

NMR approaches to understand synthesis,  
assembly and properties of novel  
carbohydrate-based materials



Valeria Gabrielli

University of East Anglia  
School of Pharmacy  
May 2021

Thesis submitted to the School of Pharmacy, University of East Anglia in fulfilment of the  
requirement for the degree of Doctor of Philosophy.

*To my grandfather, Manlio, and to my friend, Mithun*

*“Walnut tree?”*

*Norwich’s Survivals*

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include the full attribution

The doctoral work presented in this thesis has been conducted prevalently at the University of East Anglia (UEA) - School of Pharmacy - and at the John Innes Centre, which are both part of the Norwich Research Park. The PhD was sponsored by the BBSRC, during the period from 2016 to 2020 via bioDTP at Norwich Research Park. The BBSRC also financed an abroad professional internship, undertaken in the research group of Prof. Bernhard Blümich – Institute for Technical und Macromolecular Chemistry, RWTH Aachen University.

Some of the NMR experiments have been performed at the UK High-Field solid-state NMR National Research Facility in the University of Warwick (Coventry), where we were able to use the 850 MHz solid-state spectrometer. Our experimental time in the University of Warwick was sponsored by the EPSRC and BBSRC (contract reference PR140003), as well as the University of Warwick.

# Acknowledgements

Here comes the time to thank all the people that made this journey possible. It has been long, exciting, at times exhausting, challenging, always interesting and full of learning experiences – and I am extremely grateful to each one who contributed (in large or small, with a smile or a criticism). Of one thing I am sure in life, I could not have done this alone.

I would like to first thank my supervisory team – Prof. Yaroslav Khimyak, Dr Jesús Angulo and Prof. Rob Field. Yaroslav and Jesús have been a real scientific inspiration to me. They showed passion in any aspect of life-science (and beyond), perseverance in their work, and a positive attitude in overcoming challenging situations. They did not just entrust me in this PhD, they also provided the perfect environment to get the best out of it. I will never forget the day they came to the lab upstairs, together, to give me the news that I passed the interview with the BBSRC external panel, and I was actually going to start a PhD with them in few months. As well, I will never forget when they told me they were awarded with an (amazing) EPSRC grant, and a post-doc would work side-by-side with me.

This post-doc was Dr Juan Carlos Muñoz-Garcia, who I consider my real “scientific father”. Juan taught me a lot of things (everything he could, I guess). He supported me during the PhD dark times, spending hours and hours in an office without windows and always in front of a screen, with data difficult to interpret at first glance and some hard concepts to understand. From these years working together, I did not find only a colleague but a mentor and a friend.

Prof. Rob Field has been for me an extreme example. In my eyes, he combines a humble and altruistic attitude, approachability, availability and extreme smartness. I was impressed in observing how he could resolve every situation with just a few words and how everyone relied on him. I take the opportunity to also thank his research group, especially Peterson, Giulia, Sergey, Irina and Martin, for the help and support and the friendly environment they created around me.

During the PhD, I had the immense pleasure to collaborate with the GelEnz group. I would like to thank Prof. Steve Eichorn, Prof. Janet Scott and Prof. Karen Edler, as well as Rinat, Marcus, Vincenzo, Julien, Marcelo and Zakir for sharing with me part of this journey and to teach me science out of my area of expertise.

In addition, I would like to thank Prof. Bernhard Blümich for welcoming me in his research group for a three months internship, and to Chris, Jenny, Till, Anton, Denis and Alex. Prof. Blümich showed me how to combine two of my great passions, science and art, and that one does not exclude the other.

A special thanks goes to all the people that have been, are currently and will still be in our office in the basement. Susana and Karol, my friends even before starting the PhD. Serena, with whom I shared the house, the office, tears and laughs and so many life-changing events – we definitely grew up in this time together (and by the way, thanks for telling me “brava” every once in a while, it really boosted my motivation). Vani - the most hard-working person I ever met– Sam, Alex, Tom, Trey, Inaki, Kika, everyone contributed in making a special environment in the office, and Matthew, who was an extreme support to me especially in this final stage.

I would also like to thank Dr Colin Macdonald for his help with the NMR and TEM instrumentations, Dr Bertrand Lézé for his assistance with PXRD and SEM and Dr Dinu Iuga for his experimental advice on NMR pulse sequences. I would like to acknowledge the University of East Anglia and the BBSRC for the PhD scholarship, and the UK 850 MHz Solid-State NMR Facility at the University of Warwick.

Norwich is the place where I lived the longest besides home and has become a kind of home itself. Still, home is not made by places, but by people. Norwich became my home with the “Norwich’s survival” group, with whom I shared everything. It started with Matina and a bit of enthusiasm for a night out, and it grew day after day in what is today a network of friends spread around Europe and the world. I think it is impossible to summarise in few words what everyone means to me – they are just too much. I will never forget the training in the circus with Angelina, the thousands of nights spent at Matina and Sinem’s, the morning coffees with Sylvain, the Fridays at the REC centre (with almost everyone), the hours in the allotment with Ferran, Rapha and Victor, the infinite nights out, dinners, extravagant activities and the exploration of Norfolk with Sam and my a-year-and-a-half here with Baptiste.

Of course, a special thanks goes to Baptiste, not for all the things we did together, but for one little and extremely important thing: Baptiste, thanks for encouraging me in improving myself and become a better person, every day.

Marco, Rosa, Giannis, Ross, Anton, Jess, Gaëtan, Eirini, Lila, Elias, Philip, Pawel, Clémence, Thorsten, Florian, Estela, Sara, Vicenç, Leonardo, Gaetano, Marta, Elena, Gabriele, Alberto, Vlado, Natasha, Ana Victoria, Giuseppe, Maria, Nuria, Maui, Megan, Melissa, Lučka, Mar, Gerard, Noelia, Vera, Andrew, Elise, Hui, Ilaria, Paulo, Arnaud, Francesc, Remy, Cholpon – all of you are part of this experience, and all of you have a special place in my hearth.

I want to thank my family and friends from Italy. My grandparents Manlio and Vera, for giving me unconditional love and growing me up to the person I am today. My parents, for making of me a strong and independent woman, and my brothers Francesco and Mattia, one for being an example to follow, the other for giving me the challenge of being a good example. My “University friends” – Alessandra, Claudia, Ana and Ina – with whom I started this adventure of going around the world. My “Erasmus friends” – Marco, Wilma and Martina - who became my first abroad family, and with whom I keep meeting always in different countries. My closest friends Irene, Ilaria, Giulia, Mauro, Michele and Sergio, I guess the people who know me the best, that have always been at my side and always will be. Thank you all for giving me a home to come back to and a place where I belong.

# Abstract

The investigation conducted throughout this doctoral thesis covers the underpinning science of synthesis, assembly and characterisation of novel polysaccharide-based materials. Within the broad chemical family of "sugars", polysaccharides are excellent biopolymers for hydrogel preparation. Innovative glycobiology techniques such as enzymatic catalysis can be used for the functionalisation of these materials and the introduction of specific and tuneable properties.

Cellodextrin phosphorylase from *Ruminiclostridium thermocellum* (CDP) is a carbohydrate-active enzyme able to catalyse the synthesis of small fragments of cellulose. The enzyme's loose specificity towards donor and acceptor substrates allows its exploitation for the introduction of functional groups in a regiocontrolled manner. Therefore, understanding the structural details of substrates molecular recognition is fundamental for a rational design of novel cellulose derivatives.

We applied several NMR approaches and computational tools to unveil the mechanistic details of substrates recognition. The investigation highlighted differences in the binding epitopes of cognate and non-cognate donor and acceptor substrates, and the enthalpic contribution in the energy of binding played by inorganic phosphate.

CDP has been used for the synthesis of a series of cellodextrin derivatives decorated with fluorine atoms. We combined state-of-the-art NMR spectroscopy (solution and solid-state) methodologies with PXRD technique to gain the molecular details of the novel materials self-assembly. In addition, we developed an NMR-based strategy to discern between surface/disordered and core/ordered domains.

Great advances in polymer preparation for hydrogels needs to be followed by an understanding of the processes of gelation. Importantly, little is known about how the interactions between the gel network and the surrounding water control the gelation process. We developed a solution state NMR methodology to monitor water (and co-solvents) in hydrogel systems. The methodology is based on STD NMR, a ligand-based approach to monitor water-surface interactions. Our new protocol, called Spin Diffusion Transfer Difference (SDTD), enabled us to uncover changes in water structuration in hydrogels, giving insight into the role played by solvent in gelation.

## **Access Condition and Agreement**

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

# Publications generated throughout the PhD studies

- 1) De Andrade P., Muñoz-Garcia J. C., Pergolizzi G., Gabrielli V., Nepogodiev S., Iuga D., Fabian L., Nigmatullin R., Johns M. A., Harniman R., Eichhorn S. J., Angulo J. A., Khimyak Y. Z., and Field R. A. – *Chemoenzymatic synthesis of fluorinated cellodextrins identifies a new allomorph for cellulose-like materials* – Chemistry, A European Journal, **2020**. DOI: 10.002/chem.202003604
- 2) Gabrielli V., Kuraite A., Alves da Silva M., Elder K. J., Angulo J. A., Nepravishta R., Muñoz-Garcia J. C., and Khimyak Y. Z. - *Spin Diffusion Transfer Difference (SDTD) NMR: An Advanced Method for the Characterisation of Water Structuration Within Particle Networks* - ChemRxiv. **2020** Preprint. DOI: 10.26434/chemrxiv.12770813.v1
- 3) Blümich B., Baias M., Rehorn C., Gabrielli V., Jaschtschuk D., Harrison C., Invernizzi C., and Malagodi M. - *Comparison of violin back plates by nondestructive MRI depth profiling* - Microchemical Journal, **2020**, 158, 105219
- 4) Nigmatullin R., Johns M. A., Muñoz-Garcia J. C., Gabrielli V., Schmitt J., Angulo J. A., Khimyak Y. Z., Scott J. L., Edler K. J., and Eichhorn S. J. - *Hydrophobization of cellulose nanocrystals for aqueous colloidal suspensions and gels* - Biomacromolecules, **2020** 21, 5, 1812–1823
- 5) Invernizzi C., Fiocco G., Iwanicka M., Kowalska M., Targowski P., Bluemich B., Rehorn C., Gabrielli V., Bersani D., Licchelli M., and Malagodi M. - *Non-invasive mobile technology to study the stratigraphy of ancient Cremonese violins: OCT, NMR-MOUSE, XRF and reflection FT-IR spectroscopy* - Microchemical Journal, **2020**, 155, 104754
- 6) Muñoz-Garcia J. C., Corbin K., Hussain H., Gabrielli V., Koev T., Iuga D., Round A. N., Mikkelsen D., Gunning P. A., Warren F. J. and Khimyak Y. Z. - *High molecular weight mixed-linkage glucan as a mechanical and hydration modulator of Bacterial Cellulose: Characterization by Advanced NMR Spectroscopy* - Biomacromolecules, **2019**, 20, 4180 - 419

- 7) Nigmatullin R., Gabrielli V., Muñoz-Garcia J. C., Lewandowska A. E., Harniman R., Khimyak Y. Z., Angulo J. A and Eichhorn S. J. - *Thermosensitive supramolecular and colloidal hydrogels via self-assembly modulated by hydrophobized cellulose nanocrystals* –Cellulose, **2019**, 26, 529-542
- 8) Calabrese V., Da Silva M., Schmitt J., Muñoz-Garcia J., Gabrielli V., Scott J. L., Angulo J., Khimyak Y. Z., Edler K. J. - *Surfactant controlled zwitterionic cellulose nanofibrils dispersion* – Soft Matter, **2018**, 14, 7793-7800
- 9) Nigmatullin R.; Harniman R.; Gabrielli V., Muñoz-Garcia J. C., Khimyak Y. Z., Angulo J. A., and Eichhorn S. J. - *Mechanically robust gels formed from hydrophobized cellulose nanocrystals* – ACS Applied Material & Interfaces, **2018**, 10, 19318 – 19322

## Posters and Presentation at conferences

- Oral presentation webinar RSC Carbohydrate Group, Early Career Researcher Webinar Series, 2020 – SDTD: a novel solution state NMR methodology to unveil water structuration in hydrogel systems
- Poster at Pharmacy Research Colloquium – UEA, 2020: Unveiling the role of solvent in gelation via a novel solution state NMR protocol (Norwich, UK)
- Oral communication and poster presentation at EUROCARB, 2019: Insight into substrate binding in  $\beta$ -1,3 and  $\beta$ -1,4-glucan phosphorylases by STD NMR spectroscopy (Leiden, ND)
- Poster at International Carbohydrate Symposium (ICS) 2018: Unveiling multiple binding modes for Cellodextrin Phosphorylase (Lisbon, PO)
- Poster at RSC Chemical Biology and Bio-Organic Group Forum 2018: Molecular recognition of substrates by Cellodextrin Phosphorylase studied by Differential Epitope Mapping STD NMR (DEEP-STD NMR) (Norwich, UK)

## Awards

- RSC Advances Outstanding Presentation Prize at RSC Carbohydrate Group, Early Career Researcher Webinar Series

## Abbreviation and Acronyms

$\Delta$ DEEP-STD	DEEP-STD factor
$\tau_c$	Correlation time
$\tau_m$	Mixing time
$f_{LB}$	fraction of bound ligand
$f_{WB}$	fraction of bound water
2F-EpC	2-deoxy-2-fluoro Enzymatically Produced Cellodextrin
3F-EpC	3-deoxy-3-fluoro Enzymatically Produced Cellodextrin
6F-EpC	6-deoxy-6-fluoro Enzymatically Produced Cellodextrin
6F-Glc-1-P	$\alpha$ -D-6F-Flucose-1-Phosphate
2-PrOD	D <sub>2</sub> O-exchanged 2-Propanol
AFM	Atomic Force Microscopy
AGU	AnhydroGlucopyranose Units
B <sub>0</sub>	External magnetic field
CAZy	Carbohydrate-Active enZymes database
CBM	Carbohydrate Binding Module
CBP	CelloBiose phosphorylase
CDP	CelloDextrin phosphorylase
COSY	COrrelation SpectroscopyY
CP/MAS	Cross-Polarisation at Magic-Angle Spinning

CS	Corn Starch
CSA	Chemical Shift Anisotropy
D <sub>2</sub> O	Deuterium Oxide
DEEP-STD	DiffERential Epitope mapping Saturation Transfer Difference
DEPT	Distortionless Enhancement by Polarisation Transfer
DP	Degree of Polymerisation
DSC	Differential Scanning Calorimetry
DQF-COSY	Double Quantum Filter Correlation Spectroscopy
EpC	Enzymatically produced cellodextrin
EtOD	D <sub>2</sub> O-exchanged Ethanol
FID	Free Induction Decay
FWHM	Full Width Half Maximum
G'	Storage modulus
G''	Loss modulus
Gal-1-P	α-D-Galactose-1-Phosphate
Glc-1-P	α-D-Glucose-1-Phosphate
GIN-1-P	α-D-Glucosamine-1-Phosphate
GHs	Glycoside hydrolases
GLIDE	Grid-Based Ligand Docking with Energetics
GPs	Glycoside Phosphorylases
GT	GlycosylTransferase
HR-MAS	High-Resolution Magic-Angle Spinning

HSQC	Heteronuclear Single Quantum Coherence
IFD	Induced Fit Docking
ILOE	Inter-Ligand NOESY
ITC	Isothermal Titration Calorimetry
J	Dipole-Dipole coupling
$K_D$	Dissociation binding constant
LMWG	Low-Molecular Weight Gelators
Man-1-P	$\alpha$ -D-mannose-1-phosphate
MeOD	D <sub>2</sub> O-exchanged Methanol
MSD	Mean-Square Displacement
Multi 6F-EpC	Multiply 6-deoxy-6-fluoro Enzymatically Produced Cellodextrin
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
OCNF	Oxidised Cellulose nanofibrils
OPLS	Optimised Potential for Liquid Simulations
PBS	Phosphate-Buffered Saline
PDB	Protein Data Bank
PFG	Pulsed-Field Gradient
PGs	Polymeric Gelators
ppm	part per milion

PXRD	Powder X-Ray Diffraction
QM	Quantum Mechanics
rf	radiofrequency
S/N ratio	Signal-to-Noise ratio
SDTD	Spin Diffusion Saturation Transfer
SPR	Surface Plasmon Resonance
SQ	Single-Quantum
STD	Saturation Transfer Difference
STD-AF	Saturation Transfer Difference – Amplification Factor
ssNMR	Solid-state Nuclear Magnetic Resonance
$T_1$	Longitudinal relaxation time
$T_2$	Transverse relaxation time
TEM	Transmission Electron Microscopy
TMS	TetraMethylSilane
Tris	Tris(hydroxymethyl)aminomethane
$t_{\text{sat}}$	Saturation time
w/V	Weight per volume
wt%	Weight percentage
TOCSY	Total Correlated Spectroscopy
Tr-NOESY	Exchange transferred-NOESY
ZQC	Zero Quantum Coherence

## Table of Contents

<b>1. Introduction: Carbohydrates – from biological systems to building blocks for novel materials</b>	<b>2</b>
1.1 Objectives of this thesis.....	2
1.2 Introduction.....	3
1.2.1 <i>A multidisciplinary approach to science</i> .....	3
1.2.2 <i>Chemistry of carbohydrates</i> .....	4
1.3 Challenges in the synthesis of carbohydrates.....	10
1.3.1 <i>Enzymatic approaches for carbohydrates synthesis</i> .....	11
1.4 Carbohydrates-protein interactions – the molecular recognition event.....	15
1.4.1 <i>The enthalpy-entropy compensation effect</i> .....	15
1.4.2 <i>Specific binding patterns, role of water and hydrophobic effect for carbohydrate recognition in the enzyme catalytic cleft</i> .....	18
1.5 Cellulose .....	20
1.5.1 <i>Cellulose polymorphism</i> .....	22
1.6 Hydrogels.....	24
1.6.1 <i>Definition of gel</i> .....	24
1.6.2 <i>Classification of gels</i> .....	25
<b>2. Techniques and tools .....</b>	<b>28</b>
2.1 NMR Spectroscopy .....	30
2.1.1 <i>Solution state NMR<sup>67-69</sup></i> .....	30
2.1.2 <i>Ligand-based NMR spectroscopy for binding studies</i> .....	46
2.2 Solid-state NMR (ssNMR) <sup>85</sup> .....	62
2.2.1 <i>Chemical shielding and Chemical Shift Anisotropy</i> .....	63
2.2.2 <i>Homonuclear Dipolar Coupling</i> .....	65
2.2.3 <i>Heteronuclear Dipolar Coupling</i> .....	67
2.2.4 <i>Magic Angle Spinning</i> .....	67
2.2.5 <i>Heteronuclear decoupling</i> .....	69
2.2.6 <i>Homonuclear decoupling</i> .....	70
2.2.7 <i>Single pulse excitation <sup>13</sup>C NMR</i> .....	70
2.2.8 <i><sup>1</sup>H-<sup>13</sup>C CP/MAS NMR<sup>85</sup></i> .....	71
2.2.9 <i>Water-polarisation transfer – characterisation of water-network interactions in hydrogel systems by ssNMR</i> .....	77

2.3	Powder X-Ray Diffraction .....	78
2.4	Molecular modelling <sup>102</sup> .....	82
2.4.1	<i>Molecular Mechanics</i> .....	83
2.4.2	<i>Protein-ligand docking calculations</i> <sup>103, 104</sup> .....	84
<b>3.</b>	<b>Cellodextrin phosphorylase from <i>Clostridium thermocellum</i>: structural studies on substrates recognition</b> .....	<b>91</b>
3.1	Introduction.....	91
3.1.1	<i>Principal aims of the work</i> .....	91
3.1.2	<i>Importance of phosphorylases in carbohydrate synthesis</i> .....	92
3.1.3	<i>Cellodextrin phosphorylase</i> .....	97
3.2	Experimental section.....	106
3.2.1	<i>Structural basis of molecular recognition of natural and non-cognate donor substrates</i> .....	106
3.2.2	<i>Structural basis of molecular recognition of acceptor and acceptor-like substrates</i> .....	115
3.2.3	<i>Impact of phosphate on the binding of acceptors</i> .....	125
3.2.4	<i>Structural details of the CDP-donor-acceptor ternary complex interactions</i> ...	131
3.3	Final discussion and future prospective .....	136
3.4	Material and methods.....	138
3.4.1	<i>CDP expression and purification</i> .....	138
3.4.2	<i>Nuclear Magnetic Resonance</i> .....	139
3.4.3	<i>Molecular Docking:</i> .....	143
<b>4.</b>	<b>Deoxyfluorinated cellodextrin derivatives: structural characterisation</b> .....	<b>146</b>
4.1	Introduction.....	146
4.1.1	<i>Principal aims of the work</i> .....	146
4.1.2	<i>On the importance of fluorine functionalised cellulose</i> .....	147
4.1.3	<i>Combining solution and solid-state NMR approaches for material characterization</i> .....	150
4.1.4	<i>Determination of structure of cellulose by NMR spectroscopy</i> .....	153
4.2	Experimental section.....	158
4.2.1	<i>Background on the synthesis of cellodextrin and cellodextrin fluorinated derivatives</i> .....	158
4.2.2	<i>Deoxyfluorinated cellodextrin derivative multi-scale characterisation</i> .....	161
4.3	Discussion and future prospective .....	173

4.4	Material and methods.....	175
4.4.1	General materials and methods .....	175
4.4.2	Synthesis of deoxyfluorinated-cellodextrin derivatives .....	176
4.4.3	Analytical characterisation.....	180
<b>5.</b>	<b>SDTD: a novel methodological tool to study solvent in 3D gels network .....</b>	<b>185</b>
5.1	Introduction.....	185
5.1.1	Principal aims of the work .....	185
5.1.2	The role of water in hydrogels .....	186
5.1.3	NMR for solvent investigation in gels .....	190
5.1.4	STD NMR applied to hydrogel systems .....	194
5.1.5	A new methodological tool: Spin Diffusion Transfer Difference (SDTD) NMR.....	196
5.2	Experimental section.....	204
5.2.1	Model validation.....	205
5.2.2	The effect of solvent concentration on STD and SDTD build-up curves.....	207
5.2.3	The effect of gelator concentration on STD and SDTD build-up curves: SDTD reports on changes in the degree of solvent structuration.....	209
5.2.4	SDTD NMR characterisation of the role of co-solvents on the alcohol-induced gelation of OCNF hydrogels .....	213
5.3	Discussion and future prospective .....	219
5.4	Material and methods.....	221
5.4.1	Sample preparation.....	221
5.4.2	Nuclear magnetic resonance (NMR) spectroscopy .....	222
5.4.3	Simulation of the SDTD build-up curves.....	223
<b>6.</b>	<b>Final Remarks .....</b>	<b>226</b>
6.1	Cellodextrin phosphorylase from Clostridium thermocellum: structural studies on substrates recognition.....	228
6.1.1	Molecular recognition of donor and donor-like substrates: .....	229
6.1.2	Molecular recognition of acceptor and acceptor-like substrates: .....	230
6.1.3	Additional notes.....	231
6.2	Deoxyfluorinated cellodextrin derivatives: structural characterisation.....	232
6.3	SDTD: a novel NMR tool to study solvent (and co-solvents) in 3D gels network ....	234
	<b>References .....</b>	<b>237</b>
	<b>Appendix .....</b>	<b>I</b>

# Chapter 1

## Introduction

Carbohydrates - from biological systems to building blocks for novel materials

# 1. Introduction: Carbohydrates – from biological systems to building blocks for novel materials

## 1.1 Objectives of this thesis

The work presented in this thesis is the outcome of a fruitful and exciting collaboration between multiple partners across different institutions and disciplines. Since the really beginning of my PhD, I have been participating in the GelEnz scientific platform, which involves academics and industrial partners including the University of East Anglia (UEA), the John Innes Centre (JIC), the University of Bristol, the University of Bath, Unilever and Croda. Thanks to this, I have been very frequently in touch with research groups with expertise in different areas, taking small or large part in several collaborations.

Principal aim of the GelEnz network is to develop bio-based interpenetrating gels network from starch and cellulose feedstocks, otherwise destined to waste. In specific, Professor Steve Eichorn (University of Bristol and leader of the GelEnz platform), as well as Professor Janet L. Scott and Professor Karen Edler (University of Bath) have ongoing research interests in the development of novel cellulosic materials and composites combining processing and chemical modifications. Together, they joint expertise in material characterisation with a wide range of length scale – from bulk properties to supramolecular arrangement – using techniques such as rheology, Raman spectroscopy and scattering techniques (synchrotron x-ray and neutron diffraction).

In addition, the research group of Professor Robert A. Field (previously based at the John Innes Centre and recently moved the University of Manchester) has strong expertise on glycobiology and structural biology and is focused on the combination of chemical and enzymatic approaches for the tailored synthesis of oligo- and polysaccharides. In particular, they play a field leading role in the use of glycosyl phosphorylase (GPs) for the synthesis of novel oligosaccharides.

Under the framework of this wide-ranging research project, I have decided to include in this dissertation the three biggest and most relevant works from a biological and material development and characterisation prospective. Hence, we have addressed:

- 1) The structural elucidation of the binding of cognate and non-cognate substrates with the wild-type cellodextrin phosphorylase (CDP, EC 2.4.1.49:  $\beta$ -1,4-glucan linkage-dependent).
- 2) The detailed molecular level understanding and structural characterisation of novel cellodextrin derivatives combining solution and solid-state NMR spectroscopy and PXRD.
- 3) The validation and optimisation of a novel NMR-based protocol for the characterisation of the degree of solvent structuration for polysaccharides-based gels prepared in deuterated solvents.

To address the objectives of this thesis, we have taken advantage of the experience gained by the group in high-resolution Nuclear Magnetic Resonance (NMR) spectroscopy, computational methods such as protein-ligand docking and long-range characterisation of material based on PXRD. In particular, Saturation Transfer Difference NMR spectroscopy and transferred-NOE techniques have been employed for the binding studies and combined with protein-docking calculations, while  $^1\text{H}$ - $^{13}\text{C}$  CP/MAS NMR and PXRD were used to obtain material fingerprints.

## 1.2 Introduction

### 1.2.1 A multidisciplinary approach to science

During these 4 years, I had the terrific opportunity to be involved with the GelEnz project, a platform aiming to provide a multi-disciplinary approach in solving a global sustainability issue. The many projects coming and going have been for me an exceptional training in an extremely inspirational environment. All of this in the framework of helping to resolve the societal problem of food crops waste, that could be 'recycled' into sustainable alternative for ingredients currently used in manufacturing.

To observe and practice different approaches, as well as to learn and deepen concepts with a wide arm of reach, shaped my scientific mindset. All the different bits and pieces came together to shape not only this dissertation, but my future approach to science. Hence, if you wonder which was the recipe for my PhD – here it is: chemistry, biochemistry, glycobiology, structural biology, material science, physics, mathematics, geometry, a pinch of informatic, with a focus on circular economy and waste reduction and management.

Taking inspiration from Louise Pasteur, who once said: “In the fields of observation chance favours *only the prepared mind*”, I see no more prepared mind than the one able to observe things from different perspectives.

### 1.2.2 Chemistry of carbohydrates

Carbohydrates are ubiquitous in nature in different combinations, from single monosaccharides to extremely complex glycoconjugates. They are the most abundant class of bioorganic compounds in the biological world, and their complexity reflects the crucial functions they play *in vivo* (from structure and storage to specific signalling).

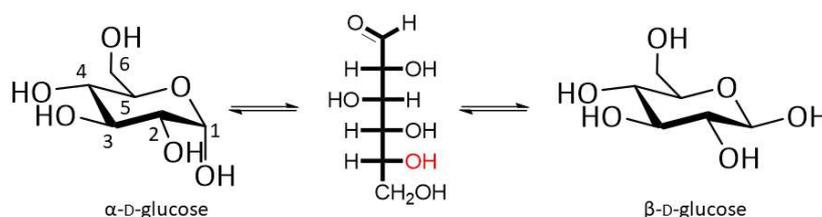
The name carbohydrate derives from “hydrates of carbon” due to the consistent 1:2:1 molar ratio between carbon, hydrogen and oxygen with empirical formula  $C_n(H_2O)_n$ . They are constituted by several hydroxyl groups and a carbonyl functional group such as an aldehyde (aldoses) or a ketone (ketoses). In addition, different numbers of sugar units can condensate to form mono-, di-, oligo- or polysaccharides.

If we take the glyceraldehyde molecule as a prototype of carbohydrate, we see that it has a stereogenic centre in C2, hence the molecule is chiral and presents two enantiomers (Figure 1.1). Based on the *Fischer projection*, the D-/L- nomenclature for the enantiomers is assigned by looking at the hydroxyl group at the stereogenic centre that is most remote of the carbonyl function: when it is at the right-hand side, the D- enantiomer is defined, while when on the left-hand side, we are in the presence of the L-enantiomer. The D-enantiomeric configuration is the most commonly found in nature. The D-/L- nomenclature does not give information on the absolute structure of the molecule. Alternatively, the R-/S- notation based on the Cahn-Ingold-Prelog (CIP) rules allows the exact assignment of each stereogenic centre. The introduction of a second or more stereogenic centres give rise to  $2^n$  ( $n$  = number of stereogenic centres) stereoisomers (two pairs of enantiomers), defined as diastereomers. Diastereomers that exhibit the opposite configuration at only one of the tetrahedral stereogenic centres are called *epimers* (i.e. D-glucose and D-mannose). Importantly, diastereomers show different physical parameters, while enantiomers exhibit the same ones (besides enantioselective interactions with other chiral molecules or polarized light).



Anomers are diastereomers and therefore have different physical properties (i.e. solubility, melting point and specific rotation).

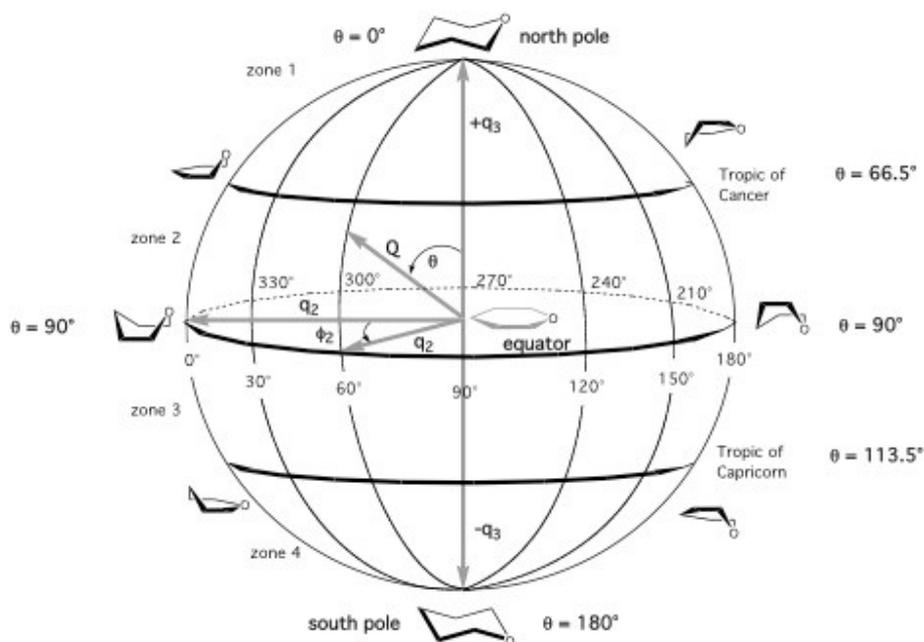
In aqueous solution, only 0.002% of D-glucose exists in the open-chain form, and  $\beta$ -D-glucose is about twice the concentration (62%) than  $\alpha$ -D-glucose (38%). The equilibrium between the two configuration is reached via the *mutorotation* phenomenon (Figure 1.2).



**Figure 1.2:** Schematic representation of the interchange between the  $\alpha$ - and  $\beta$ -configuration in D-glucose in solution achieved via the mutarotation phenomenon.

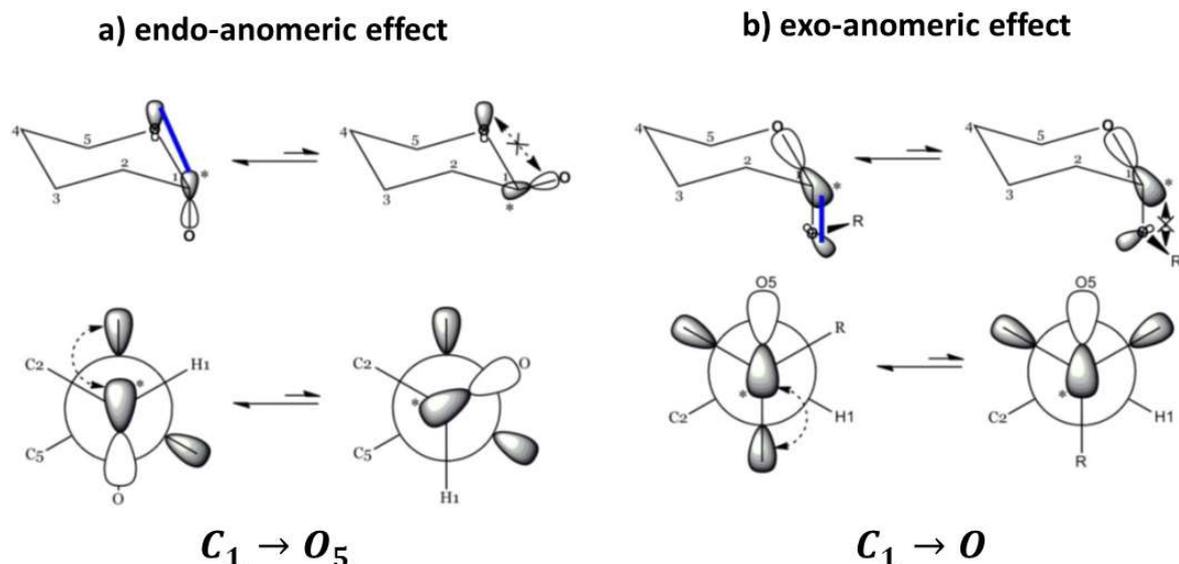
To understand why the  $\beta$ -configuration is more stable than the  $\alpha$ , there are some additional factors that we need to take into consideration. First, as for cyclohexane, D-glucopyranose can adopt several conformations (i.e., chair, half-chair, boat, twisted boat, envelope): conformational isomers that can interchange without breaking any chemical bonds (Figure 1.3). The chair conformation is normally the most stable, where the ring substituents can adopt an axial or equatorial position (parallel or perpendicular to a vertical axis drawn through the ring, respectively). The equatorial substituents are usually slightly above or below a virtual horizontal plane. The angle between axial and equatorial substituents is almost a tetrahedral angle ( $109.5^\circ$ ). Substituents in equatorial position present lower steric hindrance; consequently, the equatorial position of the bulky  $-\text{CH}_2\text{OH}$  group and the hydroxyl group in C1 in glucopyranose are favoured.

The glucopyranose chair conformation exists in two forms ( ${}^4\text{C}_1$  and  ${}^1\text{C}_4$ , where C = chair) with an energy difference of  $42 \text{ kJ/mol}^2$  and that may interchange with each other, with consequent rearrangement of the bonds from equatorial to axial (Figure 1.3). Taking the ring-oxygen and C2, C3 and C5 carbons as reference plane, the conformation  ${}^4\text{C}_1$  is defined with C1 and C4 are below and above this plane, respectively; otherwise it is labelled  ${}^1\text{C}_4$ .



**Figure 1.3:** Diagram of the pseudorotational itinerary of the pyranose ring according to Jeffrey and Yates<sup>3</sup>, based on Cremer-Pople ring puckering coordinates  $\theta$  and  $\Phi$ .<sup>4</sup> The polar  ${}^4C_1$  ( $\theta = 0^\circ$ ) and  ${}^1C_4$  ( $\theta = 180^\circ$ ) chairs, together with the 12 equatorial puckers ( $\theta = 90^\circ$ ) are shown, while the envelope (E) and half-chair (H) conformations are not shown.

