperscript{a}, Louise Kjaerulff\textsuperscript{b}, Charlotte H. Gotfredsen\textsuperscript{b}, Maria D. Mikkelsen\textsuperscript{c}, Jørn D. Mikkelsen\textsuperscript{a,*}

Address
\textsuperscript{a} Center for BioProcess Engineering, Department of Chemical and Biochemical Engineering, Building 229, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
\textsuperscript{b} Department of Chemistry, Building 201, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
\textsuperscript{c} Department of Plant and Environmental Sciences, University of Copenhagen, 1871 Copenhagen, Denmark

*To whom correspondence should be addressed:
Jørn D. Mikkelsen
jdm@kt.dtu.dk
Tel: +4545252938

This supplement contains:
Figure S1. pNP-Fuc hydrolysis vs. pH
Figure S2. Thermostability evaluation of tested fucosidases
Figure S3. Sequence of \textit{E. coli}-codon optimized Thma wildtype fucosidase gene
Figure S4. Time course of trans-fucosylation reactions performed with lactose as acceptor
Figure S5. Substrate saturation curves for tested fucosidases
Figures S6.1-6.10 Selected 1H NMR spectra
Figure S7. NMR chemical shifts of selected compounds characterized in this study
Figure S8. Comparison of 1H NMR integrals for reactions catalyzed by different α-L-fucosidases
Figure S1. pH optimum for metagenomic fucosidases. The activity of the enzymes was tested at indicated pH and initial reaction rates were measured. The buffers used were 50 mM of phosphate-citrate (pH 3-8; blue diamonds), tricine (pH 8-9; red squares); glycylglycine (pH 8-9; green triangles), and glycine-NaOH (pH 9-10; purple crosses). Enzyme concentrations used in assays with 0.1 mM substrate were as follows: 0.2 μg/mL Mfuc1, 0.2 μg/mL Mfuc2, 0.6 μg/mL Mfuc3, 0.3 μg/mL Mfuc4, 0.2 μg/mL Mfuc5, 1.0 μg/mL Mfuc6, 0.2 μg/mL Mfuc7, 0.3 μg/mL Thma fucosidase, 0.5 μg/mL Thma L322P. With exception of Mfuc2, the double concentration of enzyme was used in assays with 1 mM substrate. All reactions were done at 30 °C.
Figure S2. Thermostability of metagenomic fucosidases. The enzymes were diluted in a 5 mM pH buffer reflecting pH optimum and incubated at indicated temperatures and periods of time after which residual activity was measured in a continuous assay at pH 7 at 30 °C. Enzyme concentrations used in assays with 0.1 mM substrate were as follows: 0.8 μg/mL Mfuc1, 0.4 μg/mL Mfuc2, 1.2 μg/mL Mfuc3, 0.9 μg/mL Mfuc4, 0.7 μg/mL Mfuc5, 3.9 μg/mL Mfuc6, 0.5 μg/mL Mfuc7, 0.9 μg/mL Thma fucosidase, 1.5 μg/mL Thma L322P. Activities were normalized on basis of enzyme activity with no heat treatment.
Figure S3. Sequence of *E. coli* codon optimized sequence of *Thermotoga maritima* α-L-fucosidase gene used in this study
Figure S4. Time course of reactions catalyzed by metagenome-derived fucosidases where 25mM of pNP-α-L-fucose was used as donor and 100mM of lactose as acceptor. L-fucose (squares), pNP (circles), pNP-Fuc (triangles facing down), 3-FL (diamonds), 2'-FL (triangles).
mfuc1

mfuc2

mfuc3

mfuc4

mfuc5

mfuc8
Figure S5. Substrate saturation curves for metagenomic fucosidases along with non-linear regression fits with sigmoidal allosteric (Mfuc8; \( Y = V_{\text{max}} X^{h}/(K_{\text{prime}} + X^{h}) \)) and substrate inhibition (all except Mfuc6, \( Y = V_{\text{max}} X/(K_{m} + X(1+X/K_{i})) \)) model. The various enzymes were assayed at 30 °C at their identified optimum pH. In case of Thma WT, the optimum pH was 7 at 0.1 mM pNP-Fuc while two optima (pH 4 and 7) were observed at 1 mM pNP-Fuc, consequently substrate saturation was tested at both pH. Enzyme concentrations used were as follows: 0.2 μg/mL Mfuc1, 0.03 μg/mL Mfuc2, 0.6 μg/mL Mfuc3, 0.3 μg/mL Mfuc4, 0.2 μg/mL Mfuc5, 1.0 μg/mL Mfuc6, 0.2 μg/mL Mfuc7, 0.3 μg/mL Thma fucosidase, 0.5 μg/mL Thma L322P. The concentrations of pNP-Fuc tested were 0.1, 0.25, 0.5, 1.0, 2.5, and 5.0 mM.
Figure S6. Selected 1H NMR spectra acquired in this study