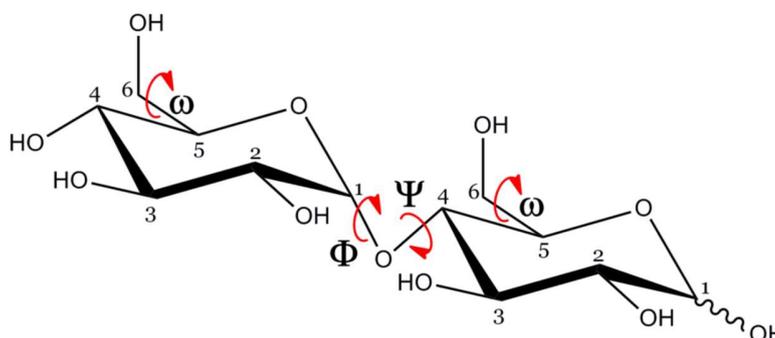
In addition, the dipole moments of the  $C(5) \rightarrow O_{ring}$  and the  $C(1)_{anomer} \rightarrow OH$  are nearly parallel and point into the same direction (Figure 1.4, a). This determines a hyperconjugation between the lone pair electrons in the non-bonding orbital ( $n$ ) of the ring oxygen and the antibonding  $\sigma^*$  orbital of the  $C(1) - O_1$  bond ( $n \rightarrow \sigma^*$ ). This so-called *endo-anomeric effect* which favours the axial configuration in C1 position and therefore the  $\alpha$ -anomer. This is compensated by the hyperconjugation between the lone pair electrons of the non-bonding orbital ( $np$ ) of the glycosidic oxygen ( $O_g$ ) and the antibonding  $\sigma^*$  orbital of the  $O_5 - C_1$  bond (Figure 1.4, b), the so-called *exo-anomeric effect*. This effect favours the  $\beta$ -configuration at the anomeric carbon.



**Figure 1.4:** Schematic representation of the stereoelectronic genesis of a) the endo- and b) the exo-anomeric effect.

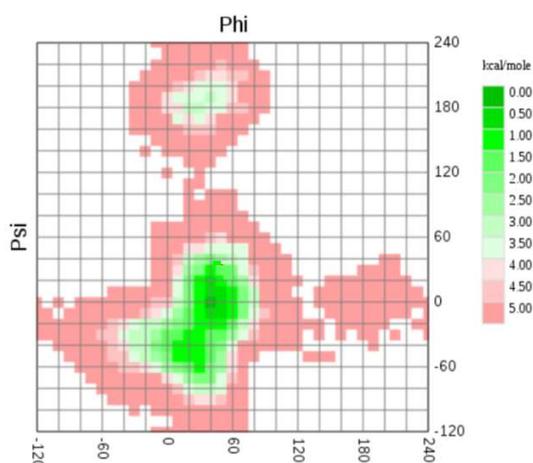
Disaccharides, oligosaccharides and polysaccharides are two or more monosaccharides linked together through glycosidic linkages. The structure of the polysaccharides depends on 1) the monosaccharide sequence, 2) the site and stereochemistry ( $\alpha$  or  $\beta$ ) of the glycosidic linkage and 3) the degree and type of substitution of hydroxyl groups (such as O-methylation or O-sulphation).

Another essential feature of carbohydrates is their high conformational flexibility. The glycosidic linkage torsional angles  $\Phi$  and  $\Psi$  define the relative orientations of two consecutive monosaccharide units in a disaccharide moiety around the glycosidic, while the  $\omega$  torsional angle is associated with rotation at the hydroxyl group (rotation around  $C_5 - C_6$  bonds, Figure 1.5).  $\Phi$  and  $\Psi$  are defined differently in crystallography and NMR spectroscopy; in the first,  $\Phi \equiv O_5 - C_1 - O_g - C_x$  and  $\Psi \equiv C_1 - O_g - C_x - C_{(x+1)}$  (where,  $O_5$  is the oxygen ring,  $C_1$  is the anomeric carbon,  $O_g$  is the glycosidic oxygen and  $C_x$  is the carbon belonging to the following residue, linked to the glycosidic oxygen), while in the NMR definition  $\Phi \equiv H_1 - C_1 - O_g - C_x$  and  $\Psi \equiv C_1 - O_g - C_x - H_x$  (where  $H_1$  is the hydrogen attached to the anomeric carbon).



**Figure 1.5:**  $\Phi$ ,  $\Psi$  and  $\omega$  dihedral angles that define the overall conformation of oligosaccharides. In the example: maltose disaccharide.

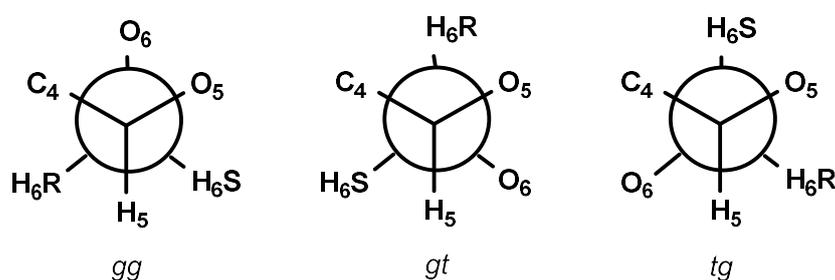
The overall conformation of the oligosaccharide chain is therefore determined by the set of  $\Phi$ ,  $\Psi$  and  $\omega$  angles. The energetically favourable conformations of a carbohydrate may be identified on Carbohydrate Ramachandran plots (CARP, Figure 1.6), which evaluates the preferred  $\Phi$ ,  $\Psi$  values of the glycosidic bond torsion angles. The energy of the linkage, the energetically favourable/unfavourable and the allowed or forbidden regions, depend on the types of monosaccharides involved, on the kind of linkage and on the degree of branching of the oligosaccharide.



**Figure 1.6:** Example of ( $\Phi$ ,  $\Psi$ ) maps for D-cellobiose with the glycosidic linkage energetic minima represented in green. Source: Lutteke *et al.*, 2004 (<http://www.glycosciences.de/>)<sup>5</sup>

As described above for the anomeric position, also the exocyclic hydroxymethyl group in hexopyranoses presents conformational preferences. Although the  $\omega$ -angle can adopt three stable

staggered rotamers, the gauche-gauche (gg), gauche-trans (gt) and trans-gauche (tg) conformers characterised to the  $O_5 - C_5 - C_6 - O_6$  and  $C_4 - C_5 - C_6 - O_6$  dihedral angles (Figure 1.7), it shows propensity to adopt a *gauche* conformations. For gluco- and mannopyranosides, almost equal populations of gt and gg rotamers with nearly complete absence of the tg rotamer (both in solid and solution phases) was showed. In contrast to the anomeric effect, the gauche effect is principally caused by solvation and electrostatic interactions<sup>6</sup>, rather than steric or stereoelectronic effects.



**Figure 1.7:** Newman projections of the gauche-gauche (gg), gauche-trans (gt) and trans-gauche (tg) rotameric conformers of the  $\omega$  dihedral angle.

Polysaccharides can form a myriad of complex structure due to the existence of two  $\alpha$  and  $\beta$  anomeric configurations, the possibility to form glycosidic-linkage at different positions (1 $\rightarrow$ 1, 2, 3, 4, 6 for hexopyranose), the changes in ring size (pyranose/furanose), and the introduction of branching and additional site-specific substitutions such as acetylation, phosphorylation or sulfation. Importantly, even a subtle change in the structure of the monosaccharide unit or the type of glycosidic linkage has a profound effect on the properties and functions of the polysaccharides. Oligosaccharides or glycans can be obtained by isolation from natural sources or prepared enzymatically and/or chemically, and the precise synthesis of a non-natural polysaccharides has potential application in the field of medicine, pharmaceuticals, cosmetics and food industries. Nonetheless, the chemical synthesis of oligosaccharides remains challenging.

### 1.3 Challenges in the synthesis of carbohydrates

Since 1941, when it was first attempted to synthesise cellulose,<sup>7</sup> many efforts have been devoted to the synthesis of polysaccharides with a well-defined structure. The great challenges in carbohydrate chemical synthesis are to achieve stereoselectivity (formation of the glycosidic

linkage with either the  $\alpha$  - and  $\beta$ - configuration) and regioselectivity (introduction of the glycosidic linkage only in one of the several available hydroxyl groups), which *in turn* depends on the choice of substrates (with appropriate leaving and protecting groups), catalysts and experimental conditions such as solvent, temperature and atmosphere.<sup>8, 9</sup> In general, the total synthesis of glycans sees the formation of a glycosidic bond between a glycosyl donor and an acceptor molecule. Important steps are: 1) the activation of a glycosyl donor by introducing a leaving group (X) at the anomeric carbon level; 2) the protection of the hydroxyl groups in both the donor and the acceptor molecules, leaving free only the one which should participate to the reaction. This second step determined the regioselectivity in the glycosidic bond formation.<sup>10</sup> One approach is the so-called cationic ring-opening polymerisation of anhydrosugar monomers.<sup>11</sup> Based on this approach, cellulose with degree of polymerisation (DP) equal to 19.3 was synthesised by polymerisation of 3,6-di-*O*-benzyl- $\alpha$ -D-glucose 1,2,4-orthopivalate and the subsequent removal of the protective group.<sup>12</sup>

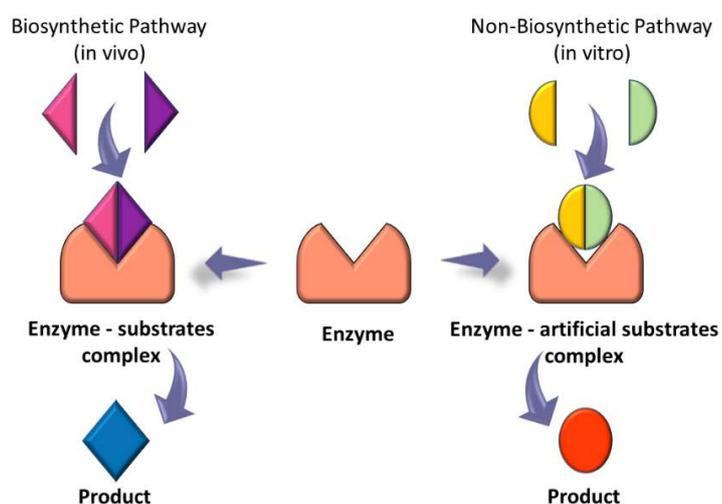
However, manual synthesis of oligosaccharides requires multistep protection-deprotection reactions, leading to undesired side-products. Furthermore, perfect stereocontrol of glycosidic linkages has not often been achieved.<sup>13</sup> To overcome the manual synthesis limits, automated approaches for oligosaccharide synthesis have been developed. Early attempts to advance in the automated chemical syntheses in solution were performed by Takahashi and Wong, and, in 2012, Seeberger *et al.*, reported the first fully automated solid-phase oligosaccharide synthesizer (Glyconeer 2.1).<sup>14</sup>

### 1.3.1 Enzymatic approaches for carbohydrates synthesis

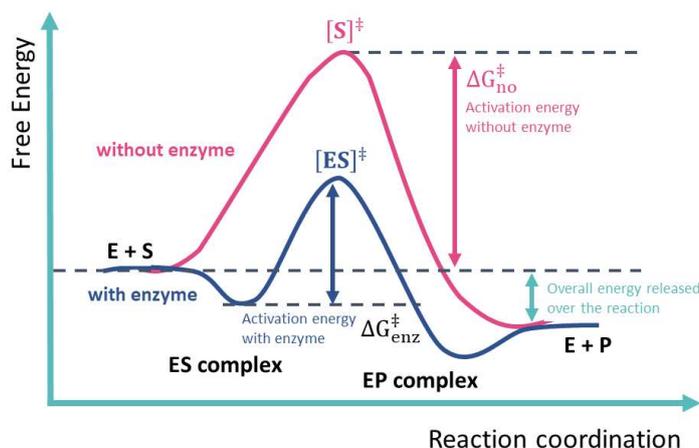
Enzymatic catalysis (*in vivo* and *in vitro*) allow the synthesis of polysaccharides with highly stereo-, regio-, and enantio-selectivity (hence, formation of structurally controlled products) in mild conditions, such as buffered water, neutral pH and atmospheric pressure. In addition, from an environmental perspective, enzymes are defined as green catalyst being natural and renewable.<sup>15</sup>

Importantly, a “key and lock” relationship (proposed by Fischer, 1894)<sup>16</sup> exists between an enzyme and its substrates, but this relationship is not absolutely strict for *in vitro* enzymatic reactions. Thus, some enzymes present loose specificity in the catalytic cleft and can recognise non-natural

substrates, form the enzyme-substrate (ES) complex and catalyse the formation of a product (Figure 1.8). This feature allowed the introduction of structural modifications in complex natural oligosaccharides in a controlled manner. In addition, during the occurrence of the catalytic reaction, the formation of the ES complex stabilises the substrate transition state and lowers its activation energy ( $\Delta G_{enz}^{\ddagger} < \Delta G_{no}^{\ddagger}$ ) (Figure 1.9), explaining why enzymes are able to perform catalysis under mild conditions and accelerate  $10^6$ – $10^{12}$  fold the reaction rate.<sup>17</sup>



**Figure 1.8:** Schematic representation of the enzyme-substrate key and lock relationship (left) and acceptability of unnatural substrates (right). In the case of in vitro enzymatic synthesis of polysaccharides, unnatural substrates can be recognised and employed for the enzymatic catalysis.



**Figure 1.9:** Energy of activation diagram (Pauling, 1946) for a chemical reaction. The comparison between an enzyme-catalyzed reaction and a reaction without enzyme. (E= Enzyme, S= Substrate, P= Product) shows that the activation energy ( $\Delta G_{enz}^{\ddagger}$ ) of the transition state  $[ES]^{\ddagger}$  is greatly lowered by the stabilizing action by the enzyme, in comparison with that ( $\Delta G_{no}^{\ddagger}$ ) of a reaction without enzyme via a transition state  $[S]^{\ddagger}$ .

Based on this background, the employment of enzymes for the precise synthesis of oligosaccharides is expected to perform: (1) perfect control of stereo- and regioselectivities in glycosidic linkages; (2) synthesis of novel; nonnatural polysaccharides without formation of side products; and (3) great turn over without the use of harmful catalysts as strong acids and bases or heavy metals.

Enzymes are categorised in six main classes: oxidoreductases (EC 1), transferases (EC 2), hydrolases (EC 3), lyases (EC 4), isomerases (EC 5) and ligases (EC 6).<sup>18</sup> Among those, carbohydrate active enzymes (CAZy, <http://www.cazy.org>) – hence involved in the cleavage, formation and rearrangement of glycosidic linkages - are categorized into two main classes: hydrolytic enzymes (hydrolases) and glycosyltransferases. The latter is further subclassified into synthetic enzymes (Leloir glycosyltransferases), sucrose-type enzymes and phosphorylases (phosphorylases).

Like in the general glycosylation, enzymes catalyse the reaction between an activated glycosyl donor and a glycosyl acceptor. The reaction proceeds as: 1) recognition of the glycosyl donor by the enzyme and formation of the glycosyl-enzyme complex (transition state); 2) attack of the glycosyl donor by the hydroxyl group of the glycosyl acceptor and formation of a glycoside;

3) liberation of the elongated glycoside and the free leaving group. Importantly, in performing enzymatic reactions *in vitro*, both the donor and acceptor molecules can be used in their unprotected forms.

Glycoside hydrolases are further divided into endo- and exo-types. Endoglycosidases catalyse natural polysaccharides hydrolysis *in vivo* but are able to perform glycosylation *in vitro* to produce the saccharide chains, probably due to a similar transition state structure. Hence, to proceed towards the synthetic direction, the enzyme substrates are designed as transition state analogues to efficiently form the enzyme-substrate complex. In specific, two different kind of monomers have been designed (i.e. glycosyl fluorides and sugar oxazolines).

Leloir glycosyltransferases (generally transmembrane-type proteins) catalyse irreversible glycosylation *in vivo*, hence have a biological relevance. The reaction transfers the glycosyl residue from a sugar donor containing a nucleoside phosphate or a lipid phosphate leaving group to an acceptor molecule such as carbohydrates, lipids, and peptides. Hence, the reaction irreversibility is dictated by the cleavage of the high-energy linkage of nucleotide substrate. Nonetheless, the utilisation of glycosyltransferases is currently limited because their instability upon isolation and purification, their low amount in nature and the high cost of both the enzyme and the activated donor substrates.

Sucrase-type enzymes are highly specialised in the recognition of sucrose as substrate and are able to transfer either the glucose or fructose moiety onto polysaccharides or appropriate acceptors, hence they are called glucosyltransferases (or glucansucrases) and fructosyltransferases (or fructansucrases), respectively.

Phosphorylase can perform the catalytic reaction towards both the phosphorolysis and the reverse phosphorolysis direction, promoting the cleavage and the synthesis of the glycosidic linkage, respectively.  $\alpha$ -Glucan phosphorylase, sucrose phosphorylase, cellobiose phosphorylase and cellodextrin phosphorylase are the most studied phosphorylases, with  $\alpha$ -Glucan phosphorylase and cellodextrin phosphorylase extensively used for the practical synthesis of poly- or oligosaccharides and related poly- and oligosaccharide-based materials.  $\alpha$ -Glucan phosphorylase is found in animals, plants, and microorganisms, where acts in the glycolytic pathway or, in other words, in the utilisation of storage polysaccharides. This enzyme catalyses the reversible

phosphorolysis of  $\alpha$ -(1 $\rightarrow$ 4)-glucans such as glycogen and starch<sup>19, 20</sup> and it has been used to synthesise unnatural saccharides<sup>21</sup> and modified nanomaterials.<sup>22</sup> Sucrose phosphorylase is found in bacterial cells and it is involved in the metabolism of extracellular sucrose. The enzyme catalyses the reversible phosphorolysis of sucrose into Glc-1-P and fructose<sup>19, 20</sup> and it has been used industrially for the kilogram-scale production of the cosmetic humectant  $\alpha$ -glucosyl glycerol.<sup>21</sup> Cellobiose phosphorylase (CBP) was proved to act only in the cellobiose moiety and do not recognise celooligosaccharides larger than cellobiose. Pfeifer & Langen GmbH & Co. KG implemented a large-scale production of D-cellobiose by CBP with concomitant formation of  $\alpha$ -Glc1-P from sucrose using sucrose phosphorylase.<sup>23</sup> More details in the reaction catalysed by phosphorylases are presented in Chapter 3, where we will pay special attention to cellodextrin phosphorylase (CDP) enzyme.

## 1.4 Carbohydrates-protein interactions – the molecular recognition event

### 1.4.1 The enthalpy-entropy compensation effect

The enzyme-substrate system (or even more generally, protein-ligand system) is a thermodynamics system with the solvent (i.e. liquid water and buffer ions) as third component. Various interactions and energy exchange processes take part in association between the system components and the spontaneity of the processes is measured by changes in the Gibbs free energy ( $\Delta G$ ). Importantly,  $\Delta G$  can be expressed based on its thermodynamics components by the equation

$$\Delta G = \Delta H - T\Delta S \quad \text{Equation 1.1}$$

where  $\Delta H$  and  $\Delta S$  correspond to changes in enthalpy and entropy, respectively, and T is the temperature in K. Negative  $\Delta G$  describes spontaneous process, while at  $\Delta G = 0$  the association reaction is at equilibrium.

The enthalpy corresponds to the total energy of the thermodynamics system, and in the binding process negative ( $\Delta H < 0$ ) or positive ( $\Delta H > 0$ ) changes of the binding enthalpy represent

formation and disruption of energetically favourable noncovalent interactions (van der Waals contacts, H-bonds, ion pairs, and any other polar and apolar interactions), respectively. The combination of (i) the disruption of noncovalent interactions between the protein and the solvent and the ligand and the solvent, (ii) the formation of noncovalent interactions between the protein and the ligand, and (iii) the reorganisation of the solvent in proximity of the complex surfaces, contribute to the net enthalpy change in the binding event.

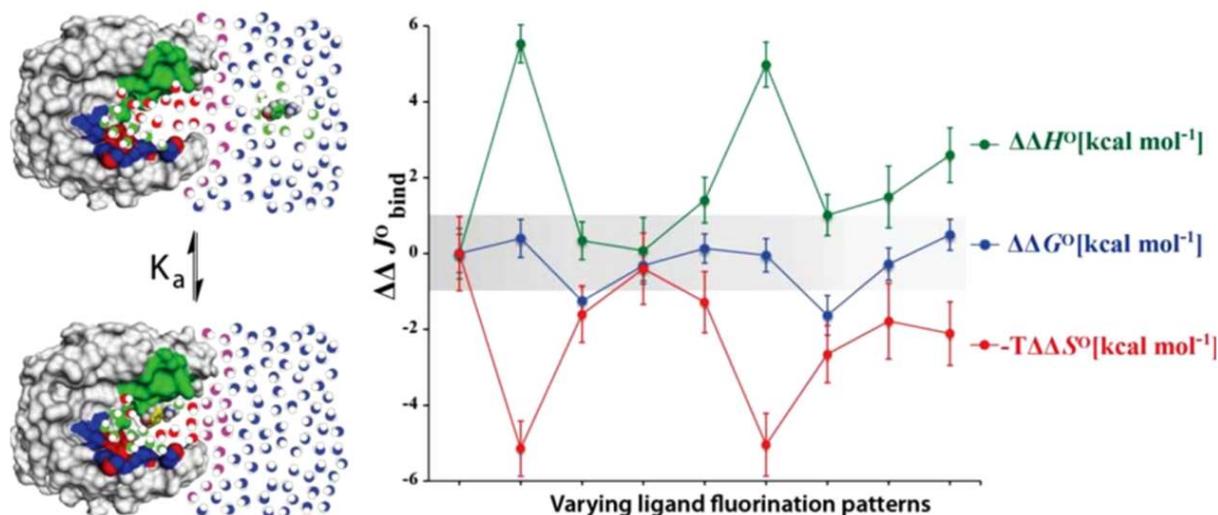
The changes in the entropic factor, instead, describe changes in the degree of freedom of the system, with the following three entropic terms contributing to the net binding entropy change  $\Delta S$

$$\Delta S = \Delta S_{solv} + \Delta S_{conf} + \Delta S_{r/t} \quad \text{Equation 1.2}$$

where  $\Delta S_{solv}$  is the solvent entropy change arising from solvent release upon binding,  $\Delta S_{conf}$  is the conformational entropy change, which represent changes of the protein and ligand conformational freedom upon binding, and  $\Delta S_{r/t}$  represents the loss of translational and rotational degrees of freedom of the protein and ligand upon complex formation. Positive and negative net entropy change contributes favourably and unfavourably to the binding free energy;  $\Delta S_{solv}$  usually makes a favourable contribution,  $\Delta S_{conf}$  can contribute either favourably or unfavourably, and  $\Delta S_{r/t}$  effects a negative contribution by selecting one single conformation, among many conformations the free ligand can assume when it is free.<sup>24</sup> Generally, for the binding event to occur, inescapable entropic penalties (i.e. the negative  $\Delta S_{r/t}$  upon binding) are overcome by either large solvent entropy gain (positive  $\Delta S_{solv}$ ) or favourable protein-ligand interactions (which lead to negative binding  $\Delta H$ ).

Entropy and enthalpy changes determine the sign and magnitude of the binding free energy ( $\Delta G$ ) and the magnitude of the negative (spontaneous) energy change determines the stability of the complex. Thus, we consider  $\Delta H$  and  $\Delta S$  as the driving factors for protein ligand binding. Generally, a complementary change between enthalpy and entropy occurs – the large negative enthalpy change resulting from the establishment of multiple favourable noncovalent interactions is usually accompanied by restriction in the mobility of the interacting patterns (conformational restrictions of the ligand and the small conformational changes throughout the protein upon binding)<sup>25, 26</sup> and therefore a negative entropy change. Similarly, the de-solvation of both the free ligand and of the

apo binding pockets is entropically favourable, but enthalpically unfavourable as disruption of noncovalent interactions is required.<sup>27</sup> This is called the enthalpy/-entropy compensation effect (Figure 1.10).



**Figure 1.10:** Differences in the enthalpy and entropy contribution measured by ITC 298.15 K for 8 different benzothiazole sulfonamide ligands with different patterns of fluorination, substrates of the human carbonic anhydrase (HCA).  $\Delta\Delta J^{\circ}$  indicates the difference in binding energy for this series of homologous compounds. Despite the differences in fluorination pattern, the ligands show similar binding affinities (similar Gibbs energies,  $\Delta G$  – in blue). Nonetheless, they also show significant and compensating changes in enthalpy ( $\Delta H$ , in green) and entropy ( $\Delta S$ , in red) of binding. The grey region demarcates the 95% confidence interval (i.e., two standard deviations) of  $\Delta\Delta G^{\circ}_{bind}$  for the reference compound. On the left panel the sketch of the association process depicts the differences in the structure (and consequently thermodynamics) properties of the water molecules surrounding the bound ligand (bottom) and filling the active site of a protein (protein). *Source:* Breiten *et al.*, 2013 (graphic abstract).<sup>28</sup>

The enthalpy-entropy compensation phenomenon has been for long a subject of debate which attributed it to experimental errors and limitations. Nonetheless, nowadays there is a much clearer evidence of this phenomenon for both dispersive and electrostatic systems. For example, Ryde (2014)<sup>30</sup> studied the enthalpy-entropy compensation effect during the association of two molecules by means of molecular mechanics (MM) and quantum mechanics (QC) calculations in

gas phase. Both the investigation of the association of a two-atom model (in which the dispersion, electrostatic and covalent interactions could be easily distinguished) and of homologous series of complexes dominated either by dispersion or hydrogen bonds resulted in the observation of the enthalpy-entropy compensation effect. Nonetheless, he pointed out that when water molecules were added to the calculations, both enthalpy-entropy compensation and anti-compensation phenomena were observed, hence demonstrating that the phenomenon is not strict in real binding studies.

### 1.4.2 Specific binding patterns, role of water and hydrophobic effect for carbohydrate recognition in the enzyme catalytic cleft

From the above discussion it is clear that the process of carbohydrates molecular recognition in the catalytic cleft is driven by noncovalent interactions, and specific elements have been identified for in protein-carbohydrate interactions. In specific, carbohydrate binding motif (CBMs, non-catalytic domains linked to the catalytic part of glycoside hydrolases)<sup>31</sup> and lectins (carbohydrate-binding proteins which lack enzymatic activity on their ligand and are distinct from antibodies and free mono- and oligosaccharide sensor/ transport proteins)<sup>32</sup> have been used as model systems for studying carbohydrates recognition mechanism.

Carbohydrates present an amphiphilic nature: they are 1) hydrophilic due to the high density of hydroxyl groups, which allows them to act both as directional H-bond donors and acceptors and to coordinate  $\text{Ca}^{2+}$  ions<sup>32</sup>, and 2) hydrophobic due to the slight electron-depletion on the CH bond (of carbons carrying hydroxyl or carboxylic groups), which enables the establishment of van der Waals, stacking, and CH/ $\pi$ -interactions, as well as changes of the solvent structure in their vicinity with enthalpic/entropic consequences.<sup>32, 33</sup>

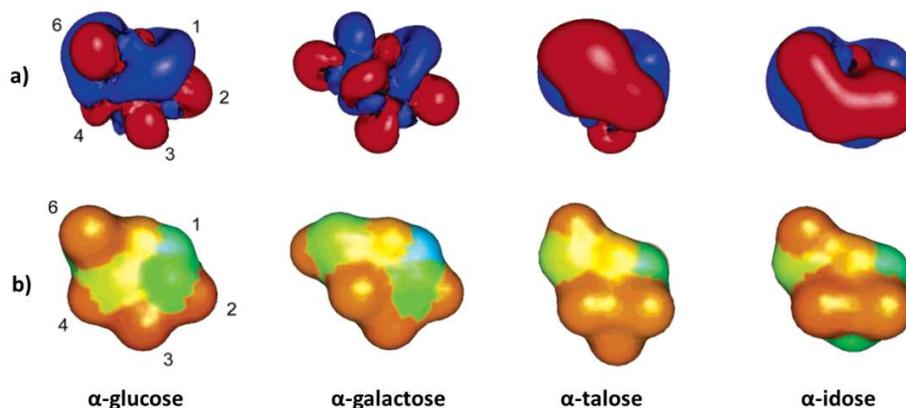
Common residues side chains in the binding pocket involved in the establishment of cooperative H-bonds are arginine, aspartate and asparagine, while residues involved in hydrophobic interactions are tryptophan, tyrosine and phenylalanine. In addition, CH/ $\pi$ - stacking is favoured for  $\beta$ -sugars ( $\beta$ -glucose, -galactose, -mannose, -glucuronic acid) due to the presence of 3 axial CH at position C1, C3 and C5.<sup>32</sup> Furthermore, epimerisation has usually a substantial consequence in the process of biorecognition, as epimers present different directionality of H-bond due to their

different electron density profile. Hence, each sugar is recognised in the binding pocket with high levels of specificity.<sup>32</sup>

In the molecular recognition process, it is important to take into consideration the role played by water, being the medium in which the interactions take place and often intrinsically present in amphiphilic binding sites. In addition, free ligands in solutions are surrounded by several layers of water molecules.

The amphiphilic nature of carbohydrates leads to anisotropic solvent densities around the molecule, with consequent formation of hydrophobic and hydrophilic patches in aqueous environments. Favourable interactions of water molecules with the hydrophilic patches result from electrostatic interactions and hydrogen bonding, whereas the interaction of water with hydrophobic surface patches is unfavourable.

Computational studies performed by the group of Vanderkooi showed the influence of hydroxyl orientation on solute hydration and surrounding water structure in aldohexopyranose stereoisomers.<sup>34</sup> The work highlighted how for a progressive conversion of hydroxyls 4, 2, and 3 from equatorial to axial configuration ( $\alpha$ -glucose,  $\alpha$ -galactose,  $\alpha$ -talose, and  $\alpha$ -idose), the establishment of intramolecular H-bond networks (OH4  $\rightarrow$  OH6  $\rightarrow$  O5 and OH6  $\rightarrow$  OH4  $\rightarrow$  OH3) and syndiaxial OH-1/3 interactions determines a more hydrophobic surface of the sugar molecule, decreasing the ability to form solute-solvent H-bonds (Figure 1.11). In addition, the syndiaxial OH-2/4 intramolecular hydrogen bonding was reported to stabilise and induce directionality in the OH4 H-bond network (as in the OH2  $\rightarrow$  OH4  $\rightarrow$  OH6  $\rightarrow$  O5 H-bond terminating network), while the OH-1 and OH-3 groups have only an indirect effect on water structure; when these groups are equatorial, a hydrophobic patch is formed by C-H groups on one face of the sugar. However, the water structure surrounding this patch is actually relatively more polar-like than that observed when the OH-1 and OH-3 are in configurations that do not support patch formation.<sup>34</sup>



**Figure 1.11:** a) Positive (blue) and negative (red) electrostatic field iso-surface contours used to illustrate how partial charge impacts the surrounding environment. Sugars with both OH-2 and OH-4 in axial configuration ( $\alpha$ -talose and  $\alpha$ -idose) are able to form a syndiaxial intramolecular H-bond. As a result, the sugar faces are dominated by one potential surface rather than a mixture of positive and negative isopotentials, which is instead the case of  $\alpha$ -glucose (OH-2 and OH-4 equatorial) and  $\alpha$ -galactose (OH-2 equatorial, OH-4 axial); b) Average water-water H-bond angles with the scale colour ranging from more linear H-bonds (27°, light blue) to more bent H-bonds (43°, orange). More distorted, less linear H-bonds are associated to a lower interaction water with the solute surface. *Source:* Adapted from Vanderkooi *et al.*, 2005.<sup>34</sup>

Finally, the so-called hydrophobic effect (i.e. the tendency of nonpolar molecules - or part of them - to aggregate in aqueous media), plays a central role in biomolecular recognition.<sup>28</sup> Even though the already discussed CH/ $\pi$ - stacking interaction is a remarkable example of hydrophobic effect,<sup>33</sup> the low occurrence of aliphatic residues in carbohydrate-recognising binding pockets has been attributed to a lower contribution of this phenomenon to protein-carbohydrate recognition.<sup>35</sup>

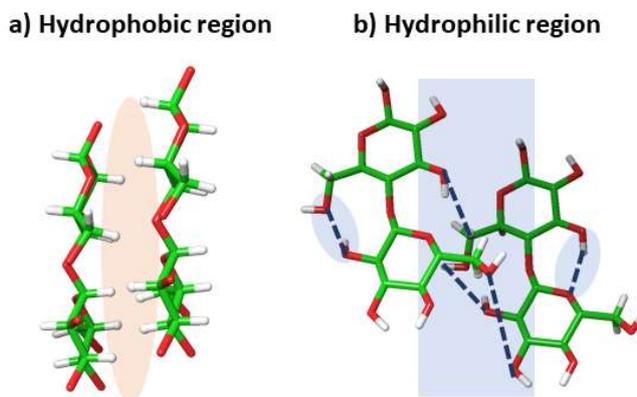
## 1.5 Cellulose

Cellulose is the most abundant biopolymer on earth, and it was firstly isolated by the French chemist Anselme Payen in 1838 by extraction from green plants.<sup>36</sup> Cellulose is a linear homopolymer composed by anhydroglucopyranose units (AGU) linked together by  $\beta$ -(1 $\rightarrow$ 4)-glycosidic linkages. The size of cellulose molecules is defined by the average degree of

polymerisation (DP), while the product of the average DP and the molecular mass of a single AGU defines cellulose molecular weight. Due to its abundance, good mechanical properties, high aspect ratio, low density, thermal stability and renewable resource origin, cellulose has become a symbolic molecule in the field of polymers and macromolecules, and many fundamental and practical studies on cellulose structure, chemical and physical properties, biosynthesis and morphology have been carried out.<sup>19,20</sup>

Cellulose is the major component of plant cell walls, but it is also produced by bacteria, algae and marine organisms such as tunicates. In the plant cell wall it is usually combined with lignin, hemicelluloses (arabinoxylan, xyloglucan, high-molecular weight mixed-linkage glucan)<sup>37</sup> and water. Native cellulose is organised in highly hierarchical fashion (from a single AGU to macro fibrils), where cellulose molecules are held together by an extensive network of intra- and inter-chains hydrogen bonds.<sup>36</sup>

The presence of three equatorial hydroxyl groups in each AGU (in position C2 and C3 there are two secondary alcohols, while in position C5 there is a primary alcohol, the hydroxymethyl group) donate hydrophilic properties to cellulose, while the numerous axial C-H present in the sugar rings give a hydrophobic character to the polymer. Thus, cellulose shows amphiphilic nature and intrinsic structural anisotropy (Figure 1.12). In addition, native cellulose is a semi-crystalline polymer and has amorphous regions (low ordered) coexisting with crystalline domains (highly ordered), the degree of crystallinity (40–60%) depending on the extraction process of the sample and on the cellulose source.



**Figure 1.12:** a) lateral and b) top view of the cellulose molecule plane. The first shows the hydrophobic region characterised by the hydrogen atoms of C–H bonds on the axial positions of the ring, the second highlights the hypothetical hydrogen bonds between the hydroxyl groups located on the equatorial positions of the ring. The hydrogen-bonds are represented as blue dashed lines.

### 1.5.1 Cellulose polymorphs

The crystal structure of cellulose has been investigated for more than a century, and Nishikawa and Ono were the first to obtain X-ray diffraction traces from wood, hemp and bamboo.<sup>38</sup> Cellulose can exist in 6 different allomorphic forms ( $I\alpha$ ,  $I\beta$ , II, IIIi, IIIii, IVi and IVii), even though cellulose IVi has been recently re-classified as  $I\beta$ .<sup>39</sup> Cellulose  $I\alpha$  and  $I\beta$  are the native forms of cellulose found in primitive microorganisms and plants,<sup>38, 40</sup> while the other allomorphs can be derived through thermochemical treatments from cellulose  $I\alpha$  and  $I\beta$ . Both native cellulose allomorphs can be converted irreversibly into cellulose II through mercerization or regeneration<sup>41</sup> and into cellulose IIIi by treatment with liquid ammonia and other amines.<sup>42, 43</sup> Cellulose IV, also called “high-temperature cellulose”, is prepared from cellulose II or III by treatment in glycerol at 260 °C, while cellulose I cannot be transformed directly into cellulose IV.<sup>39</sup> The allomorph cellulose IVi is exclusively formed from cellulose IIIi treatment, while the allomorph cellulose IVii is obtained from both cellulose II and IIIi as source material. Recently, Nishiyama *et al.* revised the crystallographic structures of cellulose and reported the H-bond networks by applying synchrotron X-ray and neutron diffraction studies.<sup>38-41, 43, 44</sup>

Cellulose can be modelled by a triclinic (cellulose I $\alpha$  and IVii) or monoclinic (cellulose I $\beta$ , II, IIIi and IIIii) unit cell. The unit cell can contain either one (cellulose I $\alpha$  and IIIi) or two (cellulose I $\beta$ , II, IIIii and IVii) chains, *in turn* labelled as ‘origin’ and ‘center’ chains. The origin chain is positioned at the corner of the unit cell parallel to the c axis direction, while the center chain passes through the center of the a/b plane. The unit cell of cellulose I $\beta$  presents two parallel and conformationally equivalent chains, while cellulose II and IIIii presents two antiparallel and conformationally inequivalent chains. In the case of cellulose IVii, two antiparallel chains were reported.<sup>45</sup> The unit cell parameters for each cellulose allomorph are reported in Table 1.1.

Cellulose allomorphs mainly differ in their conformations of the hydroxymethyl group (O5-C5-C6-O6 and C4-C5-C6-O6), with cellulose I $\alpha$  and cellulose I $\beta$  adopting a tg conformation, while cellulose II, IIIi and IIIii adopt a gt conformation (refer to Figure 1.7).

**Table 1.1:** Unit cell parameters for reported cellulose allomorphs. The values are derived from synchrotron X-ray diffraction experiments.

Cellulos	I $\alpha$ <sup>40</sup>	I $\beta$ <sup>38</sup>	II <sup>41</sup>	IIIi <sup>42</sup>	IIIii <sup>43</sup>	IVii <sup>45</sup>
<i>e</i>						
Space group	P1	P2 <sub>1</sub>	P2 <sub>1</sub>	P2 <sub>1</sub>	P2 <sub>1</sub>	P1
a(Å)	6.717(6)	7.784(8)	8.10(1)	4.450(4)	4.450(4)	7.99
b(Å)	5.962(6)	8.201(8)	9.03(1)	7.850(8)	7.640(8)	8.10
c(Å)	10.400(6)	10.380(10)	10.31(1)	10.310(10)	10.360(10)	10.34
$\alpha$ (°)	118.08(5)	90	90	90	90	90
$\beta$ (°)	114.80(5)	90	90	90	90	90
$\gamma$ (°)	333.3(6)	96.5	117.10(5)	105.10(5)	106.95(5)	90

## 1.6 Hydrogels

### 1.6.1 Definition of gel

Definitions of the term “gel” have been attempted since 1861, when Thomas Graham firstly distinguished between the rigidity of crystalline structure and the softness and fluidity of colloids.<sup>46</sup> Nevertheless, a clear definition struggled to arrive, to the extent that in 1926 Jordon Lloyd described the gel’s colloidal condition as easier to recognise than to define.<sup>47</sup> The difficulty on obtaining a clear description of gel systems by that time can be attributed to both a lack of definition of physicochemical principles and the high versatility of gel systems. Indeed, when talking about gels, we refer to semi-solid material with both solid and solution phase features. In 1974 Flory wrote the universal characteristic of gels for the Faraday Discussion of Chemical Society, identifying as gels, systems with 1) a solid-like behaviour and 2) constituted by at least two components, one of them being a liquid present at higher concentration.<sup>48</sup> Hence, a gel is a two-component colloidal dispersion, in which a gelator (the solid, continuous phase) is dispersed within a solvent (the liquid, dispersed phase).<sup>49</sup> The three-dimensional (3D) continuous network of gelators is supported by covalent or noncovalent interactions (ionic bonds, van der Waals forces, hydrogen bonds, or physical entanglements), while the solvent is entrapped by capillary forces and surface tension.<sup>49</sup>

Interestingly, gels are characterised by the elastic properties of ideal Hookean solids and the viscosity properties of Newtonian liquids. Hence, they are referred to as viscoelastic materials.<sup>50</sup> In other terms, gels deform when a force is applied and reverts to their original shape when the force is removed (elastic response); however, when the stress is higher than a certain threshold, the system loses the elastic feature and behaves as a liquid.<sup>51</sup> The gel viscoelasticity can be quantified rheologically through the application of an oscillating stress and it is described by the complex modulus  $G^*$ . The complex modulus is defined as the ratio of the amplitudes of stress (i.e. the amount of force per unit area applied to the sample) and strain (i.e. the dimensionless degree to which the material deforms) and it is expressed as:

$$G^* = G'(\omega) + G''(\omega) \quad \text{Equation 1.3}$$

where  $G'(\omega)$  (or storage module) and  $G''(\omega)$  (or loss module) are the amounts of energy stored and lost during a strain cycle, respectively.  $G'(\omega)$  describes the elastic behaviour typical of solids, whereas  $G''(\omega)$  describes liquids' viscous behaviour. For a system to be classified as gel, the inequality  $G'(\omega) > G''(\omega)$  for at least one order of magnitude must be fulfilled.<sup>52</sup> In addition, the dynamics elastic modulus ( $G'(\omega)$ ) must be relatively independent of the frequency of deformation.<sup>53</sup>

Gels (and other colloidal materials) are thermodynamically unstable systems due to excess of interfacial free energy that exists between the continuous and dispersed phases (the so-called surface effects). Hence, they present a state of higher free energy than material in the bulk.<sup>54</sup> The spontaneous transition to a lower free energy state is prevented by the existence of an energy barrier (free energy of activation -  $E_a$ ) that separates the higher free energy state (gel) from the lower equilibrium state (solid). The presence of this energy barrier kinetically traps the gels in a state of metastability.<sup>49</sup> Gels stability arises from a combination of attractive and repulsive forces at the interfaces between the electrically charged continuous and dispersed phase. Based on the DLVO (Derjaguin and Landau, Verwey and Overbeek) theory, stabilising effects arise from repulsive surface forces, thermal motion of particles and hydrodynamics resistance of the medium. Contrarily, destabilising effects are attributed to attractive surface forces, which correlate to low surface elasticity.<sup>55</sup>

### 1.6.2 Classification of gels

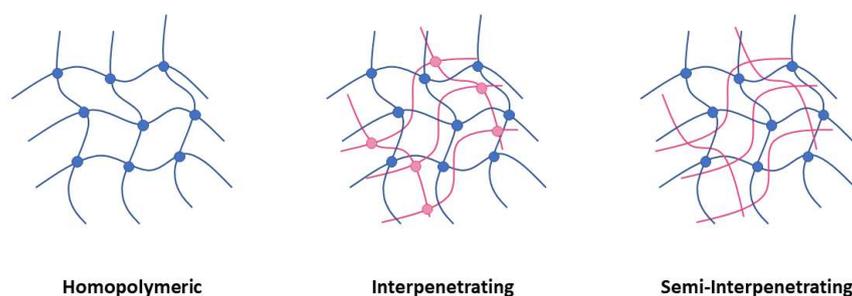
Gels are highly heterogeneous systems which can therefore be classified in several ways. A first and straightforward classification is based on the solvent involved. Gels are distinguished between hydrogels (where the solvent is water), organogels (where the solvent is an organic solvent) or xerogels (where the solvent is replaced by air). In addition, gels can be classified based on their physical appearance (matrix, film, microsphere, etc.), organisation (amorphous, semicrystalline or crystalline), and charge (neutral, ionic, amphoteric, zwitterionic).<sup>56</sup>

Gelators can be classified as i) Low-Molecular Weight Gelators (LMWG) - small molecules (with molecular weight less than 2000 Da) that self-assemble,<sup>57</sup> and ii) Polymeric Gelators

(PGs) - long-chain polymer molecules which are swollen, but not dissolved, by the solvent. <sup>58</sup>

Moreover, gels can be differentiated based on their composition into (Figure 1.13):

1. *Homopolymeric* - the polymeric network derives from a single type of monomer or polymer chain.
2. *InterPenetrating Network (IPN) gels* - two independent cross-linked polymers joint together through orthogonal crosslinking.
3. *Semi-IPN* - as IPN but with one of the polymers not crosslinked. The two components exist independently, without any covalent bonds between them.



**Figure 1.13:** Representation of different types of gels based on their compositions. The dots indicate the establishment of cross-links between the polymer chains. Both in the interpenetrating and semi-interpenetrating networks, the second polymer exists as independent unit. However, just in the semi-interpenetrating network the second polymer does not establish cross-links between its own chains.

Another important classification is based on the kind of interactions that hold the 3D network together. The two main type of interactions are i) covalent interactions for chemically cross-linked gels, and ii) non-covalent interactions for supramolecular gels. The former gives rise to strong and irreversible gels which lack stimuli responsiveness. The latter, instead, are characterised by weaker bonds which can be modulated by environmental changes. (i.e. temperature, pH and salt concentration). This environment-induced changes represent the key difference between physical and chemical gels. Importantly, reversible hydrogels display a sol–gel transition temperature, i.e. the temperature at which the system changes from non-flowing (gel) to flowing (sol) phase. This change occurs when the 3D network is disrupted by heating or other stimuli and cannot contain the water anymore.<sup>59, 60</sup>

## Chapter 2

### Techniques and tools

## 2. Techniques and tools

As has been highlighted in *Chapter 1*, this thesis is focused on three different research macro-topics: (i) the structural and energetic investigation of substrate binding in a class of enzymes able to catalyse glycosyl synthesis; (ii) the structural characterisation (with achievement of atomistic details) of novel functionally modified cellulose-like polymers and (iii) the development of a novel solution state NMR methodology to study solvent structuring at the gel solvent-network interface. In specific, we elected NMR spectroscopy technique to fulfil our research goals. Hence, this chapter includes detailed description of the fundamentals of NMR spectroscopy.

Generally, different techniques are used to study protein-ligand recognition, the most common of which are surface plasmon resonance (SPR) to measure kinetics interaction parameters,<sup>61</sup> and isothermal titration calorimetry (ITC) to achieve insight into thermodynamics properties of binding in solution.<sup>62</sup> In addition, 3D models of protein-ligand complexes can be generated by docking calculations and molecular dynamics methodologies.<sup>63</sup> In the same line, a variety of analytical techniques is normally used for material structural characterisation, covering a span of length scale that goes from bulk properties to supramolecular arrangements and atomistic details. For instance, rheology studies investigate the response of materials to an applied stress, defining the behaviour of elastic solid (Hookean solids), viscous liquid (Newtonian liquid) and viscoelastic material. On the other hand, scattering techniques allow indirect measurement of the 3D organisation of materials as a result of the scattering of an incident wave by the electrons of a crystal. In addition, atomic force microscope (AFM) is used to study material surface.

Finally, three different experimental techniques are able to reach atomistic details of protein, protein complexes, and material arrangements: x-ray diffractions, cryo-EM and NMR spectroscopy. Unfortunately, the crystallisation of carbohydrates, glycoconjugates with proteins, as well as of many supra-molecular edifices and self-assembling systems with extended dynamics domains, is not trivial and the obtained electron density map often does not present high-resolution due to the intrinsic flexibility. On the contrary, cryo-electron microscopy (cryo-EM) requires only that the protein (systems at which cryo-TEM has been applied so far) is in a purified solution. Recently,

Ka Man *et al.*, reported the structural characterisation of the protein apoferritin at atomistic details with a resolution of 1.25 Å.<sup>64</sup>

High-field NMR spectroscopy is one of the most versatile techniques: solution state NMR has been widely used for the structural characterisation of both small molecules and macromolecules in solution, as well as to probe protein-ligands intermolecular interactions with a large range of affinities. NMR spectroscopy probes molecular structure on a different length scale compared with EM. NMR provides information on the local structure by reporting on the immediate vicinity of individual atomic nuclei which present a resonance frequency sensitive to the local electronic environment, while EM provides a molecular envelope picture by probing the electronic potential of the molecule.

In addition, considerable reduction in the experimental time has been achieved with the implementation of non-uniform sampling and non-Fourier signal processing methods in multidimensional NMR, which allowed 7D experiments acquisition.<sup>65</sup> In parallel, greater resolution is obtained by the increased strength of magnetic fields available. Presently, solution state 1.2 GHz NMRs with a magnetic field of 28.2 Tesla (600,000 times stronger than the earth's magnetic field), are located in the University of Florence (Italy), ETH Zurich (Switzerland) and in the Max Planck Institute (MPI) for biophysical Chemistry in Gottingen (Germany).

In addition, solid-state NMR is commonly employed to characterise material short range ordering, the global and local dynamics of biomacromolecules as well as the interactions established with the surrounding water with a site-specific resolution. Solid-state NMR resolution is highly influenced by the rotor spinning rate. To date, high-tech spectrometer with ultra-fast MAS units (>100 kHz) allow acquisition of <sup>1</sup>H spectra in condition which mimics the solution state, highly expanding the use of solid-state NMR. The current record for reported MAS rate is 140 kHz.<sup>66</sup> A major pitfall of higher rate of spinning is that rotors with smaller diameter have to be used, which in turn means smaller sample volume. Therefore, lower quantities of sample can be analysed and longer experimental times are required.<sup>67</sup> Recently, Chen *et al.*,<sup>68</sup> have implemented a novel spherical rotor design which is expected to accommodate higher sample volume for spinning speed up to 150 kHz.

## 2.1 NMR Spectroscopy

### 2.1.1 Solution state NMR<sup>69-71</sup>

NMR is a well-established analytical tool which continues to expand its capabilities and applications. Solution state NMR has been largely preferred for the study of ‘small’ molecules, biomolecules, bio-macromolecules and metabolomics, whereas solid-state NMR has established itself for materials science and biomolecules.

When a sample containing nuclei with spin quantum number  $I = 1/2$  (as  $^1\text{H}$  and  $^{13}\text{C}$ ) is placed in a magnetic field ( $B_0$  – always applied in the z direction in the 3D Euclidean space), the spins can either align with z or with -z and will therefore present two different energy states (as defined by the  $2I + 1$  rule). Spins aligned with z are in lower energy state, called  $\alpha$ , while spins aligned with -z are in a higher energy state, called  $\beta$ . The difference in the  $\alpha$  and  $\beta$  population ( $N_\alpha$  and  $N_\beta$ , for the  $\alpha$  and  $\beta$  states, respectively) is described by the Boltzmann equation

$$\frac{N_\alpha}{N_\beta} = e^{\frac{\Delta E}{k_B T}} \quad \text{Equation 2.1}$$

where  $k_B$  is the Boltzmann constant and  $\Delta E$  is the difference in energy between the  $\alpha$  and  $\beta$  states and it is defined as

$$\Delta E = \hbar\gamma B_0 \quad \text{Equation 2.2}$$

where  $B_0$  is the applied magnetic field and  $\gamma$  is the so-called gyromagnetic ratio, which presents a defined value for each nuclide. Usually, the population of the  $\alpha$  state is slightly higher than the population in the  $\beta$  state and this excess of population results in what is called the “bulk” magnetisation vector  $M_z$  aligned with  $B_0$

$$M_z \propto N_\alpha - N_\beta \quad \text{Equation 2.3}$$

Transitions between the  $\alpha$  and  $\beta$  states can be induced by the application of an electromagnetic wave in the radiofrequency region (called radiofrequency pulse), so that the frequency of the applied pulse matches the nuclei Larmor frequency (the nuclei frequency of precession around the

z-axis deriving from the torque imposed by the static magnetic field  $B_0$  to the nuclei magnetic moment  $\mu$ ). The application of the  $90^\circ$  radiofrequency pulse initiates the movement of the magnetisation from the z to the x-y plane (also called transverse plane). This implies that at thermal equilibrium the net magnetisation on the transverse (x-y) plane to be null. All the NMR experiments can be described in terms of *pulse sequences*, the series of radiofrequency (*rf*) or field gradient pulses applied to manipulate nuclear spins to achieve the desired information.

Four different phenomena represent the core base of all NMR experiments: 1) through-bond interactions *via* scalar (J) spin coupling established by bonding electrons; 2) through-space interactions *via* NOE mediated by dipole-dipole coupling and spin relaxation; 3) chemical exchange, mediated by change of the environment of one spin between two or more states, and 4) molecular self-diffusion which describe the translational movement of molecules or complexes.

The application of a *rf* pulse excite the sample from its thermal equilibrium, but once the *rf* is stopped, the system tries to re-establish the equilibrium condition losing the energy excess by relaxation. The changes in the energy levels induced by relaxation are quite long to occur, which give us time to use a variety of pulse sequences. Consequently, understanding relaxation is pivotal to know how long the nuclei in our sample take to go back to equilibrium conditions. This allow us to timely acquire and accumulate more than one scan for each pulse sequence and increase the S/N ratio.

Importantly, relaxation is induced by fluctuations in the nuclei local magnetic field (the *rf* magnetic field is instead coherently applied to the whole sample) which are random in direction and amplitude. Those fluctuations are driven by random molecular motions, such as diffusional rotation, and are therefore connected with the molecule motional regimes.

### 2.1.1.1 Dynamics by NMR

#### Correlation time and spectral density function

The sample random fluctuations in molecular motions generate a distribution of frequencies in the correlating local magnetic field which is described by the correlation function

$$G(\tau) = \overline{B_{loc}^2} e^{-\frac{|\tau|}{\tau_c}} \quad \text{Equation 2.4}$$

where  $B_{loc}$  is the local field,  $|\tau|$  is the modulus of the time and  $\tau_c$  is the correlation time.  $\tau_c$  is defined as the average time taken for the molecule to rotate through one radian and it is affected by i) the molecular radial volume (in simpler terms, molecular size), ii) the solvent viscosity and iii) the temperature. Generally, for small fast tumbling molecules in low viscous solvent at room temperature, the  $\tau_c$  is in the order of picoseconds. Those molecules are defined in the fast motion or extreme narrowing limit. Slow tumbling macromolecules, on the contrary, are defined in the slow motion or spin diffusion limit and present a  $\tau_c$  in the order of nanoseconds.

From Equation 2.4 we can see that the exponential part of the correlation function is independent from the source of local magnetic field. This allows us to simplify the relation and define a reduced correlation function  $g(\tau)$ , independent on the size of the local magnetic field.

$$g(\tau) = e^{-\frac{|\tau|}{\tau_c}} \quad \text{Equation 2.5}$$

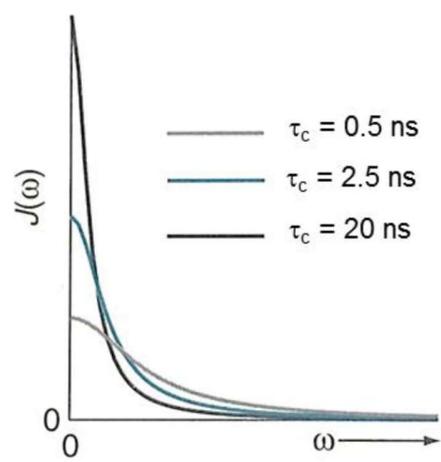
Application of the Fourier Transformation (FT) to the correlation function yields to the spectral density function,  $G(\tau) \xrightarrow{\text{FT}} J(\omega)$ , which described a distribution of molecular motions at the angular frequency expressed in radians ( $\omega = 2\pi\nu$ )

$$J(\omega) = \overline{B_{loc}^2} \frac{2\tau_c}{1 + \omega^2\tau_c^2} \quad \text{Equation 2.6}$$

The discussion proposed above for the correlation function has validity also for the spectral density function, hence the reduced spectral density function is independent from the size of local fields

$$j(\omega) = \frac{2\tau_c}{1 + \omega^2\tau_c^2} \quad \text{Equation 2.7}$$

The dependence of the spectral density for a hard sphere with frequency follows a Lorentzian function and it is dependent from the correlation time (Figure 2.1). In addition,  $J(\omega)$  represents a probability of finding a certain component of the motion at a given frequency  $\omega$ . It is therefore possible to evaluate the amount of motion at the Larmor frequency by evaluating  $J(\omega)$  at  $\omega = \omega_0$ .



**Figure 2.1:** Spectral density function  $J(\omega)$  as a function of different molecular correlation times. The area under the curves is independent from  $\tau_c$ . *Source:* Keeler, 2010<sup>70</sup>

From Figure 2.1 is visible that small molecules that tumble fast in solution have the ability to probe a wide range of motional frequencies and have a higher probability to match the Larmor frequency, reflected in a wide spectral density distribution varying with  $\omega$ . On the contrary, slow tumbling macromolecules (or medium-size molecules in viscous solvent or low temperature) can only experience lower frequencies of motion, reflected in a narrower but more intense distribution of frequencies and the spectral density function sharply peaked at  $\omega = 0$ .

### $T_1$ and $T_2$ spin relaxation

Relaxation involves two different processes: (i) the reestablishment of the equilibrium magnetisation on the z-axis ( $M_z$ ) and (ii) the decay of the transverse magnetisation ( $M_{xy}$ ) to zero. Two different mechanisms are involved:

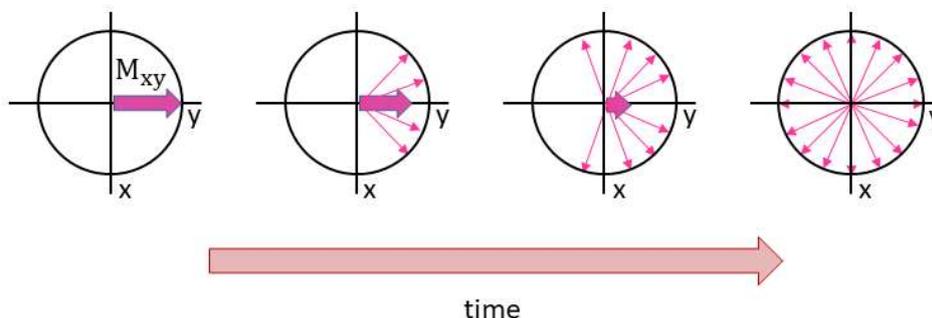
1)  $T_1$  relaxation (also called longitudinal or spin-lattice relaxation): an enthalpic process where energy is transferred to the surrounding in the form of heat. The recovery follows an exponential behaviour described by the equation:

$$M_z = M_0(1 - e^{-t/T_1}) \quad \text{Equation 2.8}$$

where  $M_z$  and  $M_0$  are the magnetization in the z-axis and at thermal equilibrium, respectively, and  $T_1$  is the first-order time constant for this process. Excited nuclear spins present long lifetime

(long  $T_1$ ) because of inefficient means to transfer the low transition energy associated with them. Indeed, spontaneous emission would require so long that is a negligible effect. On the contrary, a magnetic field matching the Larmor frequency of the excited spins can induce spin transition and  $T_1$  relaxation;

2)  $T_2$  relaxation (also called transverse or spin-spin relaxation): an entropic process which destroys the coherence generated on the x-y plane by the application of the radiofrequency pulse. Coherence destruction is induced by local magnetic fields oscillating at the Larmor frequency (as for  $T_1$ ) and by the so-called 'flip-flop' mechanism, caused by swapping of energy between spins (while a spin is excited to the  $\beta$  state, another one drops to the  $\alpha$  state). After the application of a  $90^\circ$  pulse, the resonances moved on the x-y plane will be coherent, hence they have the same phase. Nonetheless, spins are exposed to slightly different magnetic field, which causes the fanning-out of the individual magnetization vectors and finally the destruction of any net magnetization in the transverse plane (Figure 2.2).

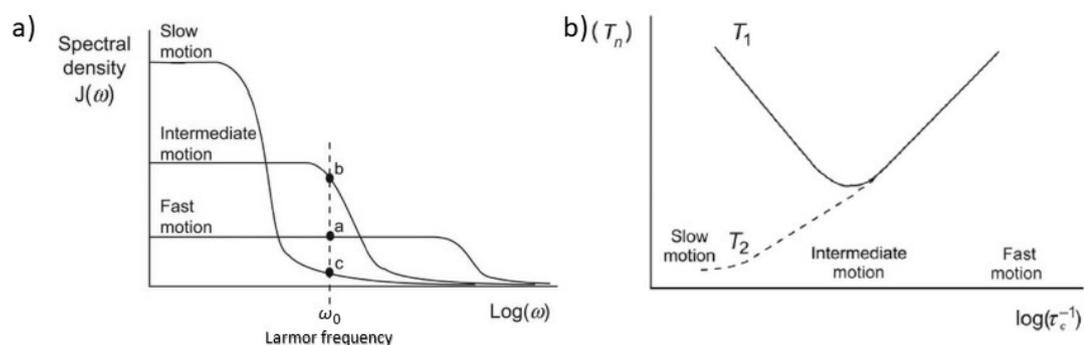


**Figure 2.2:** Fanning out of the magnetisation in the x-y plane with consequently loss in coherence.

Magnetic field differences experienced by the spins system arise from static magnetic field inhomogeneity and from intramolecular and intermolecular interactions in the sample. Generally, anything that causes  $T_1$  relaxation induces also  $T_2$  relaxation, hence  $T_2$  can never be longer than  $T_1$ . On the contrary, due to the flip-flop mechanism contribution,  $T_2$  can be shorter than  $T_1$ .  $T_2$  relaxation, as  $T_1$  relaxation, follows an exponential decay. In addition, the inverse of  $T_1$  and  $T_2$ ,  $R_1$  and  $R_2$  respectively, represent the rate at which the magnetisation recovers and fades out from the transverse plane.  $R_1$  and  $R_2$  are referred as relaxation rate constants ( $s^{-1}$ ).

Fundamental requirement for the relaxation of a spin-1/2 nucleus to occur is a time-dependent magnetic field which fluctuates at the nuclear spin Larmor frequency. This condition is generally matched by local magnetic field generated by dipole-dipole interactions, chemical shift anisotropy (CSA), spin rotation and quadrupolar interactions, while the time dependence correlates with the vibrational, rotational and diffusion motion of the molecule. It is indeed the chaotic tumbling of a molecule to induce spin relaxation. This random motion is characterised by a rotational correlation time  $\tau_c$ .

As described above, the *spectral density* function  $J(\omega)$  (Equation 2.6) describes, for a given frequency, the probability of finding a certain component of the motion. In this case, the frequency of interest is the Larmor frequency. Molecules with different motion regimes will present a different probability in creating a fluctuating magnetic field that matches the Larmor frequency (Figure 2.3) and hence induce  $T_1$  relaxation.



**Figure 2.3:** a) Spectral density function  $J(\omega)$  representing the probability of matching the Larmor frequency for different motional regimes; b)  $T_1$  and  $T_2$  curves represented as a function of molecular tumbling rates. *Source:* Claridge, 2016<sup>69</sup>

Fast tumbling molecules (as small molecules in low viscosity media) present small spectral density components at  $\omega_0$ , which correspond to a long  $T_1$  relaxation. Molecules with intermediate motion present faster  $T_1$  relaxation, but for slowly tumbling molecules the spectral density components drops again and they are correlated with long  $T_1$ . Noticeable, for slow tumbling molecules  $T_1$  is different to  $T_2$ . Indeed, in the slow-motion regime the very low frequency fluctuations stimulate the energy-conserving flip-flop process that induces  $T_2$  relaxation, hence  $T_2 \ll T_1$  (Figure 2.3, b).

The relationship between  $T_2$  and the correlation time can be expressed by the random fields relaxation mechanism,

$$\frac{1}{T_2} = \frac{1}{2}\gamma^2\langle B^2\rangle J(\omega_0) + \frac{1}{2}\gamma^2\langle B^2\rangle J(0) \quad \text{Equation 2.9}$$

where  $\frac{1}{2}\gamma^2\langle B^2\rangle J(\omega_0) = \frac{1}{2}T_1^{-1}$  and represent the  $T_1$  relaxation contribution. In the fast motion (or extreme narrowing limit,  $\omega_0\tau_c \ll 1$ ),  $T_1$  and  $T_2$  have the same value as  $(\omega_0) = J(0)$ . On the contrary, in the slow motion regime (or spin diffusion limit,  $\omega_0\tau_c \gg 1$ ), the contribution of  $T_1$  becomes neglectable and the transverse relaxation becomes proportional to the correlation time  $J(0) = 2\tau_c$ . The faster  $T_2$  relaxation determines line broadening for large molecules like polymers, biological macromolecules and supramolecular complexes with consequently disappearance of the signals.

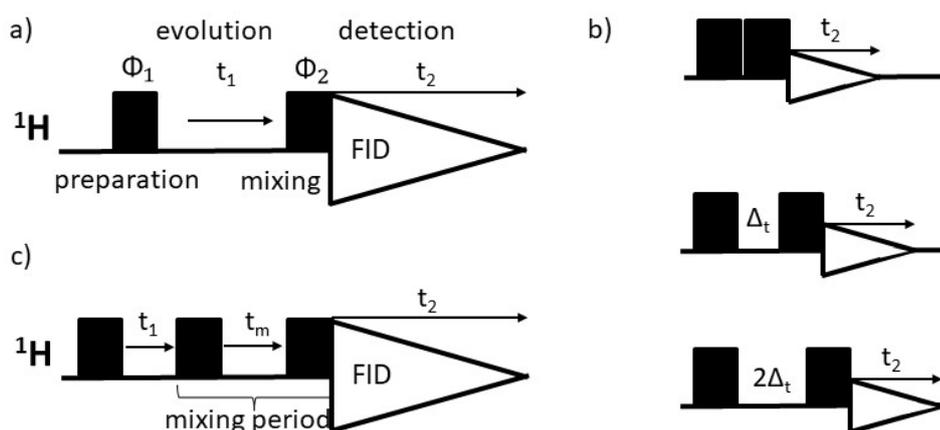
### Chemical exchange<sup>72</sup>

In solution, dynamics events too slow to induce spin relaxation still play a role in what is referred as *chemical exchange*. Examples of these kind of processes can be physical interchange of a molecule between different structural forms, proton-exchange or the on-off equilibria of a ligand-protein binding event. The dynamics regime in which the exchange takes place determines the observed resonances line shape in the spectra. Let us take as example two equally populated sites A and B with equal forward and reverse first order constant  $k$  and with no J-coupling. In a very slow regime, two sharp and well-resolved lines of equal intensity with a  $\Delta\nu = |\nu_A - \nu_B|$  frequency difference will appear in the spectra. As the kinetics constant  $k$  increases (hence faster exchange) the resonances will broaden and the lines come closer, up to the *coalescence point* in which they will merge. After the coalescence point, we enter the fast exchange regime, in which one single sharp peak in the two resonances midpoint is observed, reflecting the average properties of the species involved in the exchange. The peak line shape is defined by exchange narrowing. Hence, it can be noted that it is the exchange rate in reference to the resonance differences  $\Delta\nu_{AB}$  that defines which exchange rate regime we are observing.

Importantly, chemical shift is not the only NMR parameter influenced by chemical exchange, but also longitudinal or transverse relaxation rate constants, scalar couplings and residual dipolar couplings are sensitive to various types of chemical exchange.

### 2.1.1.2 Structural characterisation via NMR

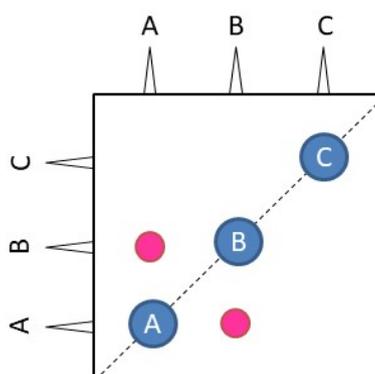
The greatest use of NMR in chemical research is molecules structure verification. In this, 2D methods find great applicability in the elucidation of molecular structure and dynamics by utilizing interactions between spins to correlate them. The standard format of 2D NMR sequences is composed by preparation – evolution – mixing – detection periods. Preparation and mixing periods are pulses or cluster of pulses, while the evolution period introduces a second time variable  $t_1$ , key for the generation of the second dimension of the spectra (Figure 2.4, a).



**Figure 2.4:** a) general scheme for 2D NMR, where the preparation and mixing periods can be either a single pulse or a cluster of pulses and delays, depending on the experiment. The coherence generated during the preparation period evolves during  $t_1$ , and after the mixing period the signal is detected during  $t_2$ ; b) Illustration of how a 2D data set is recorded for a general sequence; c) general scheme for 2D  $^1\text{H}$ - $^1\text{H}$  NOESY experiment.

To better understand how the process works, let us consider the preparation and mixing periods as two hard  $90_x^\circ$  pulses with phases  $\Phi_1$  and  $\Phi_2$ , respectively (Figure 2.4, b). If we set  $t_1$  value to zero, the preparation and mixing pulses will sum up to a hard  $180^\circ$  pulse which would invert the magnetisation from  $z$  to  $-z$  and no signal would be detected. As we increment the length of  $t_1$ , the magnetisation would be moved to transverse plane by the first hard  $90^\circ$  pulse, and here it will

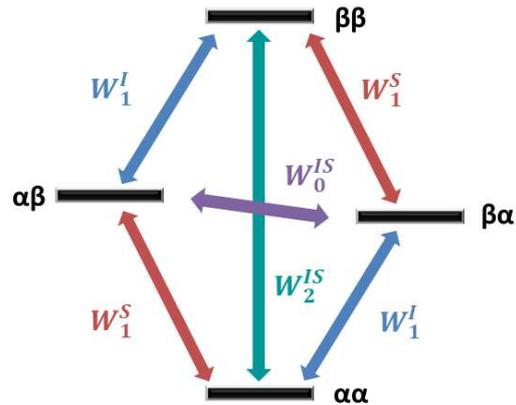
evolve based on its own chemical shift offset and move through an angle of  $360\nu t_1$  degrees. Now, we can consider this vector as composed by two components, one along the y-axis ( $M_0\cos 360\nu t_1$ ) and one along the x-axis ( $M_0\sin 360\nu t_1$ ). With the application of the second hard  $90^\circ$  pulse (mixing pulse), the y-component is moved to the -z-axis, while the x-component is unaffected and will give rise to the detected magnetisation. It is during this period that magnetisation is transferred from one spin to another. Hence, the  $t_2$  FID acquisition and Fourier transformation will give rise to a spectrum where the resonances amplitude varies as a function of  $\sin 360\nu t_1$ , reaching a maximum when the magnetisation vector had time to evolve by a  $90^\circ$  angle, and then decrease, pass through a null, and finally through a negative minimum for larger  $t_1$  (until the signal entirely decays due to relaxation). In this way, the frequency of the modulated amplitude, which corresponds to its chemical shift offset in the rotating frame, has been frequency labelled as a function of  $t_1$ , and the second dimension is indirectly generated. 2D spectra of two uncoupled spins will give rise to two 2D singlets with the same frequency in both dimensions. For two coupled spins, magnetisation transfer via coherence pathways will take place during the mixing pulse, leading to a cross-peak which correlates different shifts in the two dimensions ( $\nu_1 \neq \nu_2$ ). Thus, the resulting spectra will show *diagonal peaks* with identical frequency in both dimensions (large blue dots in the scheme in Figure 2.5), and the cross peaks featuring the two frequencies of the correlating protons (small pink in the scheme in Figure 2.5).



**Figure 2.5:** Schematic representation of a 2D NMR spectrum, the pink dots representing the cross-peak for the coupled protons A and B.

The sequence described so far allows the detection of homonuclear  $^1\text{H}$ - $^1\text{H}$  through bond correlations - the famous COSY experiment. For the performance of 2D  $^1\text{H}$ - $^1\text{H}$  through space correlation (NOESY experiment), instead, the mixing period consists of a  $90^\circ$  pulse, followed by a mixing time ( $t_m$ ) and a final  $90^\circ$  pulse before detection (Figure 2.4, c). In this sequence, the initial excitation pulse places the magnetisation vector in the transverse plane, which will evolve for the  $t_1$  evolution time. The second  $90^\circ$  pulse places one component of the magnetisation onto the  $-z$  axis, generating the population inversion required for transient NOE (further explained below). Hence, the mixing time allows the incoherent transfer of magnetisation between spins and hence NOE to develop. Finally, the new population is sampled with a final  $90^\circ$  pulse and the FID is collected.