Figure 6.1. $^1$H NMR spectrum (D$_2$O, 400 MHz) of $p$-nitrophenyl $\alpha$-L-fucose.

Figure 6.1. $^1$H NMR spectrum (D$_2$O, 500 MHz) of post-enzymatic reaction mixture by Mfuc1 enzyme with
expansion of the α-anomeric region.

Figure 6.2. $^1$H NMR spectrum (D$_2$O, 500 MHz) of post-enzymatic reaction mixture by Mfuc2 enzyme with expansion of the α-anomeric region.

Figure 6.3. $^1$H NMR spectrum (D$_2$O, 500 MHz) of post-enzymatic reaction mixture by Mfuc3 enzyme with expansion of the α-anomeric region.
Figure 6.4. $^1$H NMR spectrum (D$_2$O, 500 MHz) of post-enzymatic reaction mixture by Mfuc4 enzyme with expansion of the $\alpha$-anomeric region.

Figure 5.6. $^1$H NMR spectrum (D$_2$O, 500 MHz) of post-enzymatic reaction mixture by Mfuc5 enzyme with expansion of the $\alpha$-anomeric region.
Figure 6.6. $^1$H NMR spectrum (D$_2$O, 500 MHz) of post-enzymatic reaction mixture by Mfuc6 enzyme with expansion of the $\alpha$-anomeric region. The residual solvent signal was removed by presaturation.

Figure 6.7. $^1$H NMR spectrum (D$_2$O, 500 MHz) of post-enzymatic reaction mixture by Mfuc7 enzyme with expansion of the $\alpha$-anomeric region. The residual solvent signal was removed by presaturation.
Figure 6.8. $^1$H NMR spectrum (D$_2$O, 500 MHz) of post-enzymatic reaction mixture by *T. maritima* wildtype enzyme with expansion of the $\alpha$-anomeric region.

Figure 6.9. $^1$H NMR spectrum (D$_2$O, 500 MHz) of post-enzymatic reaction mixture by the *T. maritima* L322P enzyme with expansion of the $\alpha$-anomeric region.
Figure 6.10. $^1$H NMR spectra (D$_2$O, 400 MHz) of individual self-condensation products separated by HPLC and comparison with products of selected enzymatic reactions. From the top: $p$-nitrophenyl α-L-fucosyl-(1-4)-α-L-fucose, $p$-nitrophenyl α-L-fucosyl-(1-3)-α-L-fucose, $p$-nitrophenyl α-L-fucosyl-(1-2)-α-L-fucose, reaction catalyzed by Mfuc7, and reaction catalyzed by Mfuc6. The residual solvent signals were (partially) removed by presaturation.

$p$-nitrophenyl α-L-fucosyl-(1-2)-α-L-fucose

$^1$H NMR: δ 8.26 (d, 2H, $J = 8.8$ Hz, H3’’), 7.28 (d, 2H, $J = 8.8$ Hz, H2’’), 5.84 (d, 1H, $J = 2.8$ Hz, H1), 5.11 (d, 1H, $J = 3.9$ Hz, H1’), 4.34 (br. q, 1H, $J = 6.5$ Hz, H5’), 4.14 (m, 1H, H2), 4.13 (m, 1H, H5), 4.11 (m, 1H, H3), 4.08 (m, 1H, H4), 3.97 (dd, 1H, $J = 3.5$ Hz, 10.5 Hz, H3’), 3.83 (m, 1H, H4’), 3.82 (m, 1H, H2’), 1.23 (d, 3H, $J = 6.5$ Hz, H6’), 1.15 (d, 3H, $J = 6.7$ Hz, H6). $^{13}$C NMR derived from gHSQC and gHMBC: δ 162.2 (C1’’), 143.0 (C4’’), 126.9 (2*C3’’), 117.6 (2*C2’’), 97.6 (C1), 96.3 (C1’), 75.3 (C2), 72.7 (C4’), 70.3 (C3’), 68.7 (C4), 68.8 (C2’), 68.6 (C3), 67.9 (C5’), 66.7 (C5), 16.3 (C6’), 16.2 (C6).
\textit{p}-nitrophenyl $\alpha$-L-fucosyl-(1-3)-$\alpha$-L-fucose