To explain the coherence transfer process we can consider a coupled two-spin system,  $I$  and  $S$ , for which the transverse magnetisation created by the application of a  $90^\circ$  pulse can be represented by 4 energy levels ( $\alpha_I\alpha_S, \alpha_I\beta_S, \beta_I\alpha_S, \beta_I\beta_S$ ) with six possible relaxation pathways associated (Figure 2.6). The so-called longitudinal relaxation promotes the transition between these levels, each process presenting an associated transition probability  $W$ .  $W_1^I$  and  $W_1^S$  are denoted single quantum spin transitions, which correspond to the  $\alpha$  and  $\beta$  interconversions ( $\alpha\alpha \Leftrightarrow \beta\alpha$ ,  $\alpha\alpha \Leftrightarrow \alpha\beta$ ,  $\alpha\beta \Leftrightarrow \beta\beta$  and  $\beta\alpha \Leftrightarrow \beta\beta$ ), and are allowed based on the quantum-mechanic selection rules (which allow only transition for the total spin quantum number  $I$  corresponding to  $\Delta I = I(\text{initial state}) - I(\text{final state}) = 1$ ).  $W_2^{IS}$  and  $W_0^{IS}$  denoted the “double quantum transitions” (both spins flipped in the same direction,  $\alpha\alpha \Leftrightarrow \beta\beta$ ) and the “zero quantum transitions” (both spins flipped in opposite directions,  $\alpha\beta \Leftrightarrow \beta\alpha$ ), respectively. These are the two transitions that are not allowed by the selection rule but can be indirectly detected in 2D NMR.



**Figure 2.6:** Schematic representation of the energy level diagram of a coupled two-spin  $IS$  system and the six possible transitions.

The corresponding transition probabilities (or rates) are:

$$W_1^S = \frac{3}{40} b^2 j(\omega_{0S}) \quad \text{Equation 2.10}$$

$$W_1^I = \frac{3}{40} b^2 j(\omega_{0I}) \quad \text{Equation 2.11}$$

$$W_0^{IS} = \frac{1}{20} b^2 j(\omega_{0I} - \omega_{0S}) = \frac{1}{20} b^2 j(0) \quad \text{Equation 2.12}$$

$$W_2^{IS} = \frac{3}{10} b^2 j(\omega_{0I} - \omega_{0S}) = \frac{3}{10} b^2 j(2\omega_{0I}) \quad \text{Equation 2.13}$$

where  $j(\omega)$  is the reduced spectral density at the frequency of transition and  $b$  is the dipolar constant, defined as:

$$b = \frac{\mu_0 \gamma_I \gamma_S \hbar}{4\pi r^3} \quad \text{Equation 2.14}$$

with  $\gamma$  being the gyromagnetic ratio,  $\mu_0$  is the permeability of vacuum and  $r$  is the internuclear distance.

A rate equation for the flow of population for each level (denoted  $N_1$ ,  $N_2$ ,  $N_3$  and  $N_4$ ) can be written; for example, the rate equation for level 1 would be (similar equations can be written for the other three energy levels):

$$\frac{dN_1}{dt} = -W_1^S N_1 - W_1^I N_1 - W_2^{IS} N_1 - W_1^S N_2 - W_1^I N_3 - W_2^{IS} N_4 \quad \text{Equation 2.15}$$

At this point, to be able to follow the magnetisation of spins  $I$  and  $S$  on the  $z$ -axis, we will need to resort to the product operator formalism (even though a full explanation of this approach is not included in this thesis). From this approach two operators  $I_z$  and  $S_z$ , representing the magnetisations of spins  $I$  and  $S$  on the  $z$ -axis, are defined. The simplified form of the differential equations expressing the deviation of population of each energy level,  $dI_z/dt$  and  $dS_z/dt$ , (the so-called Solomon equations) are herein reported.

$$\frac{d(I_z - I_z^0)}{dt} = \rho_{zI}(I_z - I_z^0) - \sigma_{IS}(S_z - S_z^0) \quad \text{Equation 2.16}$$

$$\frac{d(S_z - S_z^0)}{dt} = \rho_{zS}(S_z - S_z^0) - \sigma_{IS}(I_z - I_z^0) \quad \text{Equation 2.17}$$

where  $\rho_{zI}$  and  $\rho_{zS}$  are the auto-relaxation (longitudinal relaxation) rate constants, and  $\sigma_{IS}$  is cross-relaxation rate constant which defines the magnetisation transferred from spin to spin. The presented rate constants defined as:

$$\rho_{zI} = W_0^{IS} + 2W_1^I + W_2^{IS} \quad \text{Equation 2.18}$$

$$\rho_{zS} = W_0^{IS} + 2W_1^S + W_2^{IS} \quad \text{Equation 2.19}$$

$$\sigma_{IS} = W_2^{IS} - W_0^{IS} \quad \text{Equation 2.20}$$

From the Solomon equations it is evident that, when the cross-relaxation constant is not null, the rate of recovery of spins  $I$  and  $S$  depends on each other

Assuming that  $\omega_{OI} = \omega_{OS}$ , hence the spins of interest are of the same type (e.g.  $^1\text{H}$ ), with comparable Larmor frequencies, the homonuclear cross relaxation in Equation 2.20 can be rewritten as:

$$\sigma_{IS} = W_2^{IS} - W_0^{IS} = \frac{3}{10} b^2 j(2\omega_{0I}) - \frac{1}{20} b^2 j(0) \quad \text{Equation 2.21}$$

As aforementioned, the double and zero quantum transitions are not allowed by the conventional selection rules, still they can strongly influence the observable NMR intensities associated to the transition of the spins involved in the process of cross relaxation.

### Nuclear Overhauser effect

The Nuclear Overhauser effect (NOE) is a cross-relaxation process stimulated by dipolar coupling and it manifests as a change in intensity of one resonance (called NOE enhancement even when a reduction of the signal is recorded) consequent to the perturbation of the equilibrium populations of a dipolarly coupled spins.

Let us consider a homonuclear ( $^1\text{H}$ - $^1\text{H}$ ) dipolar coupled two-spin system  $IS$  contained in a rigid molecule that tumbles isotropically in solution, i.e., it does not show any preferential axis about which to rotate. In this idealistic system both protons do not present scalar coupling ( $J_{IS} = 0$ ) but, due to their proximity, they share dipolar coupling (magnetic interaction through space). In other words, both protons can sense the presence of the other dipolar-coupled partner.

Perturbation of nucleus  $S$  resonance can be achieved by:

- 1) selective saturation and consequently equalisation of the spin populations with a weak continuous wave (CW) rf irradiation – the so-called steady-state NOE. Application of this saturating pulse results in the disappearance of the peak corresponding to  $S$ , and in an enhancement in the peak corresponding to  $I$ , proportional to  $\sigma_{IS}$ ;
- 2) a shaped  $180^\circ$  pulse, which invert the  $S$  resonance population. In this pulse sequence, after the first  $180^\circ$  pulse, a mixing time  $t_m$  is allowed for the magnetisation to evolve, hence we are in the presence of a transient process and the experiment is called transient NOE.

The sign and magnitude of the NOE enhancement ( $\eta_I\{S\}$ ) is proportional to the cross-relaxation rate constant  $\sigma_{IS}$

$$\eta_I\{S\} \propto \sigma_{IS} = W_2^{IS} - W_0^{IS} \quad \text{Equation 2.22}$$

and it is expressed as

$$\eta_I\{S\} = \frac{I - I_0}{I_0} \cdot 100 \quad \text{Equation 2.23}$$

where  $I_0$  is the intensity at the equilibrium, and  $I$  is that in the presence of NOE.

Selective saturation of  $^1\text{H}$  nucleus  $S$  perturbs the spin populations of nucleus  $I$ , which will try to go back to the initial equilibrium situation via different relaxation mechanisms. Nonetheless, only the cross-relaxation pathways, characterised by  $W_0$  and  $W_2$  transition probabilities, are responsible for the NOE development (Equation 2.22).  $W_0$  and  $W_2$  cross-relaxation pathways are in competition within each other and the dominant mechanism dictates the sign of the observed NOE. Importantly, slow tumbling molecules (short correlation time  $\tau_c$ ) generate fluctuating local magnetic fields which induce the lower energy  $W_0$  process, such that a negative NOE ( $I < I_0$ ) is effectively recorded. On the contrary,  $W_2$  (higher energy) is the dominant mechanism in fast tumbling molecules, so that a positive NOE ( $I_0 > I$ ) is observed. Hence, depending on the magnitude of the NOE developed, two limits of motions have been defined: the extreme-narrowing limit for fast tumbling molecules, and the spin diffusion limit for the slow tumbling ones. Furthermore, low-frequency are more efficient than high-frequency magnetic fields in activating the cross-relaxation pathways, due to the different spectral density function  $J(\omega)$  that describe the fast, intermediate and slow motion (Figure 2.1). Therefore, positive NOE will have lower energy compared to negative NOE.

Even though  $W_0$  and  $W_2$  cross-relaxation pathways are the ones responsible for NOE, single quantum relaxation pathways ( $W_1$ ) are also active in trying to re-establish the equilibrium population differences of the non-saturated nucleus  $I$ . Hence,  $W_1$  relaxation operates against the NOE build-up, and if it is more efficient than  $W_0$  and  $W_2$  pathways, NOE might not be observed. So, while the final sign of NOE depends on the  $W_2 - W_0$  difference, its magnitude depends on the three  $W_1$ ,  $W_0$  and  $W_2$  rates (as expressed by Equation 2.24, derived from the Solomon equation).

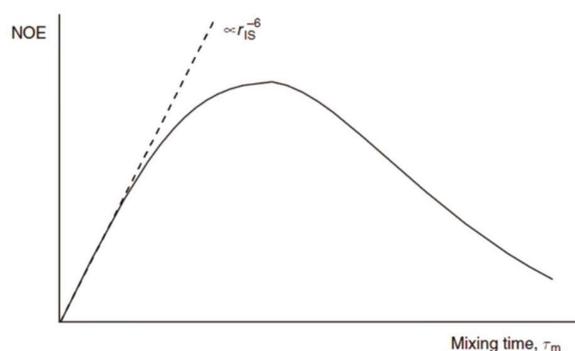
$$\eta_I\{S\} = \frac{\gamma_S}{\gamma_I} \left[ \frac{W_2 - W_0}{W_0 + 2W_1 + W_2} \right] = \frac{\gamma_S}{\gamma_I} \left[ \frac{\sigma_{IS}}{\rho_{IS}} \right] \quad \text{Equation 2.24}$$

In the case of homonuclear coupling of spin-1/2 nuclei and for fast-tumbling molecules, steady-state experiments can give rise to a maximum 50% NOE enhancement, while transient experiments develop weaker positive NOE enhancement (38% maximum). Hence, careful choice of the correct mixing time is crucial for the success of the experiments. On the contrary, for slow-tumbling molecules the efficient cross-relaxation determines that both transient and steady-state experiments will give rise to maximum negative enhancement (-100%); still, complication may arise from spin diffusion, hence transient experiments with short mixing time periods are preferred.

### Distances measurement and transfer NOESY

NOE is a distance dependent effect, and in specific it is inversely proportional to  $r_{IS}^6$ . Importantly, in the steady-state NOE,  $\eta_I\{S\}$  does not depend only on the cross-relaxation rate, but also on spin  $I$  longitudinal relaxation properties, and results from a balance between the influence of all neighbouring spins. Hence only qualitative measurement of distances can be obtained. On the contrary, transient NOE  $\eta_I\{S\}$  depends only on  $\sigma_{IS}$  and on the mixing time (which is experimentally set) and hence it is an excellent tool for the measurement of inter-nuclear distances between two spins based on the inverse and linear proportionality of  $\sigma_{IS}$  and  $r_{IS}^6$ . For this to be valid, the so-called *initial rate approximation* must be fulfilled.

In transient NOE, the NOE kinetics is monitored by inverting the spins population differences of the target resonance/s and then allowing the NOE to develop during the mixing time  $t_m$ , in which no further external interference is applied. Under those conditions, the two cross-relaxing spins will at first behave as an isolated spin pair and the growth of NOE will follow a linear dependence within the increasing mixing time. However, for longer mixing times, the NOE build-up curve will start deviating from linearity and eventually decay to zero due to relaxation of spin  $I$  (Figure 2.7).



**Figure 2.7:** Development of the transient NOE between two spins as a function of mixing time ( $\tau_m$ ).

Source: Claridge, 2016<sup>69</sup>

Hence, mixing periods comparable to  $T_1$  provide maximum enhancements. Nonetheless, to achieve a meaningful distance measurement (and *initial rate approximation* to be valid) we must use mixing times significantly shorter than the  $T_1$  relaxation time of spin I, where the NOE grows linearly. In such conditions, the NOE enhancement of spin I after inversion of spin S ( $\eta\{S\}$ ) after a period  $\tau_m$  is proportional to the cross-relaxation rate ( $\sigma_{IS}$ ) which in turn depends on  $r_{IS}^{-6}$

$$\eta\{S\} = k\sigma_{IS}\tau_m = k'r_{IS}^{-6}\tau_m \quad \text{Equation 2.25}$$

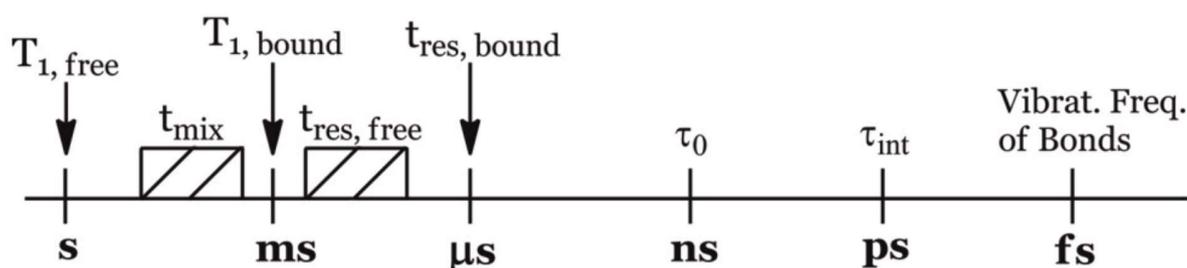
where  $k$  and  $k'$  are constants of proportionality which contain, among other known physical constant, the overall correlation time of the molecule,  $\tau_c$ . Hence, for known  $\tau_c$ ,  $r_{IS}$  can be derived. However, a more common practice is to use as reference a known internal distance between two nuclei (A and B), for which the reference NOE is measured. In this way, the direct comparison of the two NOE intensities provides the unknown internuclear distance, such as

$$\frac{\eta_I\{S\}}{\eta_A\{B\}} = \frac{r_{IS}^{-6}}{r_{AB}^{-6}} \quad \text{Equation 2.26}$$

For isotropically tumbling molecules the assumption that all internuclear vectors possess the same correlation time and therefore that the *initial rate approximation* is valid for all interactions allow measurement of internuclear distancing within a single experiment. Nonetheless, this is not the case when internal motions are the main contributors in proton-proton vectors reorientation.

### 2.1.2 Ligand-based NMR spectroscopy for binding studies

tr-NOESY and STD NMR experiments are robust and powerful ligand-based NMR techniques used to study the carbohydrates conformation in the bound state and glycan-protein interaction at an atomic level (ligand binding epitope), respectively. The medium-weak binding affinities that characterise protein-glycan interactions are the perfect range for the characterisation of the kinetics of binding, also referred as kinetics of exchange since the ligand exchange between the free and the bound state. When the residence time  $t_{res}$  of the ligand in the free state is much higher than that of the bound state, the transient interactions established within the weak-binding ligand and the macromolecular receptor allow the observation of intra- and/or inter-molecular NOE enhancements. On the contrary, when the binding is too tight, magnetisation is lost due to the efficient  $T_1$  relaxation typical of macromolecules, precluding NOE to develop prior to the acquisition period. Thus, it is important to consider the different timescales of the mechanisms involved into macromolecular recognition of low-affinity ligand (Figure 2.8).



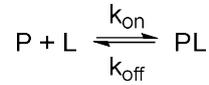
**Figure 2.8:** Timescales schematic representation of the distinct mechanism taking place during molecular recognition of low-affinity ligands to macromolecular receptors. The differences between  $T_{1, free} - T_{1, bound}$  and  $t_{res, free} - t_{res, bound}$  allow the observation of the intermolecular NOEs.

In addition, ligand-observed methods are not limited by the protein molecular size and therefore have great applicability for analysing protein–ligand interactions.

#### 2.1.2.1 The equilibrium kinetics of binding: one-site model for bimolecular association in solution

In the so-called “one-site model for bimolecular association in solution” (which does not account for any allosteric and cooperative effects), three species are involved: a protein receptor (P), a ligand (L) and the complex of their association, the complex (PL). Thus, assuming that the protein

receptor P has a single binding site for the ligand L and that they interact transiently to form the complex PL, the binding equilibrium can be described by a two states model



The temperature associated dissociation equilibrium constant  $K_D$  of the complex is expressed by

$$K_D = \frac{[P] \cdot [L]}{[PL]} = \frac{k_{off}}{k_{on}} \quad \text{Equation 2.27}$$

where the unimolecular rate constant for the dissociation reaction ( $k_{off}$ ) is inversely related to the half-life time of the protein-ligand complex or, equivalently, the residence time of the ligand in the bound state  $t_{res}$  (Figure 2.8). The bimolecular rate constant for the direct reaction ( $k_{on}$ ) measures the probability of a productive interaction between the protein and the ligand (e.g. encounter that leads to the complex formation). A  $k_{on}$  value of  $10^7 - 10^9 M^{-1} s^{-1}$  has been estimated for association process controlled by molecular diffusion. However, if the complex formation involves strong long-range forces or large conformational changes,  $k_{on}$  could be significantly larger or significantly lower, respectively. For systems in which we can assume  $k_{on}$  diffusion-limited,  $K_D$  is in direct proportion to  $k_{off}$ , hence measuring the affinity gives us direct information on ligand residence time in the binding pocket (turnover of binding). Small  $K_D$  values (strong binding) correspond to small  $k_{off}$ , indicating slow turnover of the ligand in the binding site and long residence time. On the contrary, large  $K_D$  values (weak binding) correspond to large  $k_{off}$ , hence fast turnover and a short residence time.

The fraction of the bound protein  $f_{PB}$  is given by

$$f_{PB} = \frac{[PL]}{[P] + [L]} \quad \text{Equation 2.28}$$

Combination of Equation 2.27 with Equation 2.28 allow us to express the fraction of bound receptor as a function of the concentration of free ligand, i.e. the equation of a Langmuir isotherm for the saturation of a receptor site of interaction with the ligand.

$$f_{PB} = \frac{[L]}{[PL] + K_D} \quad \text{Equation 2.29}$$

Equation 2.29 shows that increasing the ligand concentration  $[L]$ , the fractional occupation of the receptor-binding site  $f_{PB}$  will rise in a hyperbolic function of  $[L]$ . For low ligand concentration,  $[L] \ll K_D$ ,  $f_{PB}$  is directly proportional to the concentration of the free ligand  $[L]$ . In the other extreme, hence for very large ligand concentration,  $[L] \gg K_D$ , the receptor is completely saturated by the ligand ( $f_{PB} = 1$ ). Thus, for ligands with large  $K_D$  higher concentrations are needed to saturate the receptor-binding site in comparison with ligands with low  $K_D$ . When the receptor is half saturated ( $f_{PB} = 0.5$ ),  $[L] = K_D$ , from which derives the  $K_D$  definition as “the concentration of ligand in the free state that leads to half saturation of the receptor”. Importantly, in the Langmuir isotherm,  $[L]$  represent the concentration of the free ligand and not the total added concentration  $[L]_T$ .

Importantly, chemical exchange processes are established during the transient presence of the ligand in the protein receptor site, this chemical exchange being characterised by the exchange kinetics constant  $k_{ex}$

$$k_{ex} = k_{on} \cdot [P] + k_{off} \quad \text{Equation 2.30}$$

during which the ligand transiently adopts NMR parameters typical of the much larger receptor and, at the same time, transiently perturbs the microenvironment of the receptor binding site. Thus, if  $Q$  is a NMR parameter (e.g. chemical shift, relaxation rates, translational diffusion coefficient), in both cases  $Q_F \neq Q_B$ , i.e.  $\Delta Q \neq 0$  (F and B refer to the free and the bound state, respectively).

Monitoring  $\Delta Q$  and  $\Delta\nu$  upon binding, three scenarios are possible:

- 1)  $k_{ex} \ll \Delta Q$  and  $k_{ex} \ll \Delta\nu$ ;
- 2)  $\Delta\nu \gg k_{ex} \ll \Delta Q$ ;
- 3)  $k_{ex} \gg \Delta Q$  and  $k_{ex} \gg \Delta\nu$ ;

In case 1), the exchange is slow relative to both time scales, therefore we observe two resonances and can measure both  $Q_F$  and  $Q_B$ . In case 2), the exchange is slow relative to the chemical shift

time scale and fast relative to  $Q$ , so we observe two signals of which we can measure an exchange-averaged parameter  $Q$ ,  $\langle Q \rangle$

$$\langle Q \rangle = P_B Q_B + P_F Q_F \quad \text{Equation 2.31}$$

where  $P_F$  and  $P_B$  are the populations in the free and bound state, respectively.

In case 3), the exchange is fast relative to both time scales, and we have a single signal on which we can equally measure the average observed  $\langle Q \rangle$ .

All the ligand-based NMR screening experiments assume that the receptor-ligand binding is in the fast exchange regime, with the ligand presenting an average observed  $\langle Q \rangle$ . Observed differences between  $\langle Q \rangle$  and  $Q_F$  provide measurable evidence of receptor binding. To detect binding with adequate sensitivity a larger contribution from the bound state ( $P_B Q_B$ ) than the free state ( $P_F Q_F$ ) is necessary. In typical screening conditions  $[L]_T \gg [P]_T$ , and *in turn*  $P_B \ll P_F$ , hence it is necessary that  $\langle Q \rangle$  is amplified in the bound state (i.e.  $Q_B \gg Q_F$ ). However, if  $k_{on}$  is well approximated by a diffusion-limited value, then the slowest  $k_{ex}$  constants lie within the  $10^3 < k_{ex} < 10^5 \text{ s}^{-1}$  range for weak-affinity ligands with a  $K_D$  in the  $\mu\text{M}$  range. Hence for  $^1\text{H}$  observation,  $k_{ex}$  would consequently exceed most differences in intrinsic  $^1\text{H}$  relaxation rates and rotating frame precession frequencies, thus supporting the validity of the fast exchange assumption.

### 2.1.2.2 Exchange-Transferred-NOESY (tr-NOESY)<sup>73, 74</sup>

Tr-NOESY is a well-known 2D NMR technique to study the bound ligand conformation in solution. Technically, tr-NOESY is purely a NOESY experiment applied to a sample with the appropriate receptor and ligand ratio.

The tr-NOESY experiment relies on the ligand change in correlation times  $\tau_c$  between the free and bound states. The free ligand (small molecule) has short  $\tau_c$ , slow NOE build-ups and no spin diffusion, hence it shows small positive NOEs. On the contrary, the bound ligand acquires the motional properties of the macromolecule and it exhibits long  $\tau_c$ , rapid NOE build-ups, extensive spin diffusion, and strong negative NOEs (i.e. transferred-NOEs). Hence, discrimination between the NOEs of the free ligand in solution and the tr-NOEs originating from the bound state can be

accomplished easily by looking at the sign and size of the observed NOEs (Figure 2.9, a), or by looking at the NOE build-up rate (i.e. the time required to achieve NOE maximum intensity) usually in the range of 50-100 ms for tr-NOEs and four-to-ten times longer for small molecules (Figure 2.9, b).

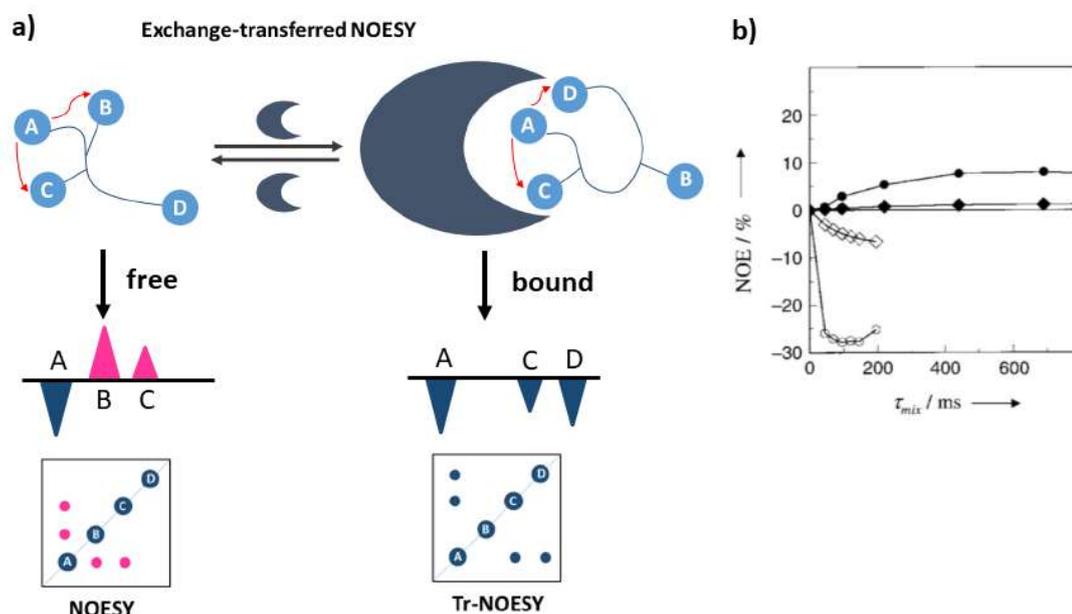
In the case of glycan-protein complexes, the fast kinetics of interaction ( $\mu\text{M} < K_D < \text{mM}$ ) is the ideal condition for the observation of tr-NOEs, while no tr-NOEs are observed for too tight bindings. In the fast exchange timescale, the perturbations in the populations of the ligand protons in the bound state (which derive from the intra-ligand longitudinal  $^1\text{H}$ - $^1\text{H}$  relaxation processes) are transferred to the bulk solution, where they can accumulate due to the smaller  $R_1$  of small molecules. Hence, it is possible to observe strong negative NOEs in the NMR signal of the free state.

The ultimately observed longitudinal cross relaxation rate or NOE ( $\sigma_{Tr-N}$ ) is given by the population-weighted average of the NOEs in the free and the bound state:

$$\sigma_{Tr-NOE} = f_F^L \sigma_F^{NOE} + f_B^L \sigma_B^{NOE} \quad \text{Equation 2.32}$$

where  $f_F^L$  and  $f_B^L$  are the fraction of free and bound ligand, and  $\sigma_F^{NOE}$  and  $\sigma_B^{NOE}$ , the NOEs in the corresponding free and bound states.

Hence, to detect NOEs from the bound state, the condition  $|f_B^L \sigma_B^{NOE}| \gg |f_F^L \sigma_F^{NOE}|$  must be fulfilled, with a ligand:protein ratio ( $[L]_T/[P]_T$ ) between 10:1 to 50:1. On the contrary, if the used  $[L]_T/[P]_T$  is too high,  $f_F^L \gg f_B^L$ , the elevated fraction of free ligand in solution will result in the detection of the small molecule positive NOEs ( $\sigma_F^{NOE}$ ) and in the reduction or even cancellation of the negative NOEs coming from the small fraction of bound ligand.



**Figure 2.9:** a) Schematic representation of a NOESY (left) and tr-NOESY (right) spectra. For the small molecule in the free state, cross-peaks show opposite sign to the diagonal peaks (positive NOEs). Upon addition of the receptor, the cross-peaks change to negative NOEs, same sign as the diagonal peaks; b) Nuclear Overhauser enhancements (NOEs) and tr-NOEs for  $\alpha$ -L-Fuc-(1 $\rightarrow$ 6)- $\beta$ -D-GlcNAc-OMe in the absence (filled symbols) and presence (open symbols) of *Aleuria aurantia* agglutinin, measured at 600 MHz as a function of the mixing time  $t_m$ . Circles and diamonds refer to proton pairs H6proR<sup>GINAC</sup>- H6proS<sup>GINAC</sup> and H1<sup>Fuc</sup>- H6proS<sup>GINAC</sup>, respectively. Source: a) Figure adapted from the Doctoral Thesis of J.C Muñoz-García<sup>75</sup> and b) Claridge, 2016<sup>69</sup>

Tr-NOESY is a widely used experiment to determine binding activity of single ligands and ligand libraries, and to determine the bioactive conformation of the ligand in the binding site. However, those protons of the receptor close to the ligand in the bound state can produce relay NOEs that can affect the intra-ligand NOEs (protein mediated spin diffusion). These indirect (protein-mediated) tr-NOE effects can give rise to negative cross-peaks between protons that are far apart and lead to errors in the analysis of the ligand bound conformation. To avoid this, we must resort to the use of short mixing times (analysis of the NOE build-up curves initial slopes), to the use of a per-deuterated receptor (if feasible)<sup>74</sup> or to tr-ROESY experiments, which allow distinction between direct and indirect NOE cross-peaks.<sup>76</sup>

### 2.1.2.3 Saturation Transfer Difference (STD) NMR<sup>77, 78</sup>

The STD NMR experiment is another ligand-based spectroscopic technique to study protein-ligand interactions in solution, and, alike tr-NOESY, it is based on the NOE effect. STD NMR has been traditionally used in pharmaceutical research to identify hits from compound library screening in drug discovery. Indeed, it can be applied to any system under chemical exchange in which the molecules involved in the exchange show relatively sharp NMR signals in the spectrum.

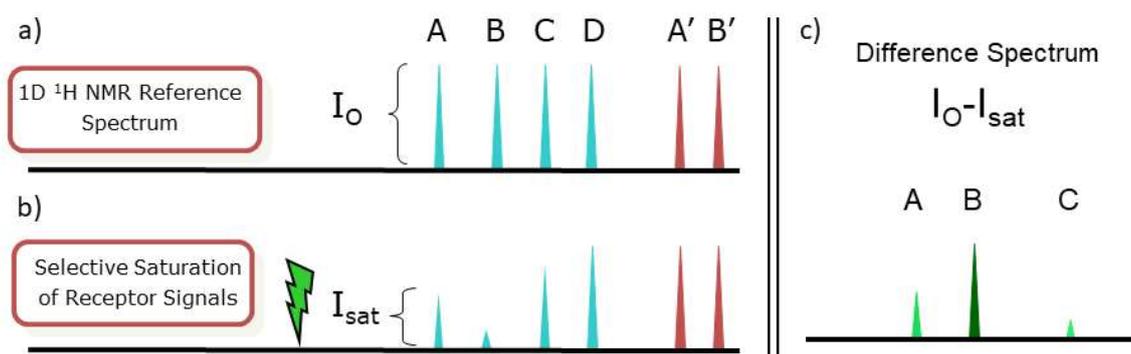
The observation of the magnetisation transfer from A to B (e.g. from protein to ligand) is implemented by selectively saturating the resonance of protons belonging to A and monitoring the transfer to the other B. From the ligand perspective, STD optimal conditions are represented by  $k_{ex} \gg \Delta\nu$  and  $k_{ex} \gg \Delta R_1, \Delta R_2$  where  $\Delta\nu$ ,  $\Delta R_1$  and  $\Delta R_2$  are the difference between the  $\nu$ ,  $R_1$ s and  $R_2$ s of the ligand in the free and bound state. The assumption  $k_{ex} \gg \Delta R_1$  is important to avoid loss of the spin perturbation before the exchange is completed. Thus, the free and bound ligand must be in a fast exchange condition.

In such fast kinetics condition, it is necessary to use a large excess of ligand to saturate the protein binding sites (usually 1:10 up to 1:1000 protein-to-ligand ratio, which means that the chemical exchange takes place with highly unequal populations). In the described conditions (fast exchange and large excess of ligand) the signals of the ligand in the NMR spectrum will mostly correspond to the free form, with the protein presenting weak broad signals on the baseline. Hence, direct observation of resonances of the bound ligand is not possible, and therefore neither their selective perturbation. This is why STD NMR resort on the selective saturation of the protein amino acid resonances, where the vast network of  $^1\text{H}$ - $^1\text{H}$  cross-relaxation pathways will allow the saturation to quickly spread over the entire macromolecule (via intramolecular NOE) and then to the bound ligand via intermolecular NOE. Due to this cross-relaxation mechanism, upon contacting the saturated protein the ligand will acquire the negative NOE characteristic of the slow-tumbling molecules, resulting in the decrease of the ligand signals.<sup>77</sup> The dissociation of the ligand will then transfer the saturation into the bulk solution, where it will accumulate during the saturation time. Assuming that  $k_{ex} \gg \Delta R_1$ , or  $k_{off} \gg R_1$ , means that the ligand does not lose the magnetisation transferred before to go back in the bulk solution. Since we monitor the signals of the free ligand in solution, saturation transfer is only detected for complexes with dissociation constant  $K_D$  in the

range of  $10^{-3} - 10^{-8}$  M. Hence, this results in the macroscopic detection of transferred saturation on the ligand signals in the saturated STD NMR spectrum.

Factors affecting the amount of saturation transferred to the binder are: i) the macromolecular mobility, ii) the lifetime of the complex and iii) the geometry of binding (binding mode, which depends on the bioactive conformation of the ligand, and the intrinsic architecture of the binding pocket).

From a practical perspective, implementation of STD NMR experiments is quite easy. A sample containing the receptor (MW >10 kDa) at low concentration ( $10^{-5}$  to  $10^{-6}$  M) and a larger concentration of a pool of small test compound is prepared. Hence two different experiments are subsequently recorded: 1) the 1D  $^1\text{H}$ -NMR *off-resonance* experiment, from which the reference spectrum is obtained (Figure 2.10, a), and 2) the 1D  $^1\text{H}$ -NMR *on-resonance* experiment, from which a spectrum with reduced intensities of the binder molecules is obtained (Figure 2.10, b). From the subtraction of spectrum 2 from spectrum 1, the difference spectrum is obtained (Figure 2.10, c).



**Figure 2.10:** Illustration of the 1D STD NMR experiment applied to a sample containing a protein receptor in the presence of a mixture of two small test compounds (a binder – resonances A, B, C and D, and a non-binder – resonances A' and B') in molar excess. Two 1D NMR spectra are recorded: a) a standard 1D  $^1\text{H}$  NMR (*off-resonance* or *reference* spectrum), and b) a 1D  $^1\text{H}$  NMR spectrum with selective saturation of receptor proton signals. The difference spectrum ( $I_0 - I_{\text{sat}}$ , c) will show only signals corresponding to the ligand, while signals of non-binder cancel out. *Source:* Figure adapted from Angulo, 2010<sup>79</sup>

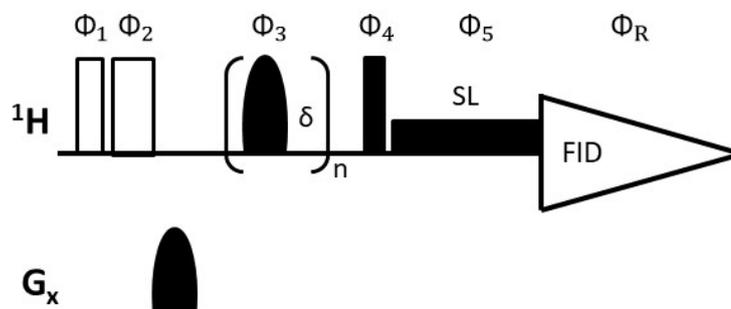
In the first recorded experiment the irradiation frequency applied to the sample under thermal equilibrium conditions is set to a value far from any ligand or protein signal (e.g. 50 ppm), thus the

spectrum presents full intensities for the ligand resonances ( $I_0$ ). On the contrary, during the acquisition of the second experiment low power Gaussian-shaped pulses selectively irradiates some protons of the receptor (but not ligand signals), e.g. the aliphatic (from 0 to -1 ppm) or aromatic region (around 7 ppm). The applied saturation time can vary from 0.1 to 20 seconds, depending on experimental needs. The existence of a non-equilibrium magnetisation (saturation) on the  $^1\text{H}$  nuclei of the receptor leads to intermolecular NOEs with those  $^1\text{H}$  nuclei of the ligand in the bound state which are at very short distance ( $r < 4 - 5 \text{ \AA}$ ) from the  $^1\text{H}$  nuclei of the receptor. The saturation transfer determines a reduction in the resonance intensities of the  $^1\text{H}$  nuclei that transiently bind to the receptor ( $I_{sat}$ ) due to the negative inter-molecular NOE and the transfer of the relaxation properties of the macromolecule to the small ligand in the bound state. Hence, the difference spectrum ( $I_0 - I_{sat}$ ) shows positive difference signals allowing identification of the binding compounds, while the intensities of non-binding molecules remain the same in both spectra and the difference spectrum cancels them out (purple red signals in Figure 2.10, a and b).

The STD absolute intensities are usually expressed as fractional STD ( $(I_0 - I_{sat}) / I_0$ ), or, in other words, the STD signal intensity is expressed as a fraction of the intensity of an unsaturated reference spectrum; Also, a blank experiment must be carried out to assure the absence of direct irradiation of the ligand.

As schematic representation of the pulse sequence applied during a STD NMR experiment is shown in Figure 2.11. Obviously, key elements is the selective radiofrequency by which the protein is saturated. Protein resonances selection could be achieved by a low power continuous wave saturating pulse (as in the presence of steady-state NOE), but it has been replaced by a train of Gaussian-shaped  $180^\circ$  selective pulses of 50 ms each (the filled half oval in Figure 2.11) to avoid undesirable spillover of the saturating field. The power is set to cover the bandwidth of about 100 Hz, while the number of pulses in the train ( $n$ ) is adjusted to obtain the desired length of total saturation time. Then, a hard  $90^\circ$  pulse (the filled bar with phase  $\Phi 4$ ) is applied, followed by a spin lock purge (usually between 10 and 40 ms), to remove the broad protein signals. Any other unwanted residual magnetisation can be cancelled by using trim pulses with a length of 2.5 and 5 ms (the empty bars with phase  $\Phi 1$  and  $\Phi 2$ , respectively), performed at a power level for spin-lock. The use of a gradient in correspondence of the trim pulses can help this process.<sup>80</sup>

The usual saturation time for STD NMR in screening experiments is 2-3 s, as a compromise between sensitivity and reasonable experimental time.



**Figure 2.11:** Pulse sequence for STD NMR. The narrow and wide-open bars correspond to trim pulses with a length of 2.5 and 5 ms, respectively, performed at the power level for spin-lock. The narrow-filled bar is a 90° hard pulse. A series of selective pulses (50 ms duration, 4 ms inter-pulse delay, frequency switched for on-resonance and off-resonance spectrum) is applied for saturation. The first two pulses as well as the gradient are applied to destroy residual magnetisation. Phase cycling:  $\Phi_1=x$ ;  $\Phi_2=y$ ,  $\Phi_3=x$ ,  $\Phi_4= \Phi_R=x$ ,  $2(-x)$ ,  $x$ ,  $y$ ,  $2(-y)$ ,  $y$ ;  $\Phi_5=4(y)$ ,  $4(-x)$ ,  $4(-y)$ ,  $4(x)$ . Source: Brand, 2005<sup>80</sup>

#### [Group epitope mapping](#)<sup>81, 82</sup>

An important application of the STD NMR is the possibility to *map* the binding epitope of the investigated ligands (Figure 2.12). Indeed, the magnetisation transfer from receptor to ligand protons in the bound state via intermolecular NOE depends on the inverse of the sixth power of their distance. Hence, in a STD NMR spectrum not all the signals of the ligand show the same amount of saturation (the shorter the protein-ligand  $^1\text{H}$ - $^1\text{H}$  distance the more intense the STD signal) and the observed differences indicate different spatial proximities between the distinct part of the ligand molecule and the protein in the bound state. Semi-quantitative structural information about the binding can be obtained following normalisation of the all measured STD intensities against the most intense signal, to which an arbitrarily value of 100% relative STD is assigned. The obtained binding epitope represents a fingerprint of the protein-ligand contacts in the bound state. It can be used to identify which segment of the ligand is in direct contact with the receptor surface and hence determine which moiety of the ligand is key for molecular recognition in the binding site.

In general, the structural information provided by the binding epitope depends on 1) differences in  $R_1$  relaxation rates of ligand protons, 2) the extent of saturation received in the first place, and 3) the kinetics of binding, while protons different ability to accumulate saturation in the free state is a source of distortions.

Binding epitopes were commonly obtained for a given saturation time. However, significantly different  $R_1$  relaxation rates of the ligand protons can produce artefacts in the obtained epitope. Indeed, increasing the saturation time the absolute STD intensity of any given proton of the ligand would generally increase, until a plateau is reached ( $STD_{max}$ ), usually at about 5-6 s, depending on the longitudinal relaxation time ( $R_1$ ) of the protons for the ligand in the free state. Hence, the closer we get to the  $STD_{max}$ , the more  $R_1$  will affect the recorded STD intensities. Consequence of this is that, for protons with slower  $R_1$  relaxation, saturation is accumulated more efficiently in solution, hence their relative STD intensity can be enhanced in comparison with faster relaxing protons and their proximity to the protein surface overrated at long saturation times.

To cancel this artefact, Mayer and James<sup>82</sup> proposed to derive STD intensities for a saturation times tending to zero, when virtually no accumulation of saturated ligand takes place. To do so, STD intensities are measured at increasing saturation times, as opposed as to a single saturation time experiment, and the experimental build-up curve is fit to the mono-exponential function:

$$STD(t_{sat}) = STD_{max}(1 - e^{-k_{sat} \cdot t_{sat}}) \quad \text{Equation 2.33}$$

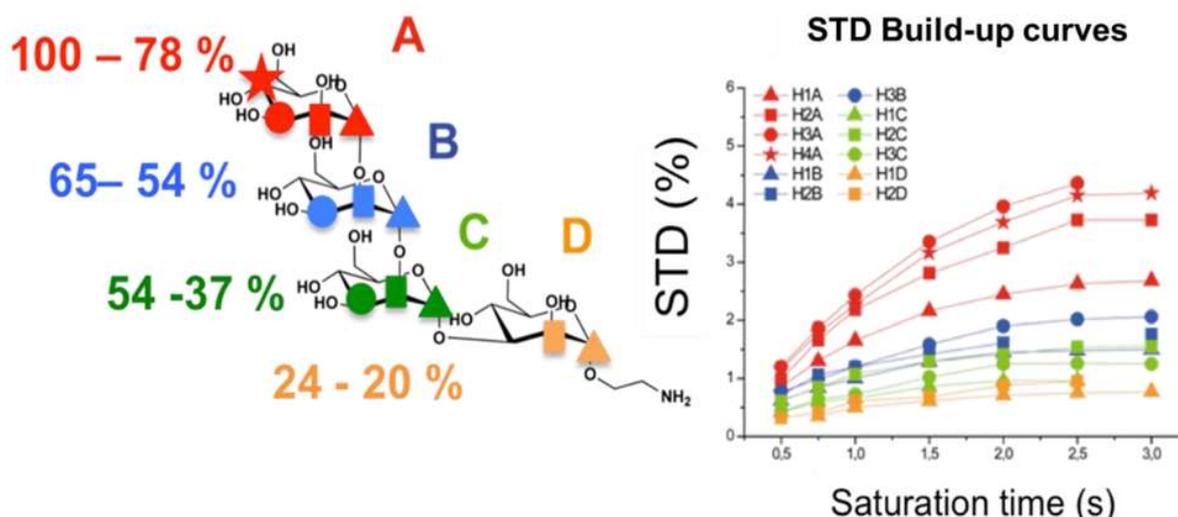
where  $STD_{t_{sat}}$  is the observed STD intensity at a given saturation time ( $t_{sat}$ ),  $STD_{max}$  is the asymptotic maximum of the build-up curve, and  $k_{sat}$  is a rate constant that measures the speed of STD build-up.

$k_{sat}$  and  $STD_{max}$  are derived by least-squares fitting, and the initial slope of the curve ( $STD_0$ ) is obtained as:

$$STD_0 = \left. \frac{dSTD(t_{sat})}{dt_{sat}} \right|_{t_{sat} \rightarrow 0} = STD_{max} \cdot k_{sat} \quad \text{Equation 2.34}$$

This approach is usually referred to as initial slopes of the build-up curves (Figure 2.12, right). In addition, possible artefacts coming from intra-molecular spin diffusion (bound state) can also be

minimised with this approximation. Thus, the group epitope mapping of a ligand is composed by the relative values of the ligand protons initial slopes, obtained by normalisation of the initial slopes values against the arbitrary chosen proton of the ligand with highest initial slope.



**Figure 2.12:** Left: group epitope mapping of a tetramannoside binding to the human anti-HIV-1 antibody 2G12 obtained by normalisation of the STD initial slopes from the experimental curves; Right: STD growth curves (absolute values) for the ligand recognition. In this example, proton H4A receives the highest saturation (thus the 100% is arbitrarily assigned) and is used as reference to calculate the STD percentages of the other protons. The distribution of saturation transfer indicates binding through the non-reducing mannose residue (mannose A).  
Source: Angulo *et al.*, 2010<sup>79</sup>

In addition, the use of large molar excess of the ligand precludes the perturbation of the absolute STD intensities due to rebinding effects (i.e., a ligand already saturated experiences another association process, without previous full relaxation), which would impede to correctly determine the group epitope mapping.

#### Differential Epitope Mapping (DEEP) STD NMR

Recently, Monaco *et al.*,<sup>83</sup> developed the Differential Epitope Mapping (DEEP) STD NMR methodology able to elucidate the nature of the protein residues around the ligand in the bound state, from which the ligand orientation inside the binding pocket can be inferred.

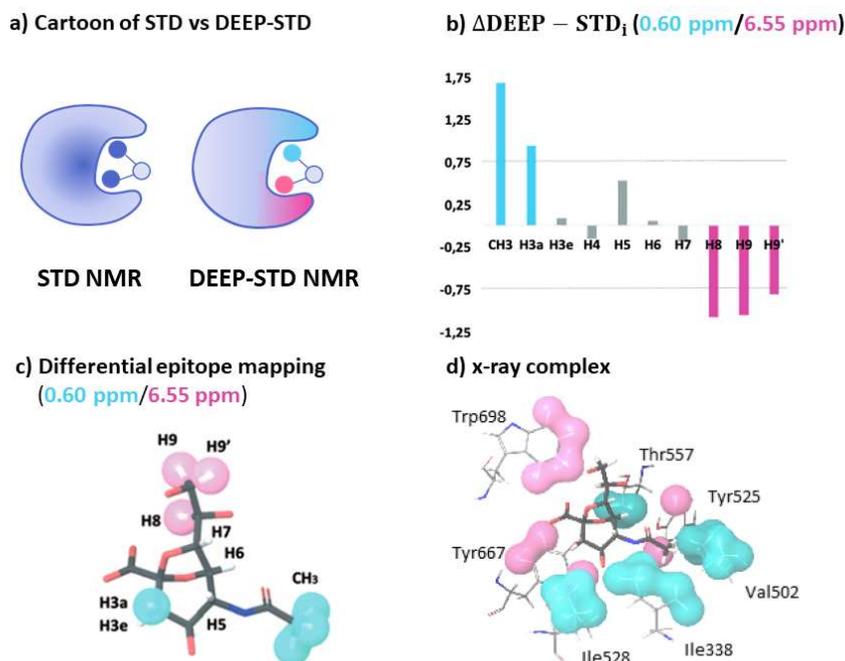
In short, the binding epitope of a ligand under two different experimental conditions (i.e., two different selective irradiation frequencies, or two different solvents, D<sub>2</sub>O and H<sub>2</sub>O) is recorded and their differences in relative STDs quantified. We will refer to the experiments performed under differential conditions as experiment-1 (exp1) and experiment-2 (exp2). Once the experiments have been recorded, it is necessary to compare the total sum of the STD values for each experiment, to determine which experimental condition produces stronger STD intensities. Hence, the experiment with largest total sum will be assigned to exp1. Next, the ratio of STD intensities ( $\frac{STD_{\%exp1,i}}{STD_{\%exp2,i}}$ ) is calculated for each proton of the ligand. The obtained values contain the contribution from differences in the global level of protein saturation, which must be removed. Hence, the average ratio of STDs over all protons, *n*, is calculated  $\left[ \frac{1}{n} \sum_i^n \left( \frac{STD_{exp1,i}}{STD_{exp2,i}} \right) \right]$  and the obtained average factor subtracted from each of the individual ratios.

The DEEP-STD factor ( $\Delta DEEP - STD_i$ ) for each proton *i* of the ligand is obtained as:

$$\Delta DEEP - STD_i = \frac{STD_{exp1,i}}{STD_{exp2,i}} - \frac{1}{n} \sum_i^n \left( \frac{STD_{exp1,i}}{STD_{exp2,i}} \right) \quad \text{Equation 2.35}$$

An increase of the relative STD values for the ligand protons will then be translated into positive values for increases in exp1, negative values for increases in exp2 (Figure 2.13). Importantly, we refer to “relative” STD values as they derive from the comparison of the binding epitopes recorded under the differential experimental conditions, and not from absolute STD values.

When the structural coordinates (PDB) of the protein are available and the binding site known, chemical shift of the amino acids resonances constituting the catalytic cleft can be obtained with SHIFTX2, a protein chemical shift prediction programs (<http://www.shiftx2.ca/>).<sup>84</sup> Hence, we can select which types of protein protons will be directly irradiated (aliphatic, aromatic, polar, etc), and the differences in binding epitopes will highlight those parts of the ligand contacting a given type of irradiated protein residues (Figure 2.13, d).



**Figure 2.13:** a) Schematic representation of the different concepts behind conventional STD and DEEP STD NMR. b)  $\Delta$ DEEP-STD histogram for the Differential Epitope Mapping (0.60 ppm/6.55 ppm) of 2,7-anhydro-Neu5Ac in complex with GH33, the glycosyl hydrolase domain of the intramolecular (IT) *trans*-sialidase from *Ruminococcus gnavus* (IT-sialidase from *RgNanH*, PDB ID: 4X4A). Positive  $\Delta$ DEEP-STD (above the limit of +0.75) after aliphatic irradiation (0.60 ppm) are in cyan, and negative  $\Delta$ DEEP-STD (below -0.75) after aromatic irradiation (6.55 ppm) are in magenta. c) Differential Epitope:  $\Delta$ DEEP-STD map of the ligand. Cyan surfaces highlights ligand contacts with aliphatic side chains; magenta, contacts with aromatic side chains. The ligand polar protons have been omitted. d) Crystal structure of the complex (PDB ID: 4X4A).  
 Source: Figure adapted from Monaco *et al.*, 2017<sup>83</sup>

Another source of differences in epitope maps is the solvent. In D<sub>2</sub>O, amino acids with polar side chains in the binding pocket will inefficiently transfer the saturation to the ligand as their exchangeable protons will be replaced by deuterium. In H<sub>2</sub>O, on the contrary, those protons will contribute to an additional transfer of saturation to ligand protons, with a major contribution expected for those protons which are in slow exchange with the bulk water. Therefore, regions of

the bound ligand in close contact with clusters of slow exchanging residues will present an increase in the relative STD intensities when the STD NMR experiment is performed in water.

In addition, Nepravishta *et al.*,<sup>85</sup> expanded the use of DEEP-STD NMR to situations when no spectral assignment of the protein is available. They proposed the use of TEMPOL-based 2D <sup>1</sup>H-<sup>1</sup>H TOCSY experiments to identify the binding pocket residues chemical shifts, and then perform DEEP-STD NMR experiments using the frequencies identified. In addition, they proposed to calculate an average DEEP-STD map by averaging all the obtained DEEP-STD factors (instead of using a single pair of frequencies), providing a more accurate depiction of the orientation and the nature of the amino acids surrounding the ligand in the binding pocket.

Dissociation constant ( $K_D$ ) measurement<sup>86</sup>

Furthermore, the thermodynamics protein-ligand dissociation constant  $K_D$  can be derived from STD NMR experiments performed at increasing total ligand concentrations. Importantly, the observed experimental STD intensities ( $(I_0 - I_{\text{sat}})/I_0$ ) are a function of the fraction of the fraction of bound ligand  $f_{LB}$

$$f_{LB} = \frac{[PL]}{[L]_T} = \frac{[PL]}{[L] + [PL]} = \frac{[P]}{K_D + [P]} \quad \text{Equation 2.36}$$

To obtain the fraction of bound protein  $f_{PB}$  necessary for the derivation of the Langmuir binding isotherm (Equation 2.38), Mayer and Meyer proposed a correction for total ligand concentration multiplying the observed STD by the molar excess of the ligand over protein to obtain the STD amplification factors (STD-AF – Equation 2.37)<sup>81</sup>, which depends on the fraction of bound protein (Equation 2.38)

$$STD - AF = STD \cdot \frac{[L]_T}{[P]_T} \quad \text{Equation 2.37}$$

$$f_{PB} = \frac{[PL]}{[P]_T} = \frac{[PL]}{[P] + [PL]} = \frac{[L]}{K_D + [L]} \quad \text{Equation 2.38}$$

Plotting the STD-AF values at increasing ligand concentration yields the protein-ligand binding isotherm, from which the dissociation constant  $K_D$  can be derived. In Equation 2.38 the concentration of the free ligand  $[L]$  is not known. However, in conditions such as very low total concentration of protein ( $[P]_T \ll K_D$ ) and an excess of ligand,  $[L] \approx [L]_T$ , so that

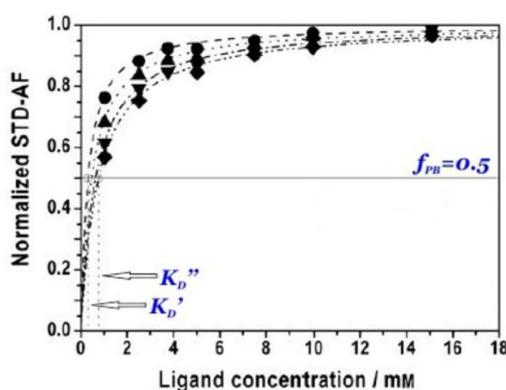
$$f_{PB} = \frac{[L]_T}{K_D + [L]_T} \quad \text{Equation 2.39}$$

Plotting STD-AF values at increasing ligand concentration  $[L]$  gives rise to a Langmuir hyperbolic dose-response curve

$$STD - AF ([L]) = \frac{\alpha_{STD}[L]}{K_D + [L]} \quad \text{Equation 2.40}$$

where  $\alpha_{STD}$  is a dimensionless scaling factor which represent the maximum STD amplification for the monitored signal. Hence, the  $K_D$  can be derived by mathematical fit of the experimental data to Equation 2.40.

The experimental parameters affecting  $K_D$  determination via STD NMR, such as the used saturation time ( $t_{sat}$ ) (Figure 2.14), the intensity of the STD signals and the fraction of bound ligand, have been thoroughly studied by Angulo *et al.*<sup>79</sup>. Importantly, they proposed to use the initial growth rates of the STD amplification factors (STD-AF<sub>0</sub>) rather than the STD-AF factors at a given saturation time, for a more accurate measurement of the  $K_D$ . This approach removes fast protein–ligand rebinding processes effects, main cause of errors in  $K_D$  determination by STD NMR spectroscopic titration experiments.



**Figure 2.14:** Example of STD binding isotherms, at different saturation times, of the WGA protein (46  $\mu\text{M}$ ) titrated with chitobiose. STD-AF values appear normalised against their corresponding plateau values. Note that different  $K_D$  values ( $K_D'$ ,  $K_D''$ ) are obtained at different saturation times. *Source:* Angulo, 2010<sup>79</sup>

## 2.2 Solid-state NMR (ssNMR)<sup>87</sup>

As previously mentioned, ssNMR is a powerful tool to obtain structural and dynamics information of macromolecules and it is commonly employed to study solid and solid-like materials.

$^1\text{H}$  detection in ssNMR is typically dominated by line broadening due to (i) strong homonuclear dipole interactions, and (ii) chemical shift anisotropy (CSA), i.e. the angular dependence of the chemical shift. CSA is eliminated via the so-called Magic Angle Spinning (MAS),<sup>88, 89</sup> while the homonuclear dipole contribution is completely averaged out only at fast spinning (ultra-fast MAS usually required, i.e. above 100 kHz).

Solid-state NMR allows the direct detection of  $^{13}\text{C}$  nuclei with high sensitivity via the cross-polarization (CP) experiment and the use of high power  $^1\text{H}$  decoupling. CP relies on the transfer of magnetisation from an abundant ( $I$ ) to a diluted ( $S$ ) spin by dipolar couplings. A thorough understanding of the parameters affecting the CP transfer is key to record of high-quality spectra, for quantitative studies and to derive molecular mobility information.<sup>87</sup>

### 2.2.1 Chemical shielding and Chemical Shift Anisotropy

The electrons circulating in the orbitals around the nuclei act themselves as magnets (they have a magnetic moment and, with their movements, they create a circulating current), resulting in the induced magnetic field ( $B'$ ). This local magnetic field modulates the effective magnetic field ( $B$ ) felt by the nuclei, which is slightly smaller than the applied external field ( $B_0$ ) – thus, we say that the electrons have a “shielding” effect

$$B = B_0 - B' = B_0(1 - \sigma) \quad \text{Equation 2.41}$$

where  $\sigma$  is the shielding constant. This shielding interaction is described by the chemical shielding Hamiltonian ( $\hat{H}_{CS}$ )

$$\hat{H}_{CS} = -\gamma \hat{I}_z B_0 (1 - \sigma) \quad \text{Equation 2.42}$$

The expression of the chemical shift ( $\delta$ ) values as part per million (ppm) of the external field derives from the fact that the local magnetic fields produced by moving electrons are considerably smaller than  $B_0$ .<sup>90</sup> Normally, the chemical shift is determined as the difference between the frequencies of the analysed nucleus ( $\nu$ ) with respect to a reference compound ( $\nu_{ref}$ ) of known frequency, to which a 0 ppm chemical shift is assigned by convention

$$\delta_{ppm} = 10^6 \left( \frac{\nu - \nu_{ref}}{\nu_{ref}} \right) \quad \text{Equation 2.43}$$

Different standard reference compounds are used for different nuclei (depending on availability and ease to handle), e.g. adamantane and tetramethylsilane (TMS) are standards for  $^1\text{H}$  and  $^{13}\text{C}$ ,  $\text{CFCl}_3$  for  $^{19}\text{F}$  and liquid  $\text{NH}_3$  for  $^{15}\text{N}$  chemical shift referencing.

Different atoms present different electron clouds densities, depending on presence of hydrogen bonds, presence of electronegative nuclei in close proximity, unpaired electrons or through space interactions within molecular clusters and hybridization state. In addition, those electrons clouds do not present a spherical symmetric distribution around the nuclear spins but rather an elongated shape (ellipsoid) along bonds or non-bonding  $p$ -orbitals (Figure 2.15). The orientation of this ellipsoid with respect to the external field (and hence the orientation of the molecule) with respect

to  $B_0$  determines the degree to which the electron density affects the size of the induced magnetic field and hence the resonance frequency. The nucleus resonance will therefore present a deviation, defined as chemical shift anisotropy (CSA). For  $^{13}\text{C}$  nucleus resonance frequency,  $\delta_{11}$  and  $\delta_{33}$  correspond to the strongest and the weakest shielding effects, respectively. The first occurs when the narrowest part of the electron cloud is oriented along  $B_0$  (Figure 2.15), while the second when the widest part is oriented along the z-axis (Figure 2.15).  $\delta_{22}$ , instead, is the shift produced by the molecular orientation perpendicular to the axes of  $\delta_{11}$  and  $\delta_{33}$  (Figure 2.15).<sup>90</sup> Importantly, for the special case when  $\delta_{11} = \delta_{22}$  (a shielding tensor with axial symmetry), the chemical shift anisotropy (CSA) Hamiltonian ( $\hat{H}_{CS}$ ) can be written as

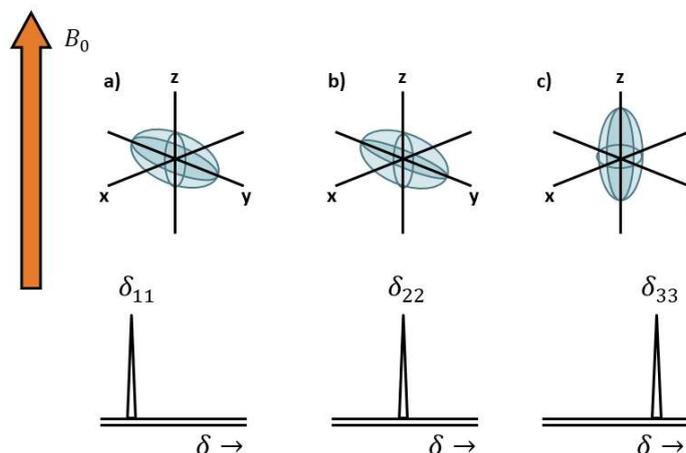
$$\hat{H}_{CSA} = \gamma B_0 I_z \left[ \delta_{ISO} + \frac{1}{2} \delta_{CSA} (3 \cos^2 \theta - 1) \right] \quad \text{Equation 2.44}$$

Where the  $\delta_{ISO}$  is the isotropic chemical shielding factor

$$\delta_{ISO} = \frac{1}{3} (\delta_{11} + \delta_{22} + \delta_{33}) \quad \text{Equation 2.45}$$

And  $\delta_{CSA}$  described the chemical shift anisotropy (CSA)

$$\delta_{CSA} = \delta_{ISO} - \delta_{33} \quad \text{Equation 2.46}$$



**Figure 2.15:** Representation of preferential orientation of the ellipsoidal electron cloud of a carbonyl with respect to the external magnetic field, with the corresponding values of chemical shift tensors, with the corresponding values of chemical shift tensors, a)  $\delta_{11}$ , b)  $\delta_{22}$  and c)  $\delta_{33}$ .

Source: Laws *et al.*, 2002.<sup>90</sup>

## 2.2.2 Homonuclear Dipolar Coupling

The homonuclear dipolar coupling is the interaction between the magnetic moments of like spins, and it is described by the homonuclear dipolar coupling Hamiltonian for a two-spin system

$$H_{I,I} = -\left(\frac{\mu_0}{4\pi}\right) \frac{\hbar\gamma_1\gamma_2}{r_{1_1 1_2}^3} \frac{1}{2} (3 \cos^2 \theta_{1_1 1_2} - 1) \left[ 2I_{1z}I_{2z} - \frac{1}{2}(I_1^+ I_2^- + I_1^- I_2^+) \right] \quad \text{Equation 2.47}$$

where  $\gamma_1\gamma_2$  are the gyromagnetic ratio of spin 1 and spin 2, respectively,  $r_{1_1}r_{1_2}$  is the internuclear distance between two like spins,  $\mu_0$  is the permeability of free space ( $\mu_0 = 4\pi \times 10^7 \text{ NA}^{-2}$ ) and  $\theta$  is the angle that describes the orientation of the internuclear vector with respect to the external magnetic field. The terms  $I_1^+ I_2^-$  and  $I_1^- I_2^+$  express the spin diffusion mechanism of magnetisation transfer, hence the energy conserving “flip-flop” transition of two like spins with similar resonance frequencies (Figure 2.16). The raising operator ( $I^+ = I_x + iI_y$ ) represents the flip of a spin from “down” to “up” orientation, while the lowering operator ( $I^- = I_x - iI_y$ ) expresses the flip of a spin from “up” to “down” orientation. By rearranging the term

$$I_1 I_2 = I_{1x} I_{2x} + I_{1y} I_{2y} + I_{1z} I_{2z} \quad \text{Equation 2.48}$$

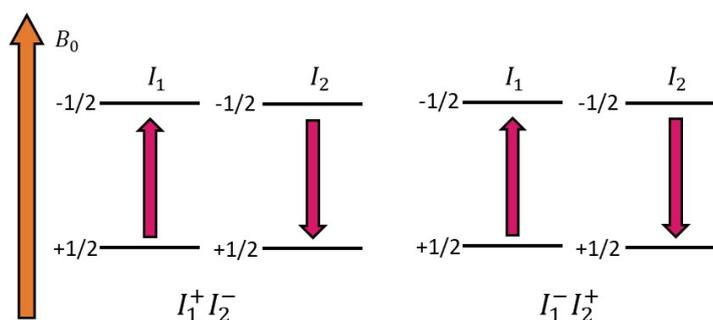
The homonuclear dipolar coupling Hamiltonian can be simplified to

$$H_{I,I} = -\left(\frac{\mu_0}{4\pi}\right) \frac{\hbar\gamma_1\gamma_2}{r_{I_1 I_2}^3} \frac{1}{2} (3 \cos^2 \theta_{I_1 I_2} - 1) [3I_{1z} I_{2z} - (I_1 I_2)] \quad \text{Equation 2.49}$$

The first left-term of the equation 47 describes to the dipolar constant (d), hence expressed as

$$d = \left(\frac{\mu_0}{4\pi}\right) \frac{\hbar\gamma_1\gamma_2}{r_{I_1 I_2}^3} \quad \text{Equation 2.50}$$

From equation 2.50 it is evident the direct proportionality between the dipolar interactions and the gyromagnetic ratio of interacting spins, as well as the inverse proportionality with the cube of the internuclear distance. Nuclei with larger magnetic moments produce stronger local magnetic fields, which *in turn* increases the dipolar coupling interaction. Therefore, for abundant nuclei with high gyromagnetic ratio, dipolar interactions are extremely strong ( $^1\text{H}$ - $^1\text{H}$  in the range of 100 kHz). On the contrary, pairs of nuclei with low gyromagnetic ratio and low natural abundance present negligible homonuclear dipolar coupling ( $^{13}\text{C}$ - $^{13}\text{C}$  do not exceed 5 kHz).



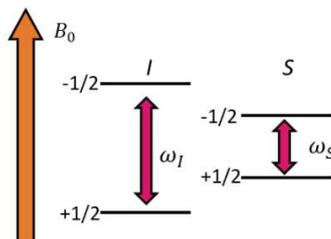
**Figure 2.16:** Energy conserving “flip-flop” interaction between two like spins occurs since  $\omega_{I_1} = \omega_{I_2}$

### 2.2.3 Heteronuclear Dipolar Coupling

The heteronuclear dipolar coupling is the interaction between the magnetic moments of unlike spins. The heteronuclear dipolar coupling Hamiltonian ( $\hat{H}_{I,S}$ ) for a two-spin system is described by

$$H_{I,S} = -\left(\frac{\mu_0}{4\pi}\right) \frac{\hbar\gamma_1\gamma_2}{r_{IS}^3} \frac{1}{2} (3 \cos^2 \theta_{IS} - 1) I_Z S_Z \quad \text{Equation 2.51}$$

where nuclear spins are labelled as  $I$  for abundant spins ( $^1\text{H}$ ,  $^{19}\text{F}$ ) and  $S$  for rare spins ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ) and  $r_{IS}$  represents the internuclear distance between the two unlike spins. The spin component ( $I_Z S_Z$ ) is significantly simplified, because the energy conserving “flip-flop” transition is not possible between two spins with different resonance frequencies (Figure 2.17). Similarly to homonuclear dipolar couplings, the strength of heteronuclear dipolar interactions depends on the gyromagnetic ratio of interacting spins, the distance between both nuclei and their orientation with respect to the external magnetic field. In specific,  $^{13}\text{C}$ - $^1\text{H}$  heteronuclear dipolar coupling has a typical strength between 20 and 30 kHz (at a distance of about 1 Å).

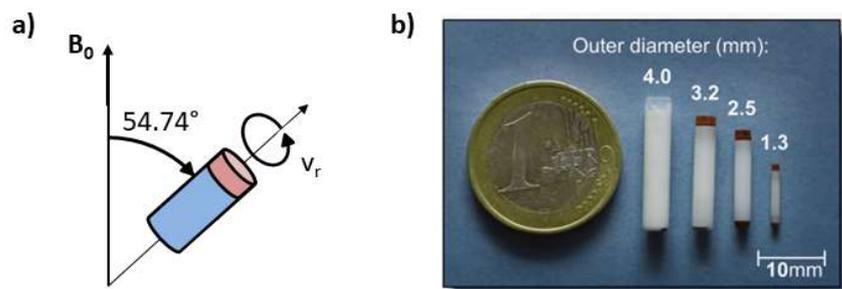


**Figure 2.17:** Energy conserving “flip-flop” transitions cannot occur between two unlike spins since  $\omega_I \gg \omega_S$

### 2.2.4 Magic Angle Spinning

The Magic Angle Spinning approach was at first proposed by Andrew *et al.* (1958)<sup>88</sup> for a single crystal of sodium chloride and revolutionised the field of ssNMR. In summary, all anisotropic interactions present orientational dependence via the term  $(3 \cos^2 \theta - 1)$ . In powder samples,  $\theta$  can assume all possible values, but when the sample is spun at an axis oriented at the angle  $\theta = 54.74^\circ$  with respect to the external magnetic field, the angular dependence in the geometric

component of dipolar and CSA interaction is cancelled to zero (Figure 2.18, a). This technique mimics the sampling of all possible orientation for rapid tumbling molecules in solution, which therefore present isotropic chemical shifts.



**Figure 2.18:** a) Representation of a rotor spun at the MAS ( $\theta = 54.74^\circ$ ) and b) Variety of NMR rotor sizes available to reach different MAS rates. *Source:* Adapted from Demers *et al.*, 2011<sup>91</sup>

In the application of the MAS to record  $^1\text{H}$  spectrum, it is possible to differentiate dynamics regimes in solids as mobile domains will appear in the spectra as sharp peak.<sup>92</sup> In addition, a spectrum obtained under  $^1\text{H}$ - $^{13}\text{C}$  Cross-Polarisation/MAS conditions (explained in subsection 2.2.8) contains *isotropic* chemical shift information in analogy to liquid-state  $^{13}\text{C}$  NMR with proton decoupling.

Nonetheless, to effectively eliminate the geometric term of the spin-spin interaction described by the Hamiltonians, MAS rates at least three times higher than the strength of the interactions are required. For example, MAS rate of 10-15 kHz effectively eliminates  $^1\text{H}$ - $^{13}\text{C}$  heteronuclear dipolar coupling. On the contrary, the much stronger homonuclear dipolar coupling (in the range of *ca.* 80 kHz) are frequently observed in solid-state NMR for nuclei with high abundance and high gyromagnetic ratio ( $^1\text{H}$ ,  $^{19}\text{F}$ ). Even though application of ultra-fast MAS rates of 120 kHz is nowadays possible with currently developed solid-state NMR probes, the homonuclear dipolar interactions cannot be fully averaged and are still present in the spectrum.

#### 2.2.4.1 Spinning sidebands

Common MAS artefacts are the so-called spinning sidebands, which arise when the spinning rate is not sufficient to average out the anisotropy. The spinning sidebands appear on both side of the isotropic chemical shift resonance and are separated by frequency distances of integer multiples of the spinning speed. In order to effectively eliminate the geometric term of the anisotropic

interactions, it is necessary to use MAS rates 3-4 times higher than the interaction strength.  $^1\text{H}$ - $^{13}\text{C}$  dipolar coupling is averaged at MAS rate of 10-15 kHz, but the stronger  $^1\text{H}$ - $^1\text{H}$  dipolar coupling (in the range of 80 kHz) are frequently observed.

### 2.2.5 Heteronuclear decoupling

When observing a dilute spin  $S$  ( $^{13}\text{C}$ ,  $^{29}\text{Si}$ ,  $^{15}\text{N}$ ), with a pool of abundant spin  $I$  nearby ( $^1\text{H}$ ), heteronuclear dipolar coupling result in a broadening of the lines. The effect of this coupling can be removed by applying high power  $rf$  irradiation at the resonance frequency of the  $I$  spin. This technique is known as high-power decoupling (HP).