$^1$H NMR: $\delta$ 8.26 (d, 2H, $J = 9.0$ Hz, H3’’), 7.28 (d, 2H, $J = 9.0$ Hz, H2’’), 5.98 (d, 1H, $J = 3.5$ Hz, H1), 5.04 (d, 1H, $J = 4.0$ Hz, H1’), 4.28 (br. q, 1H, $J = 6.7$ Hz, H5’), 4.21 (dd, 1H, $J = 3.4$ Hz, 10.4 Hz, H3), 4.14 (br. q, 1H, $J = 6.7$ Hz, H5), 4.04 (dd, 1H, $J = 3.5$ Hz, 10.4 Hz, H2), 3.92 (m, 1H, H3’), 3.91 (m, 1H, H4), 3.80 (br. d, 1H, $J = 3.3$ Hz, H4’), 3.72 (dd, 1H, $J = 4.0$ Hz, 10.2 Hz, H2’), 1.21 (d, 3H, $J = 6.7$ Hz, H6’), 1.15 (d, 3H, $J = 6.7$ Hz, H6). $^{13}$C NMR derived from gHSQC and gHMBC: $\delta$ 162.5 (C1’’), 143.2 (C4’’), 126.9 (2*C3’’), 117.7 (2*C2’’), 97.0 (C1’), 95.1 (C1), 72.7 (C2), 72.5 (C4), 72.3 (C4’), 70.1 (C3’), 68.8 (C3), 68.6 (C2’), 67.9 (C5’), 66.7 (C5), 16.3 (C6’), 16.1 (C6).

\textit{p}-nitrophenyl $\alpha$-L-fucosyl-(1-4)-$\alpha$-L-fucose

$^1$H NMR: $\delta$ 8.26 (d, 2H, $J = 9.2$ Hz, H3’’), 7.27 (d, 2H, $J = 9.2$ Hz, H2’’), 5.83 (d, 1H, $J = 3.9$ Hz, H1), 4.95 (d, 1H, $J = 4.0$ Hz, H1’), 4.54 (br. q, 1H, $J = 6.6$ Hz, H5), 4.20 (m, 1H, H3), 4.18 (m, 1H, H5’), 4.04 (dd, 1H, $J = 3.9$ Hz, 10.6 Hz, H2), 3.94 (dd, 1H, $J = 3.0$ Hz, 10.8 Hz, H3’), 3.90 (br. d, 1H, $J = 3.0$ Hz, H4’), 3.83 (br. d, 1H, $J = 3.0$ Hz, H4), 3.80 (dd, 1H, $J = 4.0$ Hz, 10.8 Hz, H2’), 1.23 (d, 3H, $J \sim 7$ Hz, H6’), 1.17 (d, 3H, $J = 6.6$ Hz, H6). $^{13}$C NMR derived from gHSQC and gHMBC: $\delta$ 162.3 (C1’’), 143.1 (C4’’), 126.8 (2*C3’’), 117.6 (2*C2’’), 101.3 (C1’), 97.7 (C1), 80.5 (C4’), 70.2 (C3’), 69.7 (C3), 69.5 (C2’), 69.5 (C5’), 68.8 (C4), 68.4 (C2), 67.7 (C5), 16.2 (C6’), 16.2 (C6).

Figure 7. NMR chemical shifts of selected compounds characterized in this study.
Figure 8. Comparison of $^1$H NMR integrals for reaction mixtures of the different enzymes. Representative integrals of selected molecules were used to compare relative amounts within the NMR samples. For all samples the Gal anomeric resonance in lactose was calibrated to 1, which enables some comparison between samples, however with relatively high uncertainty. Anomeric resonances were used for integration. For the pNP-Fuc-Fuc products (1-2-, 1-3-, and 1-4-linked) the anomeric signal at the reducing end was used, for Fuc, the $\alpha$- and $\beta$-anomeric resonances were added together.