To understand how to reduce the heteronuclear dipolar coupling contribution, we need to look once more at the heteronuclear dipolar coupling Hamiltonian (Equation 2.51), from which it is evident that the heteronuclear dipolar interactions depends also on the spin component ( $I_Z S_Z$ ), besides the gyromagnetic ratio of interacting spins, the distance between both nuclei and their orientation to the external magnetic field previously described. Briefly, the effect of this  $rf$  irradiation is to cause a repeated transition of the  $I$  spin between high and low energy states ( $\alpha \leftrightarrow \beta$ ) when the free-induction decay (FID) curve for the  $S$ -spin is acquired. Consequently, the term  $I_Z S_Z$  will oscillate between positive and negative values, producing a time averaged value of zero.

Nonetheless, application of decoupling  $rf$  pulses during long acquisition times can cause sample and probehead overheating. Molecular motions with a time scale similar to the  $rf$  irradiation can interfere with the decoupling field, resulting in an ineffective decoupling and consequent broadening of the signals in the spectrum. In addition, if the decoupling and the spinning rate are of the same order, the two processes will also interfere, resulting again in a broadening effect. For these reasons, lower power decoupling with the application of short pulses at define intervals are nowadays preferred. Common examples are TPPM (Two Phase Pulse Modulated)<sup>87, 90, 93, 94</sup>, XiX (X inverse X)<sup>95</sup> and SPINAL<sup>96</sup> pulse sequences, which present better performance than HP decoupling for similar  $rf$  power and can be used at higher spinning speed.

### 2.2.6 Homonuclear decoupling

In the same way as in heteronuclear decoupling, the homonuclear dipolar coupling can be eliminated by annulling the effect of the homonuclear dipolar Hamiltonian on the nuclear magnetisation. This can be achieved by the application of multiple pulse sequences (e.g. WAHUHA or MREV-8).<sup>87, 90</sup> Nonetheless, these decoupling sequences require the FID to be collected at particular points between cycles of pulses, as well as the application of high power pulses in order to maintain the pulse length short relative to the cycle time. Moreover, the spectral width of the resulting spectrum is the inverse of the cycle time and it must be large enough to contain all the signals; as a result, short cycles times are preferable. In addition, the decoupling sequence and the sample spinning should be synchronised as MAS can interfere with the decoupling sequence.

To overcome these problems, the homonuclear decoupling is often implemented in the indirect dimension of two dimensional (2D) experiments, which resulted in the generation of high resolution  $^1\text{H}$  spectra from 2D experiments.<sup>97</sup> Common pulse sequences are Lee-Goldburg (LG) and frequency switched Lee-Goldburg schemes (FSLG).<sup>87, 90</sup> The main advantage of this approach over the implementation of one dimensional (1D) homonuclear decoupling is that the homonuclear decoupling does not need to be rotor synchronised.

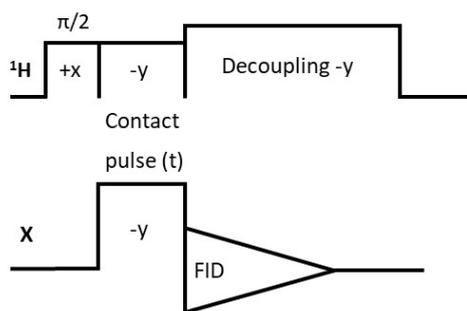
### 2.2.7 Single pulse excitation $^{13}\text{C}$ NMR

Single pulse (SP)  $^{13}\text{C}$  NMR experiments rely on the direct  $^{13}\text{C}$  excitation and subsequent spectra acquisition under conditions of high power  $^1\text{H}$ - $^{13}\text{C}$  heteronuclear decoupling. This experiment is called Single Pulse (SP) as only a  $90^\circ$  pulse on  $^{13}\text{C}$  is applied. The general pitfall of the experiment is the long relaxation delay required due to the typically long  $T_1$  times of  $^{13}\text{C}$  (tens – hundreds of seconds), leading to very long experimental times.<sup>98</sup> The long  $T_1$  times of  $^{13}\text{C}$  are due to the low efficiency of the weaker  $^1\text{H}$ - $^{13}\text{C}$  heteronuclear dipolar couplings in triggering relaxation, and the absence of the stronger  $^{13}\text{C}$ - $^{13}\text{C}$  homonuclear dipolar couplings due to the low abundance of the  $^{13}\text{C}$  isotope. On the contrary, when other nuclei (such as  $^{19}\text{F}$ ) are acquired, the more efficient homonuclear coupling allows the use of shorter relaxation delays.

### 2.2.8 $^1\text{H}$ - $^{13}\text{C}$ CP/MAS NMR<sup>87</sup>

Cross-polarization (CP) constitutes the core element in solid-state NMR pulse sequences. CP is a double-resonance solid-state NMR experiment that relies on the magnetisation transfer from abundant  $I$  nuclei (usually  $^1\text{H}$ ) to dilute  $S$  nuclei (i.e.  $^{13}\text{C}$ ,  $^{29}\text{Si}$  and  $^{15}\text{N}$ , with isotopic natural abundance of 1.1%, 4.7% and 0.03%, respectively) by applying simultaneous spin-lock pulses to both  $I$  and  $S$  nuclei (the  $S$  nuclei is often generally referred to as  $X$ ). The CP experiment is commonly performed under MAS and high-power  $I$  spin decoupling conditions, enabling the detection of well resolved  $S$  peaks. The methodology overcomes two common problems in the NMR of solids: (i) the low sensitivity for low abundance NMR-responsive isotopes with low gyromagnetic ratio (e.g.  $^{13}\text{C}$  isotopic natural abundance is 1.1% and its gyromagnetic ratio is  $\frac{1}{4}$  of  $^1\text{H}$ ); and (ii) the long  $T_1$  relaxation time of dilute spin  $\frac{1}{2}$  nuclei (tens of seconds for  $^{13}\text{C}$  in powdered organics, minutes for  $^{29}\text{Si}$  in framework silicates).

During CP, magnetisation transfer is mediated by the heteronuclear dipolar coupling between  $^1\text{H}$  and  $X$ ; hence, the kinetics of CP transfer is highly influenced by internuclear  $^1\text{H}$  -  $X$  distances and local mobility. Thus,  $^1\text{H}$  -  $X$  pairs in rigid and mobile structures will show a fast and slow CP growth, respectively. On the other hand,  $^1\text{H}$   $T_{1\rho}$  relaxation competes with CP transfer, eventually leading to equilibrium magnetisation.  $X$  nuclei shows long longitudinal relaxation times due to the absence of strong homonuclear dipolar interactions to stimulate relaxation (only the weaker heteronuclear dipolar interactions are present). Nonetheless, in CP experiment, the relaxation delay is determined by the  $^1\text{H}$   $T_{1\rho}$  relaxation time, which is in general much shorter than the  $X$  spin relaxation time. This allows the faster acquisition of spectra with a good signal-to-noise ratio. The CP pulse is shown in Figure 2.19:



**Figure 2.19:** CP pulse sequence representation

To understand the physical principles of CP, a double rotating frame of reference about  $\vec{B}_0$  is usually considered. In this frame of reference, the magnetic fields arising from the  $^1\text{H}$  and X spin-lock pulses appear static. In the first part of the pulse, the  $90^\circ$  pulse applied on  $^1\text{H}$  rotates the magnetisation along the  $-y$  axis in the  $^1\text{H}$  rotating frame. Then, two concomitant pulses are applied on  $^1\text{H}$  and X along the  $-y$  axis (contact pulse). The application of these pulses generates two different fields,  $B_1(^1\text{H})$  and  $B_1(\text{X})$ , or spin-lock fields, which act as quantization axis for  $^1\text{H}$  and X energy states. Hence, the  $^1\text{H}$  and X nuclei will present parallel and antiparallel components to the spin-lock fields, respectively, called  $\alpha^*$  and  $\beta^*$  for  $^1\text{H}$  and  $\alpha^*x$  and  $\beta^*x$  for X.

$^1\text{H}$  and X nuclei must be dipolar coupled for CP to occur, and the Hamiltonian operator describing this interaction is expressed as:

$$\hat{H}_{HX} = - \sum d_i (3 \cos^2 \theta_i - 1) \hat{I}_{iz}^H \hat{S}_z^X \quad \text{Equation 2.42}$$

where  $d_i$  is the dipolar coupling constant. As both  $^1\text{H}$  and X spins are quantized in the x-y plane of the rotating frame, the dipole-dipole coupling operator  $\hat{I}_{iz}^H \hat{S}_z^X$ , which acts in a direction perpendicular to the quantization axes, does not affect the net energy of the spin system or alter the net spin polarization (given by the sum of  $^1\text{H}$  and X spin polarizations) parallel to the transverse plane. For CP to occur, the magnitude of the contact pulses applied on  $^1\text{H}$  and X must meet the Hartmann-Hahn (HH) matching condition:

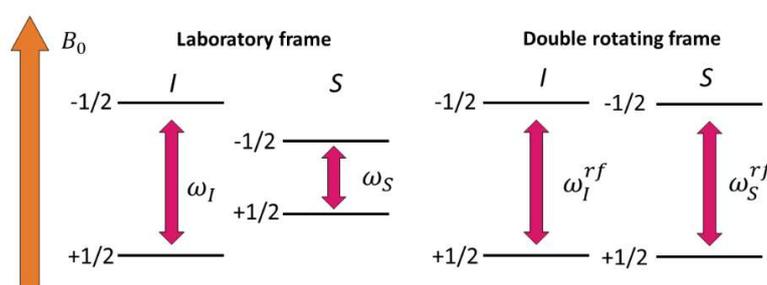
$$\gamma_H B_1(H) = \gamma_X B_1(X) \quad \text{HH matching at the centerband, static case} \quad \text{Equation 2.43}$$

$$\gamma_H B_1(H) = \gamma_X B_1(X) + \omega_{MAS} \text{ HH matching at the 1}^{\text{st}} \text{ order sidebands,}$$

Equation 2.44

MAS case

where  $\omega_{MAS}$  is the magic angle spinning speed. Hence, the CP process is an energy conserving process, where the energy gaps between the I and S spins in their respective rotating frame becomes equal (Figure 2.20). The dipolar coupling between the  $^1\text{H}$  and X spins allows redistribution of energies between them, while the total energy and the net magnetisation are maintained constant.



**Figure 2.20:** Equalisation of the energy differences between the transitions of the I and the S spins in a double rotating frame when the nutation frequencies on both spins are equal ( $\omega_I = \omega_S$ )

For  $^1\text{H}$ , the initial  $90^\circ_x$  pulse rotate the magnetization in the transverse plane, hence the spin population in the direction of the  $B_1(\text{H})$  field will be the same as the one along  $B_0$  in the laboratory frame. As the  $B_1(^1\text{H})$  field is much smaller than  $B_0$ , it cannot sustain the same energy difference between the  $\alpha^*$  and  $\beta^*$  spin states, leading to a reduction of the population difference ( $\alpha^* \rightarrow \beta^*$ ). At the same time,  $\beta^*x \rightarrow \alpha^*x$  transitions in the rotating frame of the X spin occur to conserve the energy, leading to a large magnetisation along  $B_1(\text{X})$

Thermodynamics offers another elegant explanation of the CP as a heat-exchange process.<sup>99</sup> When the contact pulse is applied, the large spin population difference in I rotating frame could be compared to a high spin temperature, while the zero magnetisation of spin S in the rotating frame correspond to zero spin temperature. Thus, the HH matching condition brings both spin systems in thermal contact, enabling the “heat” transfer from the hot I spin population to the cold S spin population. Hence, it is the difference in spin temperature which enables the flow of energy and the polarisation transfer to occur.

Another important phenomenon to consider during CP is homonuclear  $^1\text{H}$  spin diffusion, which leads to magnetisation redistribution within the  $^1\text{H}$  network. A more exhaustive explanation on its effect on CP will be given in the following CP kinetics section.

### 2.2.8.1 CP kinetics<sup>99</sup>

The CP kinetics methodology relies on monitoring peak intensity changes of X nuclei during CP experiments performed at variable contact times. As CP relies on heteronuclear dipolar interactions, it is sensitive to the local mobility and internuclear distances. CP kinetics is thus a powerful method to characterise local structural and dynamics between dipolar coupled nuclei in solid-like materials and provides crucial information to correctly interpret single CP spectra. Indeed, different peaks can present maximum intensity at distinct contact times, hindering the application of single CP spectra for quantitative analysis. On the other hand, CP kinetics allows to identify the optimum contact time for each peak and, therefore, to design quantitative CP experiments.

Over the years, two main models have been used to explain the behaviour of CP with contact time. The classical *I-S* model was constructed around homogeneous solids, thus systems presenting weak *I-S* heteronuclear and strong *I-I* homonuclear dipolar couplings. This favours the occurrence of homonuclear spin diffusion and homogenisation of the *I* magnetisation. The general CP kinetics equation for abundant and diluted spin  $\frac{1}{2}$  systems is:

$$I(t) = I_0 \left( 1 + \frac{T_{IS}}{T_{1\rho}^S} - \frac{T_{IS}}{T_{1\rho}^I} \right)^{-1} \left\{ \exp\left(-\frac{t}{T_{1\rho}^I}\right) - \exp\left[-t\left(\frac{1}{T_{IS}} + \frac{1}{T_{1\rho}^S}\right)\right] \right\} \quad \text{Equation 2.44}$$

where  $I_0$  is the absolute intensity,  $T_{IS}$  the CP time constant,  $T_{1\rho}^I$  and  $T_{1\rho}^S$  the relaxation constant in the rotating frame for *I* and *S*, respectively, and *t* is the contact time.

For slowly relaxing *S* nuclei, where  $\frac{T_{IS}}{T_{1\rho}^S} \approx 0$ , the equation simplifies to:

$$I(t) = I_0 \left( 1 - \frac{T_{IS}}{T_{1\rho}^I} \right)^{-1} \left[ \exp\left(-\frac{t}{T_{1\rho}^I}\right) - \exp\left(-\frac{t}{T_{IS}}\right) \right] \quad \text{Equation 2.45}$$

This equation describes the double exponential behaviour of CP, where the initial growth is governed by the  $\exp\left(-\frac{t}{T_{IS}}\right)$  term, and the curve exponential decrease is governed by the  $-t/T_{1\rho}^I$ .

The maximum intensity ( $I_0$ ) is decreased by the CP time constant factor  $\left(1 - \frac{T_{IS}}{T_{1\rho}^I}\right)^{-1}$ .

On the other hand, when  $T_{1\rho}^I$  relaxation is slow in comparison to  $T_{IS}$ , i.e.  $T_{IS}/T_{1\rho}^I \approx 0$ , the CP kinetics expression simplifies to:

$$I(t) = I_0 \left[ 1 - \exp\left(-\frac{t}{T_{IS}}\right) \right] \quad \text{Equation 2.46}$$

which describes a monoexponential curve that grows with a time constant  $T_{IS}$  and reaches the plateau corresponding to the amplitude of  $I_0$ .

Importantly, while  $T_{IS}$  and  $T_{1\rho}^S$  are characteristic for individual resonances,  $T_{1\rho}^I$  represent a volume property averaged over a distance of ca. 2 nm, as assumed by the fast  $I-I$  magnetization redistribution due to the efficient spin diffusion. This highlights a limitation in the  $I-S$  model for the description of CP kinetics in solids with heterogeneous populations of spin source.

On the contrary, the  $I-I^*-S$  model takes into account spin diffusion efficiency. The model originated from the study of stationary single crystals which revealed a transient harmonic oscillation for the CP of a C-H group.<sup>99</sup> This was explained with the existence of two different proton populations:  $I^*$  corresponding to the proton directly bound to the diluted spin system  $S$ , and  $I$  for the rest of the proton network.

In the single crystal case, spin diffusion is not fast enough to prevent the oscillatory CP transfer and hence the CP curve will present two different CP growing rates. At first, a fast rise of the intensity is observed due to the  $I^*-S$  close proximity. Then, the remote spins  $I$  will transfer polarisation to  $I^*$  via spin diffusion, which in turn will transfer it to  $S$ . This is characterised by a slow rise of the CP intensity or damped oscillation. Finally, for long contact times, magnetisation will eventually decay due to  $T_{1\rho}^H$  relaxation.

The general equation that describes the oscillatory CP behaviour of the  $I-I^*-S$  model for isolated pairs of spin-1/2 in static conditions is:

$$I(t) = I_0 \left[ 1 - \frac{1}{2} \exp(-Rt) - \frac{1}{2} \exp(-3Rt/2) \cos(bt/2) \right] \quad \text{Equation 2.47}$$

where  $R$  represents the spin-diffusion rate,  $b$  the dipolar coupling, and  $t$  the contact time. The equation was then modified to include both the effect of  $T_{1\rho}^I$  relaxation and of low-probability spin diffusion between  $S$  nuclei.

Generally, for isolated  $S$ - $I_n$  spin pairs in a powder sample under MAS conditions, the following equation applies:

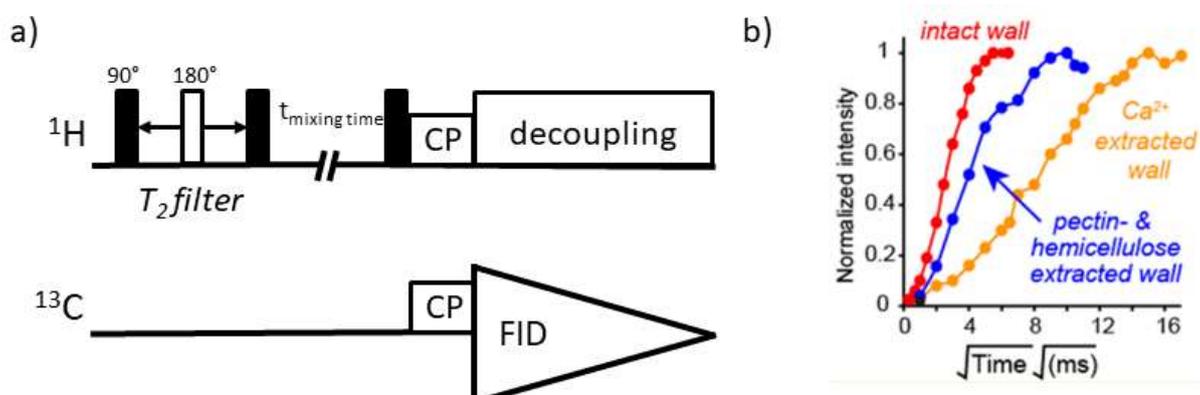
$$I(t) = I_0 \exp\left(-\frac{t}{T_{1\rho}^I}\right) \left[ 1 - \lambda \exp\left(-\frac{t}{T_{df}}\right) - (1 - \lambda) \exp\left(-\frac{3}{2} \frac{t}{T_{df}}\right) \exp\left(-\frac{1}{2} \frac{t^2}{T_2^2}\right) \right] \quad \text{Equation 2.48}$$

where  $T_{1\rho}^I$  is the  $I$  spin lattice relaxation constant in the rotating frame,  $T_{df}$  is the  $I$  spin-diffusion time constant describing the strength of the homonuclear dipolar interactions and the homogeneity of the  $I$  spin pool;  $T_2$  is the spin-spin relaxation constant, and  $\lambda$  is defined by the number of  $I$  spins ( $n$ ) attached to the  $S$  spin under study ( $\lambda = 1/(n + 1)$ ). Due to its dependence on mobility, the theoretical values of 1/2 and 1/3 for CH and CH<sub>2</sub> groups, respectively, does not always match with the experimental data and  $\lambda$  must be fitted. Usually, fitting gives values that do not deviate more than 0.1 from the theoretical values for CH and CH<sub>2</sub> groups, while for quaternary and methyl carbons,  $\lambda$  values in the range of 0.7-0.8 and 0.4-0.6, respectively, have been reported.<sup>99</sup>

To summarise, the  $I$ - $I^*$ - $S$  model applies when heteronuclear  $I^*$ - $S$  dipolar interactions are sufficiently strong compared to homonuclear  $I$ - $I$  dipolar interactions, which is often the case when <sup>1</sup>H and <sup>13</sup>C are in the same functional group and this is relatively immobile. On the contrary, for weak  $I$ - $S$  heteronuclear interactions and strong  $I$ - $I$  homonuclear dipolar coupling, the CP kinetics will most likely follow the classical  $I$ - $S$  model. The classical model is usually employed for polymers with substantial segmental chain motion.

### 2.2.9 Water-polarisation transfer – characterisation of water-network interactions in hydrogel systems by ssNMR

In solid-state NMR, water polarization transfer CP experiment (WPT-CP) is a well-established technique to gain site specific information on the hydration profile of solid particles.<sup>100</sup> WPT-CP experiments begins with a  $^1\text{H}$   $T_2$  filter sequence, optimized to select the  $^1\text{H}$  magnetisation of the mobile component (e.g. water in hydrogels) while removing the transverse magnetization of the immobile components of the system (e.g. cellulose fibrils). Hence, after the initial  $90^\circ$  pulse on the  $^1\text{H}$  channel, a  $180^\circ$  pulse in the middle of the  $T_2$  filter is applied to refocus the isotropic chemical shift evolution and  $B_0$  field inhomogeneity. Next, a  $^1\text{H}$   $90^\circ$  pulse stores the mobile component magnetisation along the z-axis and allows the transfer of  $^1\text{H}$  polarisation to the immobile components via spin diffusion (hence, distance- and mobility-dependent  $^1\text{H}$ - $^1\text{H}$  dipolar coupling) and chemical exchange mechanisms during the mixing time period. Finally, a CP building block transfers the  $^1\text{H}$  to  $^{13}\text{C}$  magnetisation for detection (Figure 2.21, a).



**Figure 2.21:** a) WPT-CP pulse sequence and b) WPT curves representing the normalized growth of peak area (peak intensities at each mixing time is normalized against the spectrum acquired with the longest mixing time) against the square root of mixing time. *Source:* Adapted from White *et al.*, 2014.<sup>100</sup>

The experiment is repeated for varying mixing times to obtain the so-called WPT curves (normalised peak intensity vs square root of mixing time) for each carbon peak (Figure 2.21, b). To avoid a strong contribution of spin diffusion during CP, the application of a short contact time (500  $\mu\text{s}$ ) is recommended. In this way, the acquired signal is mostly affected by water proximity and mobility.

Importantly, it should be noted that the intensities obtained in the WPT-CP experiments ultimately depends on  $^1\text{H}$ - $^{13}\text{C}$  cross-polarization efficiency (last step before FID acquisition); thus, only components that are immobile enough to cross-polarise will contribute to the spectra, while the most mobile components might not be “picked-up” by this experiment. In other words, WPT-CP curves report only on water interactions with rigid components.

### 2.3 Powder X-Ray Diffraction

PXRD is an analytical technique able to characterise the long-range ordering of crystalline and semi-crystalline materials based on their diffraction pattern. Crystals with precise periodicities over long distances give rise to sharp and clear diffraction peaks, while crystals with long range ordering defects (such as impurities, dislocations, planar faults, internal strains) produce diffraction patterns which present broadened, distorted and/or weakened diffraction peaks. Contrarily, amorphous materials lack a well-defined three-dimensional organisation, which results in strong peak

broadening and, hence, the lack of a distinguishable diffraction pattern. Importantly, each specific crystalline arrangement of atoms produces a unique diffraction pattern.

Crystals can be defined as solids presenting an ordered aggregation of atoms, ions or molecules, enclosed by symmetrically arranged plane surfaces and intersecting at defined and characteristic angles. In crystals, atoms repeat periodically in all three dimensions (crystal lattice), the smallest repeating pattern being defined as the crystal unit cell. Hence, we describe the unit cell in terms of the lattice (set of identical points). Seven possible shapes of unit cells have been defined, each one described by three side lengths  $a$ ,  $b$ , and  $c$  and the angles  $\alpha$ ,  $\beta$ , and  $\gamma$  (Table 2.1).

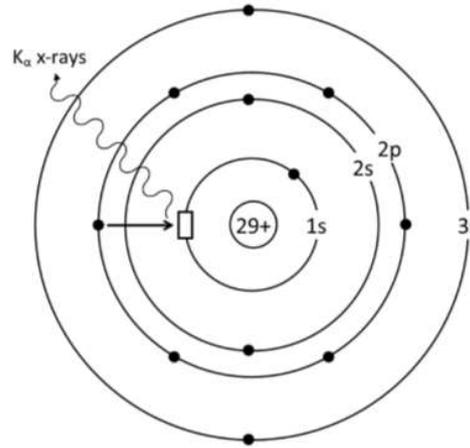
**Table 2.1:** The cell parameters reported by the Seven Crystal System for the possible shapes of unit cells.

Crystal system	Cell lengths	Cell angles
<b>Cubic</b>	$a=b=c$	$\alpha=\beta=\gamma=90^\circ$
<b>Tetragonal</b>	$a=b\neq c$	$\alpha=\beta=\gamma=90^\circ$
<b>Orthorhombic</b>	$a\neq b\neq c$	$\alpha=\beta=\gamma=90^\circ$
<b>Monoclinic</b>	$a\neq b\neq c$	$\alpha=\gamma=90^\circ\neq\beta>90^\circ$
<b>Triclinic</b>	$a\neq b\neq c$	$\alpha\neq\beta\neq\gamma\neq90^\circ$
<b>Hexagonal</b>	$a=b\neq c$	$\alpha=120^\circ\ \beta=\gamma=90^\circ$
<b>Rhombohedral</b>	$a=b=c$	$\alpha=\beta=\gamma\neq90^\circ$

Two main techniques are employed to access structural information of ordered materials. Single crystal X-ray diffraction provides precise information on unit cell parameters (position of atoms, bond lengths and angles between them), but it is limited to molecules or materials able to form stable and high-quality crystals of sufficient size. Powder X-ray diffraction (PXRD) enables the analysis of powdered materials reporting on the long-range planes of symmetry that characterise the material. In PXRD spectra, the X-ray intensities are plotted against the scattering angle.

X-rays are emitted by an anode bombarded by electrons generated from an electrically heated tungsten filament and accelerated in vacuum by high potential voltage of 20-60 kV. The anode can be composed of different metals, generally Cu. The electron emitted by the filament displace the electrons on the inner shell (K) of Cu. Hence, electrons from the outer shell descend emitting

energy in the form of radiation (Figure 2.22). Electrons that descend from the L and M shell emit  $K_{\alpha}$  and  $K_{\beta}$  radiation, respectively.  $K_{\alpha}$  radiations usually present a wavelength of 1.54056 Å.



**Figure 2.22:** Generation of X-rays through changes in the electronic structure of Cu.  
 Source: Lesley *at al.*, 2005<sup>101</sup>

X-rays electromagnetic waves present a wavelength ranging from 0.01-10 nm, which is on the same order of magnitude of inter-atomic distance of crystalline materials (1 Angstrom = 0.1 nm). X-rays can be scattered by a point object resulting in a new wave which spread in all directions. In the case that no loss of energy occurs, the wave maintains the same frequency. Two waves scattered by different points, will interfere with each other. The resulting amplitude depends on the phase shift ( $\delta\phi$ ) between the two waves (when the two waves are not coincident, hence do not have the same phase). This can be described on a linear scale in the wavelength units ( $\Delta$ ) or as a phase shift ( $\delta\phi$ ) on an angular scale:

$$\frac{\Delta}{\lambda} = \frac{\delta\phi}{2\pi} \Rightarrow \delta\phi = \frac{\Delta}{\lambda} 2\pi \quad \text{Equation 2.52}$$

In the two extreme cases, the detected intensity ( $I$ ) of two sine waves ( $A_1$  and  $A_2$ ) can present either the same ( $\delta\phi=0$ ) or  $\delta\phi=\pi$  shifted phase by constructive or destructive interference, and can be written as:

$$\text{For } \delta\phi = 0: I = (A_1 + A_1)^2 \quad \text{Equation 2.53}$$

For  $\delta\varphi = \pi$ :  $I = (A_1 - A_2)^2$  Equation 2.54

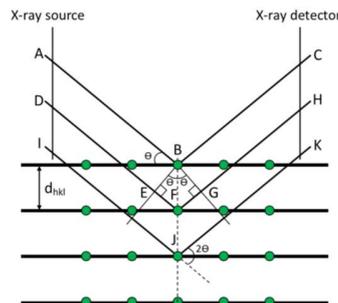
The above equations for two and more (j) waves can be written as:

$I = (A_1 + A_2 \exp(i\delta\varphi))^2$  Equation 2.55

$I = \left[ \sum_j A_j \exp(i\varphi_j) \right]^2$  Equation 2.56

Hence, the detected intensity corresponds to the square of the sum of the amplitudes of the X-rays waves scattered by the electrons of the atoms distributed in the three-dimensional space, meaning that the phase shift of the detected waves contains information about relative atomic positions. Application of the Bragg's Law then leads to structural information of crystalline solids.

Bragg's Law treats crystals planes of atoms acting as mirrors. To understand this, we need to look at Figure 2.23. When a monochromatic beam of parallel X-rays (ADI) is directed to a crystal (the green dots aligned in planes in Figure 2.23) with an angle  $\Theta$ , the ray A is scattered by atom B, while ray D and I penetrates the material and are scattered by atom F and J, respectively. As mentioned, the crystal is organised in planes, characterised by the Miller indices (hkl). When parallel, those planes are equally spaced by a distance  $d_{hkl}$ , and present the same Miller index.



**Figure 2.23:** Example of X-ray diffraction resulting in reflections at a particular angle  $\Theta$  used to derive Bragg's equation. *Source:* Dinnebier *et al.*, 2008<sup>101, 102</sup>

From Figure 2.23 it is evident that ray D travels longer distance before (EF) and after (FG) being reflected, in comparison with ray A. Nonetheless, as both rays are reflected in the same direction,

superimposition occurs, with a constructive interference arising only when the difference in distance  $\Delta = EF + FG$  is a multiple ( $n$ ) of wavelength ( $\lambda$ ):

$$\Delta = EF + FG = n\lambda \text{ and } EF = FG \quad \text{Equation 2.57}$$

In geometrical terms (based on Figure 2.3)

$$\sin \alpha = \frac{EF}{d_{hkl}} \Rightarrow EF = d_{hkl} \sin \alpha \quad \text{Equation 2.58}$$

$$\Delta = EF + FG = 2d_{hkl} \sin \alpha \quad \text{Equation 2.59}$$

$$n\lambda = 2d_{hkl} \sin \alpha \quad \text{Equation 2.60}$$

Equation 2.60 is known as Bragg's equation and it relates the spacing between the individual planes ( $d_{hkl}$ ) with the Bragg's angle  $\Theta_{hkl}$  at which the reflections from those planes are observed. The Miller indices (100), (110), (111) are used to label the observed reflections in PXRD according to the plane from which they diffract. However, in the case when the reflected waves are out of phase, destructive interferences occur, and no diffraction patterns are detected.

Importantly, powder XRD patterns can be simulated from previous diffraction data using *Mercury*, a high-quality crystal structure visualization and investigation software developed by the Cambridge Crystallographic Data Centre (CCDC). A Crystallographic Information File (.cif) containing atoms coordinates can be loaded in the software and the simulated powder diffraction pattern calculated and displayed using the Mercury Powder Diffraction Pattern tool. The pattern is plotted as diffraction intensity against  $2\Theta$  for a given wavelength of radiation (for example,  $\text{CuK}_{\alpha 1} = 1.54059$ ).<sup>103</sup>

## 2.4 Molecular modelling<sup>104</sup>

Molecular modelling is a broad field that spreads from theoretical and semi-empirical quantum mechanical (QM) methods to semi-empirical Molecular Mechanics (MM) methods. QM methods aim to an accurate description of the molecule electronic environment. On the contrary, the MM approach is based on the Born-Oppenheimer approximation, which treats the motions of nuclei

and electrons separately. Due to their smaller size, the motions of the electrons can be neglected, and the atomic nuclei can be described as spherical masses that follows classical mechanical physic.

Application of QM methods to large biological molecules (i.e. proteins, DNA and lipid membrane) is computationally intense and most of the time not feasible, while the simplified MM approach is widely used to calculate molecular geometries and energies.

### 2.4.1 Molecular Mechanics

In MM, atomic nuclei are described as “balls” of different sizes (depending on the atom type) and given coordinated, which are connected by springs. The balls interact with each other through processes of bond stretching, dihedrals distortion, angle opening and closing.

The basic concept of MM is that bonds have equilibrium lengths and angles which are defined in the so-called “force field”. Deviations from these standard values results in an increase of the total energy of the molecule, described by the potential energy function. In fact, the overall potential energy  $U$  for a molecular system as a function of the internal coordinates  $\vec{R}$  of the  $N$  particles constituting the system can be expressed in classical mechanic terms as

$$\begin{aligned}
 U(\vec{R}) = & \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 \\
 & + \sum_{dihedral} k_\chi(1 + \cos(\eta\chi - \delta)) \\
 & + \sum_{Coul\ i,j} \frac{q_i q_j}{\epsilon r_{ij}} + \sum_{LJ\ i,j} \left( \frac{A_{ij}}{r_{ij}^{12}} + \frac{B_{ij}}{r_{ij}^6} \right)
 \end{aligned}
 \tag{Equation 2.61}$$

where the constants  $k_b$ ,  $k_\theta$  and  $k_\chi$ , as well as the bonds and angles reference volumes  $b_0$  and  $\theta_0$  are extracted from experimental data (hence it falls into the semi-empirical methods category) or QM calculation and expressed in the force field. The different energy terms are resolved by harmonic functions which penalise distortion from the reference values. Not only the contribution of bond stretching ( $\sum_{bond}$ ), angles ( $\sum_{angles}$ ) and proper (and improper) dihedrals re-arrangement ( $\sum_{dihedral}$ ) are considered, but also non-bond terms are taken into account: the electrostatic interactions ( $Coul\ i, j$ ) and the Lennard-Jones ( $LJ\ i, j$ ) repulsion-dispersion potential energy terms.

Thus, in the second line of the equation,  $q_i$  and  $q_j$  are the charges of the interacting particles,  $\epsilon$  is the solvent dielectric constant,  $A_{ij}$  and  $B_{ij}$  are the Lennard-Jones coefficient and  $r_{ij}$  is the distance between the particles. Importantly, the nature of each particle and bond and the parameters relative to them (e.g., the atoms mass and volume) must be specified. To find the optimal molecular geometry, the values of the coordinates of each atom are searched by minimising the force field equation.

### 2.4.2 Protein-ligand docking calculations<sup>105, 106</sup>

Molecular docking is a computational procedure (based on the force field theory) used to predict the preferred orientation and conformation of a ligand bound to its target protein (protein-ligand docking). Although several docking software packages exist (e.g., DOCK, FlexX, AutoDock, AutoDock Vina, etc.), they all share common features to:

- 1) Locate regions on the protein surface for favourable interactions with the ligands
- 2) Sample the conformational space of the ligand
- 3) Compute the interaction energy between the protein and ligand

Recently, the group of Imberty conducted a comparison study between three docking software (DOCK, AutoDock and Glide) to predict the binding mode of flexible carbohydrate in shallow lectins binding site. They found that Glide is the most successful to account for the flexibility of the oligosaccharides and to represent the complexity of their binding modes.<sup>107</sup>

Ideally, in performing docking calculation both the docked protein and ligands should be set free to rotate, translate and change conformation so that a more realistic interaction is reproduced. Nonetheless, calculations performed in this way are not feasible due to the high-number of degrees of freedom characterising the protein-ligand system. Thus, approximations about the protein and ligand flexibility states have been introduced. In rigid docking (the simplest approximation possible), the protein and the ligand are treated as two distinct rigid bodies and are enabled only to translate and rotate. In more computationally demanding techniques, only the receptor is treated as a rigid body, while the ligand is enabled to explore its conformational degree of freedom (those are the most widely used docking algorithms).

In both cases, to perform docking calculations the same practical steps must be followed (shortly described below). In the Schrödinger suite docking calculations are performed by the Glide program, to which we will refer henceforth.

### 1) Protein preparation and ligand conformational minimisation

Before to start any molecular modelling study, some preparations steps are necessary from both the protein and the ligand prospective. Common sources for protein 3D molecular structures are X-ray crystallography data, NMR studies or homology modelling. Whatever the source is, the protein structure needs to be checked and prepared to ensure that 1) all the atom types and bond lengths, angles and orders are correctly specified, 2) no hydrogen atom is missing, and 3) all the residues in the binding pocket have defined orientation (in some cases, amino acid side chains can present several conformations in the X-ray structure). Then, selection of the number and type of water molecules to keep in the binding pocket is possible. Finally, a round of minimisation is performed.

For the ligand, instead, a conformational search and a minimisation step must be performed. In this way, the starting conformation of the molecule is in a low relative minimum and represents the free molecule in solution. Notably, during docking calculations, the docking algorithms usually generates new ligand conformations. Nonetheless, those are based only on torsional variations, while bond lengths and angles in the input ligand structure will be kept consistent among the docked poses. For these reasons, docking several conformations of each ligand with variations in bond lengths and bond angles is a reasonable strategy to reduce input dependence. In the Schrödinger suite the receptor and ligand preparations are carried out by the protein preparation wizard and the ligand preparation wizard, respectively.

### 2) Grid generation

The space of the receptor in which the ligand will be docked is defined by generating a grid in which the ligand is confined (Figure 2.24). The grid is generated by the Receptor Grid Generation application. The centre of the grid is defined using a reference ligand in the binding pocket or by choosing any residues of interest. In glide, the grid has a default cubic shape with 20 Å sides length, but this value can be changed (30 Å maximum length for each side). The length of each side can be set to different values; hence the grid can have different shapes. In addition, an inner grid can be

defined, which encloses the centre of the ligand (defined as the midpoint of the longest segment which can be drawn between any two atoms of the molecule). Each sides of the inner box are set to a default value of 10 Å, but it can be modified into the 6-14 Å range.

The size of the grid strongly affects the outcome of the docking calculations: larger grids allow to explore a wider space and it is useful when probing unusual and asymmetric binding modes in the active site; on the other hand, smaller grids will reduce the number of results and exclude odd poses (saving calculation time).

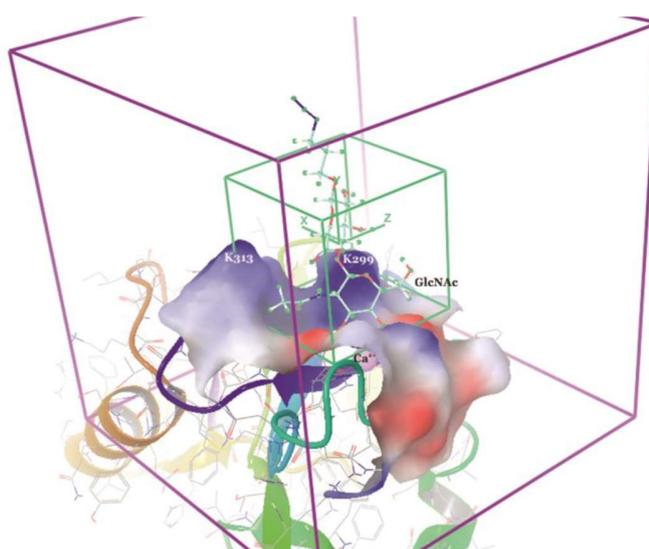


Figure 2.24: 3D image showing the grid box (purple-lined cube) and the ligand diameter midpoint box (green-lined cube) in glide.

### 3) Docking calculations

Several docking algorithms can be applied to generate ligand structures. The algorithms can be grouped into deterministic and stochastic approaches, the first is reproducible, while the second presents random factors and is not fully reproducible. The most common algorithms used in docking are the incremental construction, the genetic and the hierarchical algorithms. They all aim to define the best location for the ligand on the protein surface (within the grid), to sample the conformational space of the ligand and to compute the energy of the protein-ligand interaction, giving higher score to the most favourable.

Glide docking program uses a hierarchical algorithm, also defined as a “funnel” process, where the number of poses decreases stepwise (Figure 2.25). Initially a set of ligand conformations is generated and deterministically screened over the entire space available to the ligand (rough scoring). The identified promising ligand poses are then refined in torsional space in the field of the receptor using a standard molecular mechanics energy function with a distance-dependent dielectric model (Glide SP & XP use OPLS3 as force field). Finally, the lowest-energy poses are subjected to a Monte Carlo procedure with full ligand flexibility to minimize the poses within the field of the receptor (post-docking minimization or PDM).

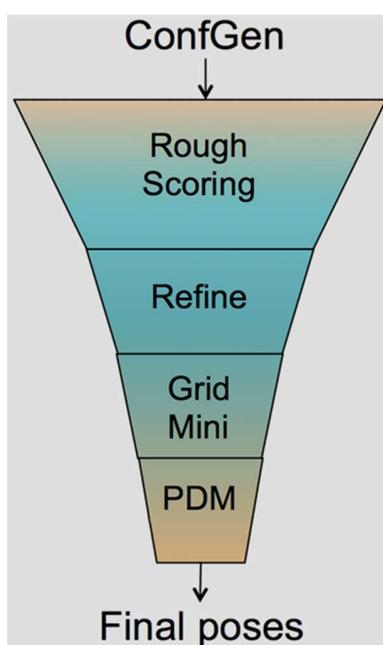


Figure 2.25: Glide docking “funnel” showing the protocol followed to generate docked poses.

Source: Harder *et al.*, 2016<sup>108</sup>

#### 4) Scoring functions

The poses produced after the calculation are then scored and ranked. Energy scoring functions evaluate the free energy of binding ( $\Delta G$ ) of the ligand-receptor interaction, described by the Gibbs free energy equation

$$\Delta G = \Delta H - T\Delta S$$

Equation 2.62

where  $\Delta G$  represents the energetic changes between the free and bound states of both ligand and receptor,  $\Delta H$  is the enthalpy,  $T$  is the temperature expressed in Kelvin and  $\Delta S$  is the change in entropy of the system.  $\Delta G$  is also related to the binding association and dissociation constants ( $K_a$  and  $K_d$ , respectively), allowing to estimate binding affinities

$$\Delta G^\circ = -RT \ln K_a = RT \ln K_d \quad \text{Equation 2.63}$$

where  $R$  is the ideal gases constant.

Nonetheless, calculation of  $\Delta G$  for each of the generated ligand-receptor interactions poses is computationally demanding. Thus, empirical scoring functions have been developed to enable a faster ranking. Terms such as solvation, ionic interactions and hydrogen bonds are used to construct an equation that predicts binding affinities. These terms are then optimised via multilinear regression using a set of known protein-ligand complexes. Several functions are available, and they serve diverse needs. “GlideScore” in Glide, for example, has been optimised to compare binding of different ligands to one receptor (e.g., to define the best binder among many), while the “Emodel” scoring function weights more the electrostatic and van der Waals energetic component in the force field and is more suitable to compare conformations of the same ligand to its receptor. Glide uses Emodel to select the best poses of the ligand and then ranks these best poses with GlideScore. Therefore, the ranking of poses for a given ligand does not reflect the actual ranking that Glide used for poses selection. Therefore, the Emodel score needs to be considered to determine the highest ranked pose for a ligand.

### 2.4.2.1 *Induced Fit Docking*

While consideration of the protein and/or the ligand as rigid bodies allow a reduction in the computational costs, it does not reflect reality and can give rise to misleading results. Indeed, induced fit phenomenon within the binding site (i.e. rearrangements of the protein side chain or backbone upon binding to better adapt to the shape and binding mode of the ligand) might take place. The performance of Induced Fit Docking (IFD) protocols allows screening of the ligand pose against several receptor conformations with refined active site geometries in the presence of the ligand.<sup>109</sup> This should 1) allow the docking of known active ligands that cannot be docked in the rigid procedure, and 2) rescue rigid docking false negatives (poorly scored true binders).

## Chapter 2 - Techniques and tools

First, the active ligand is docked with Glide using reduced van der Waals (vdW) radii and an increased Coulomb-vdW cutoff. In addition, highly flexible side chains are temporarily removed during the docking step. Hence, a protein structure prediction program (Prime) is used to reorient protein side chains nearby the ligand and minimisation of both the ligand and the protein residues is performed. Each ligand pose is then redocked into its corresponding low energy protein structure. The resulting energies are ranked based on GlideScore and Prime energies scoring functions.

Importantly, this protocol can be used to construct models of protein-glycans interactions. Glycans generally present weak binding constants and rigid docking produces a high output of poorly scored true binders.

## Chapter 3

Cellodextrin phosphorylase

from *Clostridium thermocellum*:

structural studies on substrates recognition

## 3. Cellodextrin phosphorylase from *Clostridium thermocellum*: structural studies on substrates recognition

### 3.1 Introduction

#### 3.1.1 Principal aims of the work

The scientific community is increasingly using enzymatic routes for the synthesis of oligosaccharides and glycoconjugates, although with some limitations on enzymes availability. Glycoside phosphorylases (GPs) are a class of enzymes able to build up short-to-medium length sugar chains starting from activated donor and acceptor substrates. Wild-type cellodextrin phosphorylase (CDP, EC 2.4.1.49:  $\beta$ -1,4-glucan linkage-dependent) tolerates a number of non-cognate sugar-1-phosphate donors and several acceptor substrates, proving to be a powerful biological tool for the synthesis of new sugars with functional groups introduced in a regio- and stereo-selective manner. The enzyme demonstrated reduced kinetics efficiency for non-natural substrates.

Herein, we have characterised the molecular recognition of a series of cognate and non-cognate CDP substrates by NMR spectroscopy and docking calculations. To gain a complete picture of the structural details of the molecular recognition of substrates by CDP under the dynamics conditions existing in solution, and hence complementing solid state structures (i.e. from X-ray crystallography), other techniques are valuable, such as Saturation Transfer Difference NMR spectroscopy (STD-NMR) and transferred NOESY (tr-NOESY) experiments. Specifically, STD NMR allows one to map the key ligand protons interacting with the protein (binding epitope mapping)<sup>81</sup> and to determine protein-ligand dissociation constants ( $K_D$ ).<sup>79</sup> Tr-NOESY experiments are used to measure intramolecular proton-proton distances of ligands in the state bound to the protein and hence they report on the bioactive conformation of the ligand and probe conformational rearrangements upon binding.<sup>110</sup>

In addition, we generated 3D-models the binary complexes CDP/donor and CDP/acceptor and the ternary complex CDP/donor/acceptor by molecular modelling (protein-ligand SP docking and

## Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

induced fit docking IFD calculations run in MAESTRO Schrödinger using Glide as force field) and analysed the resulting structures on the basis of their correlation with experimental data from NMR spectroscopy, X-ray crystallography and with literature.

This study reveals atomistic details of substrate recognition and structural features that contribute to donor and acceptor specificity. In addition, we focused in the characterisation of the role played by inorganic phosphate (enzymatic co-factor) on substrates binding properties.

### 3.1.2 Importance of phosphorylases in carbohydrate synthesis

The number of reported phosphorylases constantly increases, proving how these enzymes are useful tools for the practical synthesis of  $\alpha$ - and  $\beta$ -glucans. Table 3.1 lists all the phosphorylases belonging to the EC 2.4.1 subclass (Transferases, Glycosyltransferases and Hexosyltransferases) stated in ExPASy (Bioinformatic Resource Portal - <https://enzyme.expasy.org/EC/2.4.1.->) to date.<sup>111</sup> Phosphorylases classification can be based on: (1) the anomeric forms of the substrates glycosidic linkages, (2) the anomeric forms of the glyucose 1-phosphate, and (3) the retention or inversion mechanism of the carried reaction.<sup>22, 112</sup>

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

**Table 3.1:** Catalogue of known phosphorylases and their characteristics. The reaction mechanism catalysed is classified based on anomeric retention or inversion in the reaction. Generally, reversible phosphorylation determines the liberation of monosaccharide 1-phosphate from the nonreducing-end of specific glycosides in the presence of inorganic phosphate. The enzymes are named combining the name of the substrate and phosphorylase (continuous next page).

<u>EC</u>	<b>Enzyme</b>	<b>Mechanism</b>	<b>Substrate</b>	<b>Product</b>
<a href="#">2.4.1.1</a>	Glycogen phosphorylase	retention	((1→4)-α-D-glucosyl)(n)	α-D-Glc-1-P
<a href="#">2.4.1.7</a>	Sucrose phosphorylase	retention	Sucrose	α-D-Glc-1-P
<a href="#">2.4.1.8</a>	Maltose phosphorylase	inversion	Maltose	β-D-Glc-1-P
<a href="#">2.4.1.20</a>	Cellobiose phosphorylase	inversion	Cellobiose	α-D-Glc-1-P
<a href="#">2.4.1.30</a>	1,3-β-oligoglucan phosphorylase	inversion	((1→3)-β-D-glucosyl)(n)	α-D-Glc-1-P
<a href="#">2.4.1.31</a>	Laminaribiose phosphorylase	inversion	Laminaribiose	α-D-Glc-1-P
<a href="#">2.4.1.49</a>	Cellodextrin phosphorylase	inversion	((1→4)-α-D-glucosyl)(n)	α-D-Glc-1-P
<a href="#">2.4.1.64</a>	α,α-trehalose phosphorylase	inversion	α,α-trehalose	β-D-Glc-1-P
<a href="#">2.4.1.97</a>	1,3-β-D-glucan phosphorylase	inversion	((1→3)-α-D-glucosyl)(n)	α-D-Glc-1-P
<a href="#">2.4.1.211</a>	1,3-β-galactosyl-N-acetylhexosamine phosphorylase	inversion	((1→3)-β-D-glucosyl)(n)	α-D-Glc-1-P
<a href="#">2.4.1.216</a>	Trehalose 6-phosphate phosphorylase	inversion	α,α-trehalose-6-P	β-D-Glc-1-P
<a href="#">2.4.1.230</a>	Kojibiose phosphorylase	inversion	2-α-D-glucosyl-D-glucose	β-D-Glc-1-P
<a href="#">2.4.1.231</a>	α,α-trehalose phosphorylase	retention	α,α-trehalose	α-D-Glc-1-P
<a href="#">2.4.1.247</a>	β-D-galactosyl-(1→4)-L-rhamnose phosphorylase	inversion	β-D-galactosyl-(1→4)-L-rhamnose	α-D-Gal-1-P

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

**Table 3.1:** continued

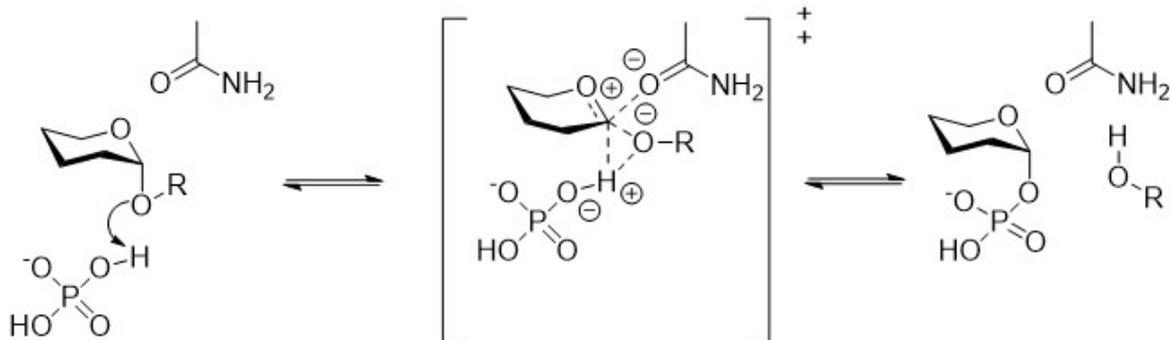
<b>EC</b>	<b>Enzyme</b>	<b>Mechanism</b>	<b>Substrate</b>	<b>Product</b>
<a href="#"><u>2.4.1.279</u></a>	Nigerose phosphorylase	inversion	3-O- $\alpha$ -D-glucopyranosyl-D-glucopyranose	$\beta$ -D-Glc-1-P
<a href="#"><u>2.4.1.280</u></a>	N,N'-diacetylchitobiose phosphorylase	inversion	N,N'-diacetylchitobiose	$\alpha$ -D-GINAc-1-P
<a href="#"><u>2.4.1.281</u></a>	4-O- $\beta$ -D-mannosyl-D-glucose phosphorylase	inversion	4-O- $\beta$ -D-mannopyranosyl-D-glucopyranose	$\alpha$ -D-Man-1-P
<a href="#"><u>2.4.1.282</u></a>	3-O- $\alpha$ -D-glucosyl-L-rhamnose phosphorylase	inversion	3-O- $\alpha$ -D-glucopyranosyl-L-rhamnopyranose	$\beta$ -D-Glc-1-P
<a href="#"><u>2.4.1.319</u></a>	$\beta$ -1,4-mannooligosaccharide phosphorylase	inversion	((1 $\rightarrow$ 4)- $\beta$ -D-mannosyl)(n)	$\alpha$ -D-Glc-1-P
<a href="#"><u>2.4.1.320</u></a>	1,4- $\beta$ -mannosyl-N-acetylglucosamine phosphorylase	inversion	4-O- $\beta$ -D-mannopyranosyl-N-acetyl-D-glucosamine	$\alpha$ -D-Man-1-P
<a href="#"><u>2.4.1.321</u></a>	Cellobionic acid phosphorylase	inversion	4-O- $\beta$ -D-glucopyranosyl-D-gluconate	$\alpha$ -D-Glc-1-P
<a href="#"><u>2.4.1.329</u></a>	Sucrose 6(F)-phosphate phosphorylase	retention	Sucrose-6(F)-phosphate	$\alpha$ -D-Glc-1-P
<a href="#"><u>2.4.1.332</u></a>	1,2- $\alpha$ -glucosylglycerol phosphorylase	inversion	2-O- $\alpha$ -D-glucopyranosyl-glycerol	$\beta$ -D-Glc-1-P
<a href="#"><u>2.4.1.333</u></a>	1,2- $\beta$ -oligoglucan phosphorylase	inversion	((1 $\rightarrow$ 2)- $\beta$ -D-glucosyl)(n)	$\alpha$ -D-Glc-1-P
<a href="#"><u>2.4.1.334</u></a>	1,3- $\alpha$ -oligoglucan phosphorylase	inversion	((1 $\rightarrow$ 3)- $\alpha$ -D-glucosyl)(n)	$\beta$ -D-Glc-1-P
<a href="#"><u>2.4.1.339</u></a>	$\beta$ -1,2-mannobiose phosphorylase	inversion	$\beta$ -D-mannopyranosyl-(1 $\rightarrow$ 2)-D-mannopyranose	$\alpha$ -D-Man-1-P
<a href="#"><u>2.4.1.340</u></a>	1,2- $\beta$ -oligomannan phosphorylase	inversion	((1 $\rightarrow$ 2)- $\beta$ -D-mannosyl)(n)	$\alpha$ -D-Man-1-P
<a href="#"><u>2.4.1.352</u></a>	Glucosylglycerate phosphorylase	retention	2-O- $\alpha$ -D-glucopyranosyl-D-glycerate	$\alpha$ -D-Glc-1-P
<a href="#"><u>2.4.1.359</u></a>	Glucosylglycerol phosphorylase	retention	2-O- $\alpha$ -D-glucopyranosyl-glycerol	$\alpha$ -D-Glc-1-P

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

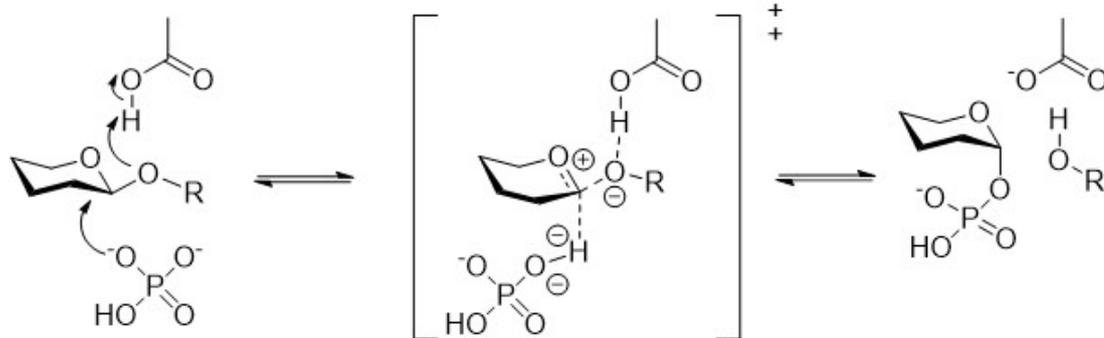
Phosphorylases can also be classified, based on their amino acid sequence similarity, into the glycoside hydrolase families (GH) 13, 65, 94, 112, 130, or glycosyltransferase families (GT) 4 or 35, as reported in the Carbohydrate-Active Enzymes database (<http://www.cazy.org/>).<sup>113</sup>

The enzymes catalyse the cleavage of a glycosidic linkage at the non-reducing end of a saccharide chain in the presence of inorganic phosphate (also called exo-wise phosphorolysis). The reaction produces the corresponding glucose-1-phosphate and the saccharide chain with one smaller DP. In the reverse reaction the glucose unit is transferred from the glycosyl 1-phosphate (donor) to the non-reducing end of an appropriate glycosyl acceptor with the formation a stereo- and region- controlled glycosidic linkage and inorganic phosphate liberation. The reversibility of the reaction is dictated thermodynamically by the comparable bond energy of the glycosyl-phosphate and the glycosidic-linkage.<sup>22, 112</sup> A schematic representation of the retaining and inverting reaction mechanisms is shown in Figure 3.1.

## a) Direct front-side nucleophilic displacement mechanism



## b) Phosphorolysis by inverting phosphorylase



**Figure 3.1:** a) schematic representation of the phosphorylases catalysed  $S_N2$  with retaining mechanism; b) schematic representation of the phosphorylases catalysed  $S_N2$  with inverting mechanism. Inverse phosphorolysis has been described as a general acid/base-catalysed displacement reaction, with the phosphate operating a nucleophilic attack on the anomeric carbon of the scissile glycosidic bond. Concomitantly the general acid catalyst in the binding site donates a proton to the glycosidic oxygen.<sup>114</sup> In contrast to the typical glycosyl hydrolase mechanism, the catalytic reaction requires only a single carboxylate side-chain as the phosphate is ionised at physiological pH. *Source:* Nakai *et al.*, 2013.<sup>112</sup>

Phosphorylases present strict stereo- and regio-specificities, important characteristics for the synthesis of oligo- and polysaccharides with well-defined structure. As mentioned in Chapter 1, section 1.3.1, while several phosphorylases have been employed for the synthesis of poly- and oligosaccharides, others are just able to catalyse the reversible phosphorolysis from and towards the disaccharide substrates. For example, cellodextrin phosphorylases, laminaridextrin

## Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

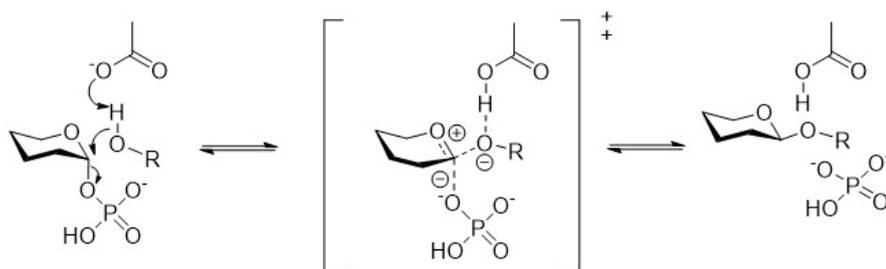
phosphorylase, and kojibiose phosphorylases have been reported to efficiently catalyse the linear extension of carbohydrate acceptor, while maltose phosphorylase, nigerose phosphorylase, trehalose phosphorylase, cellobiose phosphorylase, and chitobiose phosphorylase catalyse the formations of the corresponding glycobioses.<sup>112</sup>

### 3.1.3 Cellodextrin phosphorylase

Cellodextrin phosphorylase, a  $\beta$ -1,4-oligoglucan orthophosphate glucosyltransferase, was firstly isolated in 1967 from *Ruminiclostridium thermocellum* by Sheth and Alexander.<sup>115</sup> To date CDPs have been isolated from several bacterial sources, such as *Ruminiclostridium thermocellum*<sup>115</sup>, *Clostridium stercorarium*<sup>116</sup>, *Ruminococcus albus*,<sup>117</sup> *Thermosiphon africanus*<sup>118</sup> and *Ruminiclostridium cellulolyticum*.<sup>119</sup> These bacteria are able to degrade cellulose and metabolize the soluble cellodextrins as part of the energy catabolism. A number of GH94 CDP have been cloned, expressed and characterised.<sup>116, 118-121</sup> Among these, CDP from *Ruminiclostridium thermocellum* and *Clostridium stercorarium* are the most studied. Interestingly, although the two enzymes are far away in the GH94 family phylogenetic tree and share only *ca.* 25% amino acid sequence identity,<sup>120</sup> Petrović *et al.* reported no differences in the catalysis of cellobiose and Glc-1-P.<sup>122</sup>

CDP performs a reverse phosphorolysis reaction (Figure 3.2) to synthesise cellooligosaccharides with average DP of 7 or 9, depending on the primer used.<sup>123</sup> Importantly, the complete dissociation of the enzyme and the product after addition of each glucose unit explains why precipitation of the substrate interrupts the elongation.<sup>122</sup> Indeed, cellooligosaccharides with DP >7 are poorly soluble in water and readily precipitate, becoming not accessible to further chain elongation.<sup>124</sup>

Reverse phosphorolysis



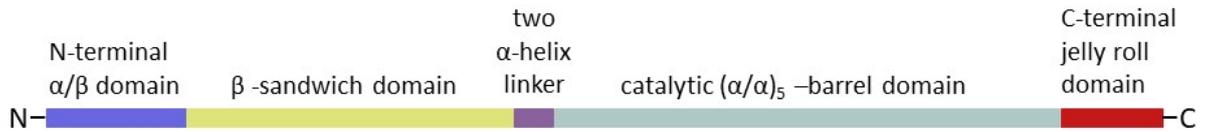
**Figure 3.2:** Schematic representation of the reverse phosphorolysis performed by CDP in the presence of a general glucose-1-phosphate (donor) and a general acceptor molecule.

Recombinant CDP from *Ruminiclostridium thermocellum* (henceforth depicted as CDP) was expressed with high yield in *E. coli*.<sup>120, 125</sup> It is stable up to 60 °C and presents highest activity at pH 7.5.<sup>120</sup> The protein consists of 1009 amino acids, with a MW of 114.364 kDa per monomer (in solution, it forms a dimer), as reported by the GF elution profile.<sup>126</sup> The enzyme catalyses the phosphorolysis of cellooligosaccharides longer than cellobiose, and its permissiveness towards a wide range of donors and acceptors has been assessed. This will be discussed in the following subchapter.

The publication of the CDP crystal structures in the apo and bound to D-cellobiose acceptor ligand states (PDB: 5NZ7 apo CDP structure and 5NZ8 ligand bound CDP structure), by O'Neill *et al.*, opened the avenue for an in-depth CDP structural characterisation and provided valuable information on key interactions involving substrate recognition.<sup>126</sup>

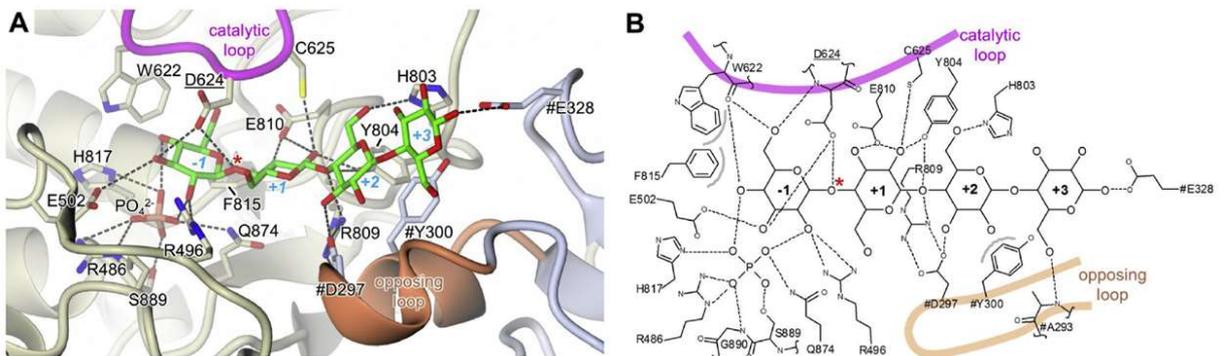
The apo structure (2.3 Å resolution) revealed that CDP adopts a homodimeric structure<sup>127</sup> with the dimer interface formed by a large β-sandwich domain. Each monomer is composed by the N-terminal α/β domain, the β-sandwich domain that is connected by a two α-helix linker to the (α/α)<sub>6</sub>-barrel catalytic domain and a final two layered-jelly roll fold peripheral domain (Figure 3.3). The N-terminal arm and the α/β domain were reported to interact with the β-sandwich domain of the opposing monomer, maintaining a more open active site in CDP and leading to a dimer interfacial area of ca. 4800 Å<sup>2</sup> (in comparison, CBP misses the N-terminal arm and the α/β domain and has ca. 3300 Å<sup>2</sup> of dimer interfacial area).

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*



**Figure 3.3:** Schematic linear representation of CDP domain structures.  
 Source: O’Neil *et al.*, 2017.<sup>126</sup>

CDP structure bound to D-cello-tetraose was obtained by soaking 10 mM D-cellohexahose into a crystal which had been grown in the presence of 10 mM phosphate buffer. The electron densities collected (3.0 Å resolution) revealed the presence of D-cello-tetraose adjacent to a phosphate ion in the active site cleft. All the sugar rings were treated as  $\beta(1\rightarrow4)$ -linked  ${}^4C_1$  chairs and the glucan was oriented with the non-reducing terminal in the inner side of the catalytic cleft, located between Asp624 (expected general acid catalyst) and the phosphate ion. The glucan spanned from subsite -1 to subsite +3 of the active side pocket, occupying the -1 donor site (Figure 3.4, a and b).



**Figure 3.4:** Representation of the subsites in the catalytic cleft. The absence of electron density after the +3 subsite did not exclude the presence of the additional two residues of D-cellohexaose, but rather delimits the extent of acceptor ordered binding. The catalytic loop terminates at the +1 subsite level, the acceptor site opens out and the glucan chain extends across a wide U-shaped canyon formed at the dimer interface. Source: O’Neil *et al.*, 2017.<sup>126</sup>

The -1 subsite terminates with the “hydrophobic platform” formed by Trp622 sidechain, structurally conserved in the close homologues and important to exclude water around the scissile bond region. A lobe adjacent to the active site cleft accommodate the phosphate ion. Since the binding of the glucan to the -1 subsite closes the access to the phosphate lobe, the co-substrate

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

must bind before the sugar in the phosphorolytic reaction, coherent with the sequential Bi Bi mechanism. Beyond the +1 subsite the acceptor site opens out. Nonetheless, the homodimeric structural organisation of the protein determines interaction of the residues occupying the +2 and +3 subsites with the sidechains of the  $\beta$ -sandwich domain in the opposite monomer unit. Important interactions for substrate recognition are highlighted in Figure 3.4, b. More specifically, (i) the carboxylate of the catalytic Asp624 hydrogen bond with both the O3 of the -1 subsite sugar and the oxygen of the scissile glycosidic linkage; (ii) the 6OH of the -1 subsite sugar hydrogen bonds with both the carbonyl of Trp622 and the amide of Asp624 (main chains), while C6 makes van der Waals interactions with Trp622 and Phe815 side chains; (iii) Cys625, the adjacent residue within the catalytic loop, hydrogen bonds with the O2 of the +1 subsite sugar; (iv) from the opposite loop, Asp297 hydrogen bonds within O3 of the +2 subsite sugar, while Tyr300 forms a stacking interaction with the sugar ring; (v) Glu328 sidechain, also from the opposing subunit but from a different loop, hydrogen bond with the O1 of the +3 subsite sugar.

#### 3.1.3.1 CDP specificity towards donor and acceptor substrates

CDP presents the same regiospecificity as cellobiose phosphorylase but differs in the specificity towards the acceptor DPs. Indeed, the large difference in acceptor reactivity between D-glucose and longer cellooligosaccharides is well accepted among the scientific community. Nonetheless, in several works the use of D-glucose or  $\beta$ -D-glucoside as CDP acceptors has been reported. In addition, CDP present a very low phosphorolysis rate on D-cellobiose.<sup>128</sup>

CDP is very effective in the synthesis of functionalised cellooligosaccharides. Samain *et al.* pioneered the effort to apply unnatural substrates for CDP and established the use of 4-Thiocellobiose, methyl  $\beta$ -cellobioside, and methyl 4-thio- $\alpha$ -cellobioside as acceptor substrates.<sup>129</sup> CDP presents loose substrate specificities, and a complete list of the variety of donors and acceptors recognised by the enzyme is presented in Table 3.2 and Table 3.3, respectively, while the molecules which did not report any turn-over are reported in Table 3.4.

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

**Table 3.2:** CDP specificities towards donor-like substrates.

#### Donors

$\alpha$ -D-glucosyl 1-fluoride<sup>114</sup>

$\alpha$ -D-xylose 1-phosphate<sup>126, 130</sup>

$\alpha$ -D-galactose 1-phosphate<sup>126, 131</sup>

$\alpha$ -D-glucosamine 1-phosphate<sup>126, 131</sup>

$\alpha$ -D-galactosamine 1-phosphate<sup>131</sup>

**Table 3.3:** CDP specificities towards acceptor and acceptor-like substrates. The red frame indicates D-glucose or  $\beta$ -D-glucoside derivatives as carbohydrate moiety of the acceptor-like substrates (continuous next page).

Acceptors
D-Glucose <sup>123, 124, 128, 132-137</sup>
Radioactive D-Glucose <sup>138</sup>
1-azide-1-deoxy- $\beta$ -D-glucopyranoside ( $\beta$ -glucosyl azide) <sup>132</sup>
Oligo(ethylene glycol) (OEG) bearing $\beta$ -D-glucose <sup>139</sup>
Alkyl $\beta$ -D-glucoside <sup>140</sup>
2-aminoethyl- $\beta$ -D-glucoside <sup>141</sup>
2-(glucosyloxy)ethyl methacrylate <sup>142</sup>
vinyl glucosides <sup>143</sup>
4-O- $\beta$ -D-Glucopyranosyl-D-glucose (cellobiose) <sup>114, 120, 122-124, 128, 144</sup>
Methyl $\beta$ -cellobioside <sup>124</sup>
Phenyl $\beta$ -cellobioside <sup>124</sup>

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

**Table 3.3:** continued

Benzophenone  $\beta$ -cellobioside<sup>124</sup>

4-Thiocellobiose<sup>124</sup>

Methyl 4-thio- $\alpha$ -cellobioside<sup>124</sup>

Methyl  $O$ - $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 3)-S- $\beta$ -D-glucopyranosyl-(1 $\rightarrow$ 4)- $O$ -(4-thio- $\beta$ -D-glucopyranosyl)-(1 $\rightarrow$ 4)- $\beta$ -D-glucopyranoside<sup>129</sup>

N,N-bis( $\beta$ -D-cellobiosyl)succinamide<sup>145</sup>

tris(aminoethyl N-carbonylmethyl  $\beta$ -D-cellobiosyl)amine<sup>145</sup>

$\beta$ -D-cellobiosyl polyamidoamide (PAMAM) dendrimers<sup>145</sup>

2- $O$ - $\beta$ -D- Glucopyranosyl-D-glucose (sophorose)<sup>114</sup>

4- $O$ - $\beta$ -D- Glucopyranosyl-D-altrose<sup>146</sup>

4- $O$ - $\beta$ -D-Glucopyranosyl-D-2-deoxy- D-glucose<sup>146</sup>

4- $O$ - $\beta$ -D- Glucopyranosyl-D-mannose<sup>146</sup>

4- $O$ - $\beta$ -D- Glucopyranosyl-D-xylose<sup>130, 146</sup>

4- $O$ - $\beta$ -D- Glucopyranosyl-1-deoxynojirimycin<sup>125</sup>

4- $O$ - $\beta$ -D-Glucopyranosyl-D-2-deoxy- D-glucose<sup>146</sup>

4- $O$ - $\beta$ -D- Glucopyranosyl-D-mannose<sup>146</sup>

4- $O$ - $\beta$ -D- Glucopyranosyl-D-xylose<sup>130, 146</sup>

4- $O$ - $\beta$ -D- Glucopyranosyl-1-deoxynojirimycin<sup>125</sup>

$\beta$ -D-2-ammine-glucopyranosyl- $\beta$ -(1 $\rightarrow$ 4)-D-glucopyranosyl- $\beta$ -(1 $\rightarrow$ 4)-D-glucopyranose<sup>147</sup>

4- $O$ - $\beta$ -D-Xylopyranosyl-D-xylose (xylobiose)<sup>130</sup>

4- $O$ - $\beta$ -D- Xylopyranosyl-D-glucose<sup>130</sup>

Cellobitol<sup>146</sup>

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

**Table 3.4:** List of tested monosaccharides, disaccharides and trisaccharides acceptors for which CDP Glc-1-P did not show any turn-over. In addition, the only donor-like substrate which did not show any catalytic efficiency is reported in red.

Xylose <sup>114, 115</sup>	N-Acetyl-glucosamine* <sup>114</sup>	Mannobiose* <sup>114</sup>
L-Rhamnose* <sup>114</sup>	$\beta$ -methyl glucoside <sup>115</sup>	Sucrose* <sup>114</sup>
Salicin <sup>115</sup>	Melibiose* <sup>114</sup>	Talose* <sup>114</sup>
Lactulose* <sup>114</sup>	Isomaltose* <sup>114</sup>	L-Fucose* <sup>114</sup>
Maltose <sup>114, 115</sup>	Mannotriose <sup>126</sup>	N-Acetyl-mannosamine* <sup>114</sup>
Arabinobiose* <sup>114</sup>	Galactose* <sup>114</sup>	Lactose <sup>114, 115</sup>
Xylose <sup>114, 115</sup>	Fructose* <sup>114</sup>	Trehalose* <sup>114</sup>
Mannose* <sup>114</sup>	N-Acetyl-galactosamine* <sup>114</sup>	Turanose* <sup>114</sup>
Arabinose* <sup>114</sup>	Gentibiose <sup>114, 115</sup>	<b>Mannose-1-P<sup>126, 131</sup></b>

\* The acceptor specificity was determined exclusively at 4.4 mM CDP concentrations for 3 h using 200 mM  $\alpha$ -Glc1-P and 200 mM of the tested acceptor at pH 5.4 and 40 °C.

In general terms, the turn-over for the reported acceptors was tested with Glc-1-P as donor, with only Shintate and Singh reporting on the transfer of unnatural donor to unnatural acceptors. Indeed, Shintate demonstrated the efficiency of Xyl-1-P towards xylose containing disaccharides.<sup>130</sup> Recently, Singh *et al.* reported for the first time the CDP-catalysed  $\beta$ -1,4-galactosylation and  $\beta$ -1,4-glucosamination of both  $\beta$ -1,3-linked and mixed-linked oligosaccharide acceptors.<sup>131</sup>

Importantly, CDP ability to synthesise derivatised  $\beta$ -(1 $\rightarrow$ 4)-oligosaccharides has been used for the synthesis of 1,3: 1,4- $\beta$ -d-Glucanases from *Bacillus licheniformis*<sup>129</sup> and cellulase inhibitors<sup>125</sup>, and for the subsite mapping of (1 $\rightarrow$ 3,1 $\rightarrow$ 4)- $\beta$ -d-glucan endohydrolases.<sup>148</sup> Table 3.5 and Table 3.6 report the kinetics parameters of various cognate and non-cognates CDP donor and acceptor substrates, respectively.

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

**Table 3.5:** Kinetics parameters of CDP donor substrates reported towards the synthetic direction. Donor specificity was determined by the measurement of phosphate release keeping the acceptors concentration constant (5 mM). Source: the data were taken from Singh (2020)<sup>131</sup> and O’Neill (2017)<sup>126</sup>

Reverse phosphorolysis				
Donors	Acceptor	$k_{cat}$ (s <sup>-1</sup> )	$K_M^{app}$ (mM)	$k_{cat}/K_M^{app}$ (mM <sup>-1</sup> s <sup>-1</sup> )
<b>Glc-1-P</b>	D-cellobiose	16.4 ± 0.7	3.0 ± 0.6	5.5
	D-laminaribiose	15.9 ± 0.4	3.0 ± 0.3	5.3
<b>Gal-1-P</b>	D-cellobiose	0.6 ± 0.02	9.3 ± 1.1	0.06
	D-laminaribiose	0.6 ± 0.01	10.7 ± 0.7	0.06
<b>GlcN-1-P</b>	D-cellobiose	0.08 ± 0.003	1.6 ± 0.2	0.05
	D-laminaribiose	0.14 ± 0.013	5.1 ± 1.2	0.03
<b>Man-1-P</b>	D-cellobiose	NA	NA	NA
	D-laminaribiose	NA	NA	NA

**Table 3.6:** Kinetics parameters of CDP acceptor substrates reported towards the synthetic direction. Acceptor specificity was determined by the measurement of phosphate release keeping Glc-1-P concentration constant (10 mM).

Reverse phosphorolysis			
Acceptors	$k_{cat}$ (s <sup>-1</sup> )	$K_M^{app}$ (mM)	$k_{cat}/K_M^{app}$ (mM <sup>-1</sup> s <sup>-1</sup> )
<b>Glucose</b> <sup>120, 126, 128</sup>	NA/nd	NA/nd	NA/nd
<b>Phenylβ-D-glucopyranoside</b> <sup>126</sup>	15.0 ± 6.3	24 ± 13	0.63
<b>Cellobiose</b> <sup>126</sup>	17 ± 0.50	2.6 ± 0.18	6.5
<b>Cellotriose</b> <sup>126</sup>	9.5 ± 0.35	0.68 ± 0.076	14
<b>Cellotetraose</b> <sup>126</sup>	5.0 ± 0.25	0.54 ± 0.13	9.3
<b>Cellopentaose</b> <sup>126</sup>	4.3 ± 0.47	0.36 ± 0.076	12
<b>Cellohexaose</b> <sup>126</sup>	7.6 ± 1.8	1.9 ± 0.76	4.0
NA not applicable; nd not determined			

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

Comparison of Glc-1-P with the non-cognate donor substrates reveals a *ca.* 25 fold and *ca.* 125-fold drop in  $k_{cat}$  for Gal-1-P and GlcN-1-P, respectively, and a slight increase of  $K_M^{app}$  for both the donors. These data indicate a lower catalytic efficiency for Gal-1-P and GlcN-1-P, as expressed by  $k_{cat}/K_M^{app}$ . Man-1-P did not show turnover.

According to the literature, D-glucose was found to be a poor CDP acceptor and the kinetics parameters could not be determined.<sup>128</sup> On the contrary, phenyl  $\beta$ -D-glucopyranoside was proved a better acceptor, probably due to the fixed anomeric configuration. The tested cellooligosaccharides show a decrease of  $K_M^{app}$  from 2.6 mM to 0.36 mM with DP increase up to 5, while cellohexahose presented a higher  $K_M^{app}$  of 1.9 mM.  $k_{cat}$  follows a similar pattern with respect to the acceptor chains length.

#### 3.1.3.2 *Obtaining ligand SAR from X-ray crystal structure and substrate specificity*

Comparison between the CDP-bound D-cellobiose crystal structure (PDB: 5NZ8)<sup>126</sup>, and the turn-over capacity of the tested donor and acceptor substrates enabled us to rationalise the enzyme catalytic efficiency. We can first look at the contacts established by D-cellobiose non-reducing ring, which occupies the -1 subsite in the catalytic cleft.

The hydroxyl group in C2 position establishes hydrogen bonds with Arg496, explaining the lack of turn-over demonstrated for Man-1-P, mannose, manno-1,6-bisphosphate and mannose-6-phosphate and the lower catalytic efficiency (drop in  $k_{cat}/K_M^{app}$ ) of GlcN-1-P. For mannoglycosides, the configurational switch of the hydroxyl group at C2 from equatorial to axial could either break the hydrogen bonds established with the side-chains of Arg496 and Gln874 and, in case of acceptor substrates, with the phosphate group, or create a clash with the preserved catalytic residue Asp624, with severe impact on the overall interactions. GlcN-1-P reduced catalytic activity, instead, could be explained by the positive charge carried by Arg496, which would interfere with the donor NH<sub>2</sub> group. In addition, the good acceptor activity of  $(\beta(1\rightarrow4)\text{-GlcN}(\text{Glc})_2)$  toward extension with Glc-1-P demonstrates that a change of the oxygen with a nitrogen at C2 of the +1 subsite acceptor sugar does not alter significantly the acceptor binding.<sup>126</sup>

The contacts established by the hydroxymethyl group, the OH sugar hydrogen bonds with the carbonyl of Trp622 and the amide of Asp624 backbones and the C6 van der Waals interactions with

Trp622 and Phe815 side chains, explain the poor donor activity of xylose 1-phosphate, which lacks both C6 and O6.

## 3.2 Experimental section

### 3.2.1 Spectral assignment of donor and acceptor substrates

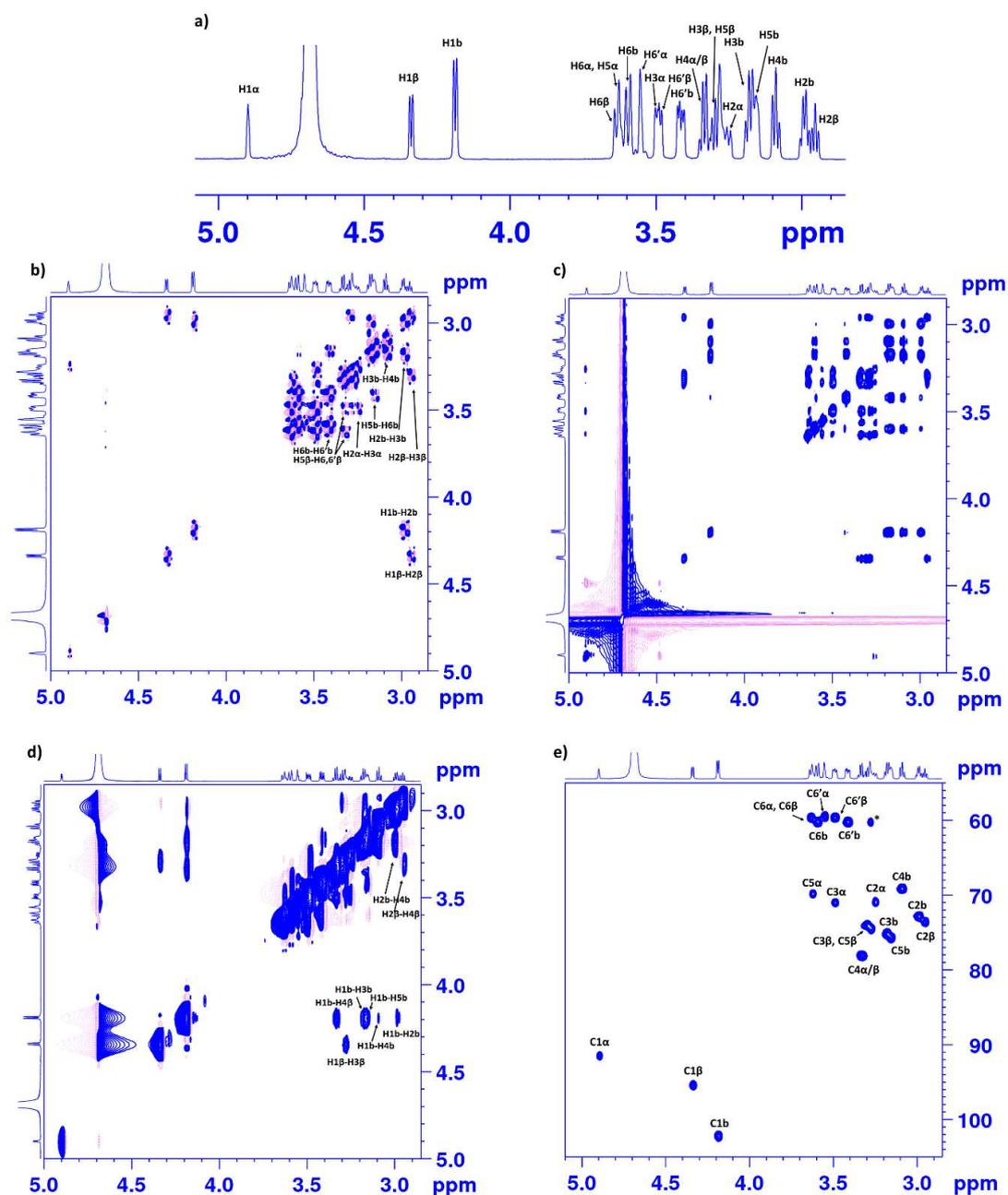
Before proceeding with the investigation of donor and acceptor substrates binding, a full assignment of the  $^1\text{H}$  1D spectrum for each studied ligand was necessary (i.e. Figure 3.5, a – representing the proton spectrum of D-cellobiose 3 mM, in  $[\text{D}_{11}]\text{Tris}$  25 mM). The assignment of the different proton signals was obtained by comparison of the connectivity patterns in 2D spectra ( $^1\text{H}$ - $^1\text{H}$  COSY,  $^1\text{H}$ - $^1\text{H}$  TOCSY,  $^1\text{H}$ - $^1\text{H}$  NOESY and  $^1\text{H}$ - $^{13}\text{C}$  HSQC).

For example, the  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts of D-cellobiose could be assigned in sequence: the H1 proton resonances, at 4.90, 4.35 and 4.18 ppm for H1 $\alpha$ , H1 $\beta$ , and H1b, correlate with the directly coupled H2 signals at 3.25, 2.95 and 2.99 ppm in the  $^1\text{H}$ - $^1\text{H}$  COSY spectrum (Figure 3.5, b). The acquisition of  $^1\text{H}$ - $^1\text{H}$  TOCSY experiment allowed the identification of each individual spin system (Figure 3.5, c). The  $^1\text{H}$ - $^1\text{H}$  NOESY spectrum shows the cross-relaxation peak between H1 (4.18 ppm) and H4 (3.33 ppm), belonging to the glycosidic linkages (Figure 3.5, d). In addition, as shown in the HSQC spectrum, the C6 resonance at ca.  $\sim$  60 ppm and correlate with the proton peaks at 3.63 ppm, 3.59 ppm, 3.55 ppm, 3.49 ppm, 3.41 ppm and 3.28 ppm (Figure 3.5, c). The C2 (73.5, 72.8 and 70.8 ppm) and C3 (75.1, 74.0 and 70.9 ppm)  $^{13}\text{C}$  resonances in the HSQC spectrum of the D-cellobiose correlate with the directly coupled H2 (2.95, 2.99 and 3.25 ppm) and H3 (3.2, 3.3 and 3.48 ppm) signals, respectively (Figure 3.5, e).

Using a similar method, we were able to assign the chemical shifts of the other studied substrates. For the less complex structure of the monosaccharide donors, the spectra assignment was possible by only the acquisition of COSY and HSQC spectra.

The spectra acquired for each donor and acceptor substrates are reported in the Appendix.

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*



**Figure 3.5:** a) COSY, b) TOCSY, c) NOESY (mixing time 400 ms) and d) HSQC spectra recorded for D-cellobiose (3 mM, in [D<sub>11</sub>]Tris 25 mM) at 278 K.

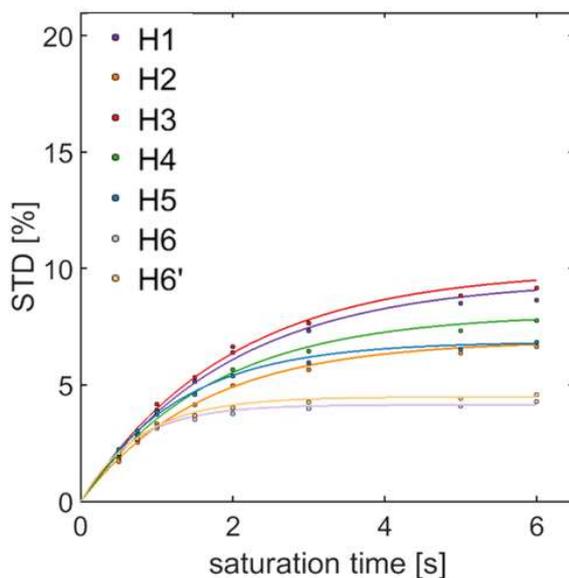
### 3.2.2 Structural basis of molecular recognition of natural and non-cognate donor substrates

The first part of this study focused on the investigation of Glc-1-P (CDP natural donor), and a series of sugar 1-phosphate molecules (i.e. the glucose isomers Gal-1-P and Man-1-P, and the functionalised glucose analogues GlcN-1-P and 6-deoxy-6-fluoro- $\alpha$ -D-glucose-1-phosphate 6F-Glc-1-P). The choice of these donor-like substrates was driven by the kinetics studies reporting that modifications on the hexopyranose ring of the sugar 1-phosphate ligands affect the enzymatic activity of CDP dramatically. As for these sugar 1-phosphate series binding to CDP is expected to occur at the medium/weak affinity range (as a guide,  $K_M$  values fall in the low mM range),<sup>114, 120, 126, 128, 131</sup> i.e. within the so-called fast chemical exchange conditions, the NMR technique of choice was saturation transfer difference (STD) NMR spectroscopy. This technique allows to identify differences in ligand-enzyme contacts between the non-cognate substrates and the natural substrate Glc-1-P, at atomic detail.

#### 3.2.2.1 Detection of binding by STD NMR

First, we confirmed that binding to CDP was detectable by STD NMR for all the small molecules chosen. After confirmation of binding, we carried out series of STD NMR experiments at different saturation times, in order to gain structural information on the complexes. To that aim, we monitored the growth of saturation transfer for every proton of the ligands (STD build-up curves), and from these curves we determined the corresponding ligand group epitope mappings using the initial growth rates approach (described in Chapter 2, section 2.1.2.3).<sup>82</sup> Figure 3.6 shows the STD build-up curves with the monoexponential fit for Glc-1-P, from which  $STD^{max}$ ,  $k_{sat}$ ,  $STD_0$  and normalised percentages (Table 3.7) were derived. The fitted STD build-up curves and corresponding  $STD^{max}$ ,  $k_{sat}$ ,  $STD_0$  and normalised percentages for all the other investigated ligands are reported in the Appendix.

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*



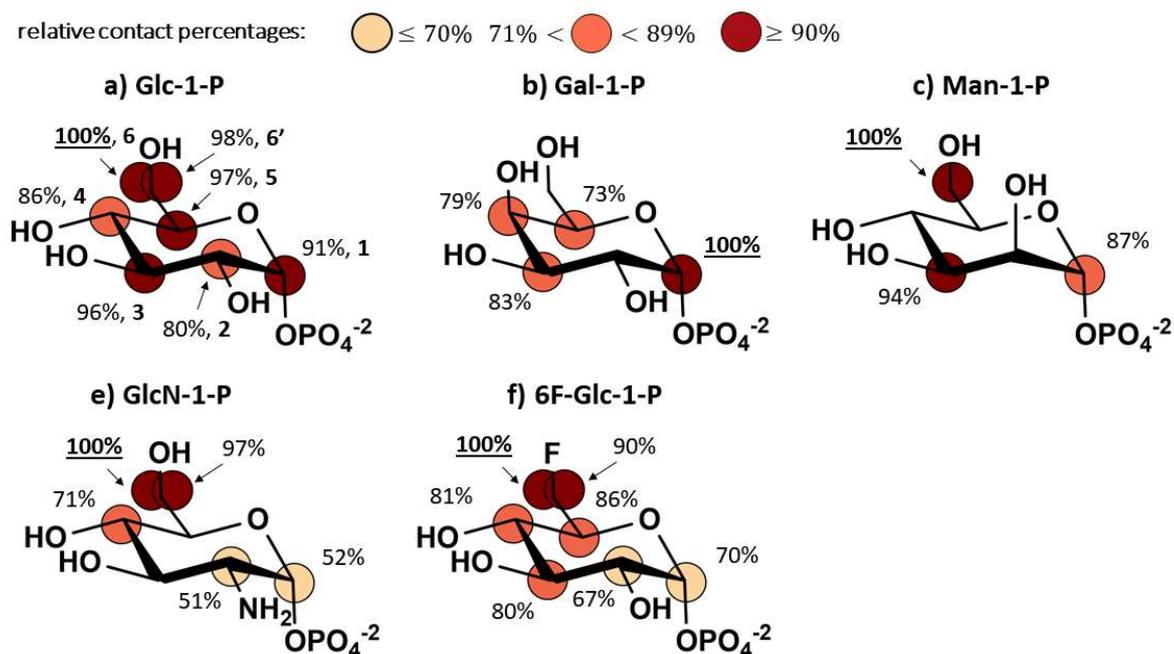
**Figure 3.6:** Glc-1-P STD NMR build-up curves recorded at increasing saturation time at 278 K. 15  $\mu$ M binding unit was used for a ligand concentration of 3 mM.

**Table 3.7:**  $STD_{max}$ ,  $k_{sat}$  and  $STD_0$  for Glc-1-P. Relative STD (%) were obtained through normalisation against the largest ligand STD initial slope (H6; 100%)

	$STD_{max}$	$k_{sat}$	$STD_0$	STD (%)
H1	9.70	0.52	5.06	91
H2	7.01	0.64	4.45	80
H3	9.95	0.54	5.34	96
H4	8.12	0.59	4.80	86
H5	7.03	0.77	5.42	98
H6	4.31	1.29	5.56	100
H6'	4.77	1.15	5.48	98.5

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

Figure 3.7 shows the group epitope mappings for Glc-1-P and the other four non-cognate sugar 1-phosphate ligands. The percentages on different regions of the ligands report on their distinct proximity to the surface of the CDP enzyme in the bound state (higher normalised STD values correspond to shorter ligand-enzyme distances).



**Figure 3.7:** Group epitope mapping of a) Glc-1-P, b) Gal-1-P, c) Man-1-P, d) GIN-1-P and e) 6F-Glc-1-P from STD NMR for the interactions with CDP determined by using the initial slope approach of each STD build-up curve. The assigned relative contact percentages were obtained through normalisation against the maximum ligand STD initial slope for each ligand and report on the saturation received by those protons from the enzyme (higher values corresponding to shorter ligand-enzyme distances). Numbering of the glucopyranose atom positions are reported in Glc-1-P in bold after the STDs relative percentages in Glc-1-P. The COSY and HSQC spectra recorded for peaks assignment (Figures A.1, A.2, A.4, A.6 and A.8), the STD build-up curves with the monoexponential fit (Figures A.3, A.5, A.7 and A.9), the derived  $STD^{max}$ ,  $k_{sat}$ ,  $STD_0$  and normalised percentages (Tables A.1, A.2, A.3 and A.4) are reported in the Appendix. All the experiments were run at 278 K.

### 3.2.2.2 *Molecular recognition of the cognate donor substrate Glc-1-P*

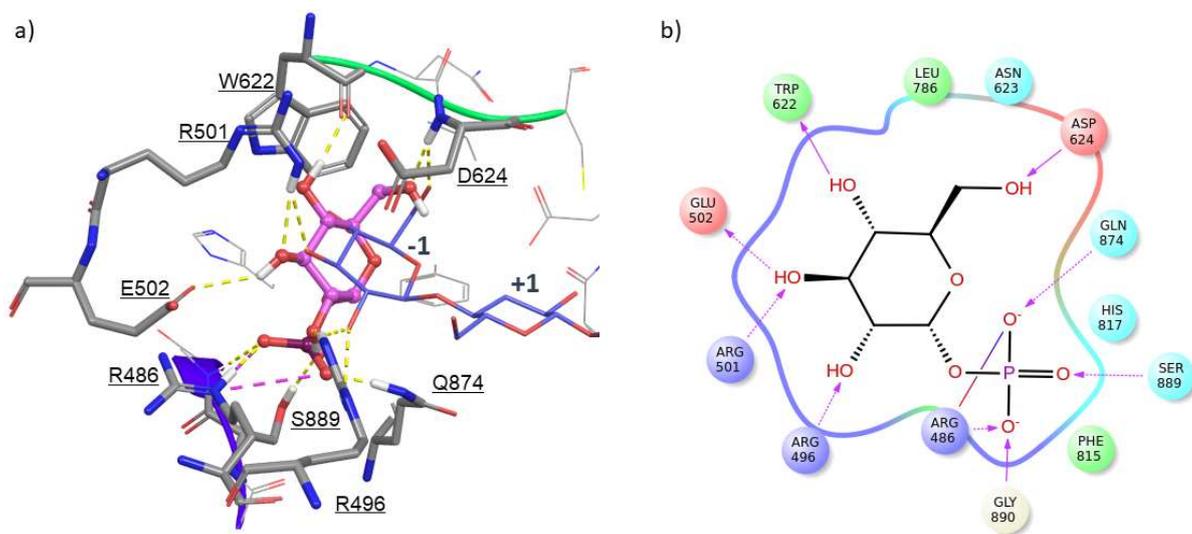
For the natural CDP donor substrate, Glc-1-P, the analysis of the STD initial slopes revealed that the whole glucopyranose ring is in very close contact with CDP in the bound state. All the non-exchangeable ring protons of Glc-1-P show normalised STD values above 80% (Figure 3.7, a), which supports the intimate recognition of Glc-1-P in CDP donor subsite. The large STD on H1 indicates a close recognition by CDP of the  $\alpha$ -configuration at the anomeric centre, in good agreement with the known specificity of the enzyme (the enzyme does not use the anomer  $\beta$ -D-glucose-1-phosphate as a donor).<sup>20, 22, 112</sup> The largest saturation transfer occurs to protons H5 and H6 of the glucopyranose ring. This result suggests a major role of the hydroxymethyl group at C5 of Glc-1-P in binding to CDP and reflects the poor kinetics and low turnover reported for Xyl-1-P as donor substrate for CDP. Finally, H2 and H4 showed lower contribution to donor substrate recognition, with H2 showing the lowest STD.

To deepen our understanding of the molecular recognition of Glc-1-P, we generated a 3D model of the CDP/Glc-1-P complex by molecular modelling (protein-ligand SP docking calculations run in MAESTRO Schrödinger<sup>149</sup> using Glide as force field<sup>105</sup>). The resulting structure was analysed on the basis of its correlation with the experimental data from STD NMR. The most populated solution from the docking calculations is shown in Figure 3.8, where a superimposition of the obtained pose to the terminal non-reducing glucose ring of the published CDP-bound D-cellotetraose structure is presented.<sup>126</sup>

In the model, Glc-1-P accommodates a position compatible to the performance of the nucleophilic attack, with the phosphate group binding in a lobe adjacent to the -1 subsite of the active site. This position is the same occupied by the inorganic phosphate in the crystal structure of bound CDP.<sup>126</sup> As shown in Figure 3.8, the phosphate establishes hydrogen bonds with four different residues (Arg486, Gln874, and Ser889 sidechains, and Gly890 backbone) acting as an H-bond acceptor and a salt bridge with Arg486. The glucopyranose ring is anchored within the binding pocket residues at the -1 subsite, with the hydroxyl groups in C2, C3 and C6 acting as H-bonds acceptors with the side chains of Arg496 and Arg501 and the backbone NH of Asp624, respectively. Additionally, the hydroxyl groups in C3 and C4 act as H-bond donors with the sidechain of Glu502 and the backbone CO of Trp622.

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

The described model is in very good agreement with the experimental binding epitope from STD NMR experiments (Figure 3.8, a). The glucopyranose ring is in very close proximity with CDP, with protons H2 and H4 being more water exposed due to a slight tilt of the sugar ring in comparison to the non-reducing terminal ring of D-celotetraose (Figure 3.8, a). Remarkably, the hydroxymethyl group is located between Trp622 and Asp624, making close contacts with the Trp622 side chain. This explains the most notable transfer of saturation towards protons H6, H6' and H5 (Figure 3.7, a). The function of Trp622 side chain as a “hydrophobic platform” has been described already,<sup>150</sup> and O’Neil *et al.*<sup>126</sup> reported the establishment of van der Waals contacts between the hydroxymethyl group of D-celotetraose non-reducing ring and the residues Trp622 and Phe815, as well as the establishment of H-bond between 6OH of the -1 subsite sugar within both the carbonyl of Trp622 and the amide of Asp624 (main chains). In this study we show for the first time the atomic details that make these contacts key elements for the recognition of the natural donor substrate. In addition, the proximity to Asp624 at the CDP catalytic cleft agrees with its known pivotal role in enzyme catalysis due to its ability to act as proton donor/acceptor in the  $S_N2$  reaction mechanism.<sup>150</sup>



**Figure 3.8:** a) 3D docking model of the CDP/Glc-1-P complex. The D-celotetraose complex determined by X-ray crystallography is superimposed for comparison (PDB ID 5NZ8,<sup>126</sup> purple wire representation). For simplicity, only two rings of D-celotetraose are shown (sites -1 and +1). The protein residues establishing interactions with Glc-1-P are represented as thick tubes. Glc-1-P is represented as ball-and-stick and painted in pink. Non-bonded interactions are represented as

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

dash-lines; H-bonds are coloured in yellow and salt bridges in magenta; b) Ligand interaction diagram reported to facilitate the visualisation of the set of interactions established by Glc-1-P with the binding pocket residues in the -1 subsite. The arrows indicate H-bond going from donor to receiver, dash lines represent H-bond with residues sidechain, whereas the solid line with residues backbone. The solid line shaded from red to blue represents a salt bridge.

#### 3.2.2.3 *Molecular recognition of non-cognate donor-like epimer ligands: Gal-1P and Man-1-P*

The NMR-validated 3D molecular model of the CDP/Glc-1-P complex (Figure 3.8) shows that the binding pocket has enough room to accommodate some configurational or functional changes on the hexopyranose ring. This is in agreement with the already proven ability of CDP to use both non-cognate donor and acceptor substrates, albeit with reduced catalytic efficiency in comparison to Glc-1-P. The investigation of glucose epimers at position C4 (Gal-1-P) and position C2 (Man-1-P) allowed us to decipher how a configurational inversion at these positions affects donor binding in the -1 subsite of the catalytic pocket. Importantly, the performed STD NMR experiments provide a clear evidence of the ability of Man-1-P to bind CDP. Hence, the binding of Man-1-P to bind CDP is herein demonstrated for the first time, while this was only suggested previously by its impact on the enzymatically produced cellodextrin oligomer length.<sup>126</sup>

The group epitope mapping of Gal-1-P (Figure 3.7, b) shows that H1 establishes the closest contact with the CDP binding site. Interestingly, the configurational change at C4 gives rise to a rearrangement of the hexopyranose ring in comparison to Glc-1-P, inferred from the observed reduced saturation transfer to protons H3 and H5 in Gal-1-P. On the contrary, the group epitope mapping of Man-1-P (Figure 3.7, c) supported that the configurational change at C2 does not affect the molecular recognition of the hexopyranose ring by CDP significantly. Indeed, within the experimental error, the group epitope mapping of Man-1-P is very similar to that of Glc-1-P, with protons H1, H3 and H6 receiving large saturation transfer, supportive of an intimate recognition by CDP as in the case of Glc-1-P.

These results are relevant as Man-1-P is not processed by CDP.<sup>126, 151</sup> Our STD NMR study thus demonstrates that the C2 configurational change, although detrimental for the catalytic activity, neither impairs binding nor affects the binding mode. In this way, the results show that the null

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

catalytic efficiency for Man-1-P cannot be explained by a change in the binding mode in the donor site, in comparison to Glc-1-P, but rather by the effect that the configurational change at C2 has on the network of interactions with the catalytic residues of CDP in the donor binding subsite -1. Most likely, the axial orientation of the hydroxyl group at C2 of the mannose ring precludes the catalytic loop from shifting towards the substrate, remaining far from the ligand as in the holo-enzyme structure.<sup>126</sup> In this conformation, the carboxylate side chain of Asp624 is not within H-bond distance of the anomeric carbon, and cannot act as the general acid catalyst in the proposed  $S_N2$  reaction mechanism. Previous studies already supported the presence of a steric clash of the axial hydroxyl with the catalytically competent Asp624.<sup>126, 131</sup>

#### 3.2.2.4 *Molecular recognition of functionalised donor-like ligands: GlcN-1-P and 6F-Glc-1P*

The group epitope mapping of the functionalised GlcN-1-P (Figure 3.7, d) and 6F-Glc-1P (Figure 3.7, e) revealed close contact of hydroxymethyl group and significantly reduced in H1 and H2 contacts in comparison with Glc-1-P. In the case of GlcN-1-P, the reduction in contact reported for H2 can be explained by steric hindrance and electrostatic repulsion of the amine group with the positively charged side chain of Arg496, located in the -1 subsite of the catalytic cleft (see Figure 3.8).<sup>126, 131</sup> In addition, the close contact for H6s in 6F-Glc-1-P indicate recognition of the fluorine group of this derivative, supporting tolerance of CDP for a group at position 6 isosteric to OH acting as an H-bond acceptor. This is in perfect agreement with the proven ability of CDP to use 6F-Glc-1-P as donor-like substrate.<sup>152</sup>

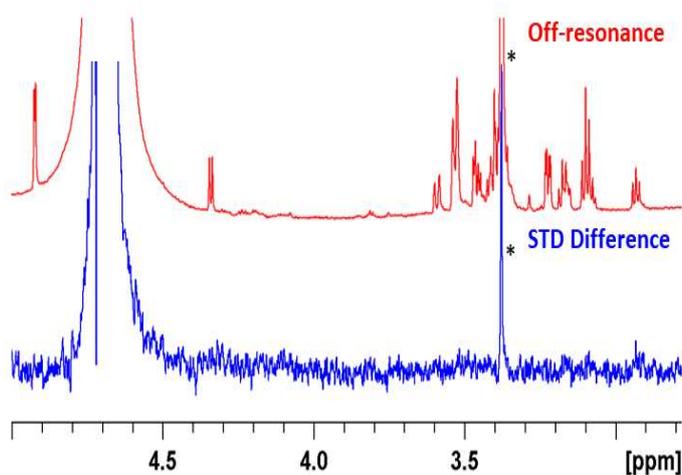
The NMR validated 3D model of the CDP/Glc-1-P complex along with the comparison of relative STD values for all the non-cognate donor-like ligands (Figure 3.7) strongly support: (i) the importance of the presence of an H-bond acceptor at position 6 of the hexopyranose ring, and (ii) a key relevance of the equatorial hydroxyl at position 4. As regarding the latter, Gal-1-P received the lowest saturation transfer in comparison to the other donor-like ligands (see Appendix, Figures A.3, A.5, A.7 and A.9), compatible with a lower affinity, and shows a significantly different binding epitope. These data correlate very well with previous studies in which epimerization at C4 led to an increase in the  $K_m^{app}$  from 3 mM (Glc-1-P) to 9.3 mM (Gal-1-P), whereas for functionalised Glc-1-P analogues, like GlcN-1-P, only a slight decrease on the  $K_m^{app}$  was reported.<sup>131</sup>

Binding of the functionalized glucose-1-phosphate analogues correlates very well with the observed ability of CDP of using them as active donor substrates. The introduction of an amine group at C2 (GlcN-1-P) results in a reduction of the glucopyranose ring contacts (Figure 3.7), while the catalytic activity is preserved (CDP has previously shown to perform the addition of one GlcN residue to the acceptor substrate using GlcN-1-P as donor).<sup>126</sup> 6F-Glc-1-P has been recently proven to act as a donor and used for the CDP catalysed enzymatic synthesis of multiply 6F-cellodextrin chains.<sup>152</sup>

### 3.2.3 Structural basis of molecular recognition of acceptor and acceptor-like substrates

#### 3.2.3.1 Detection of binding by STD NMR

We next investigated the molecular recognition of acceptors by CDP using a set of four acceptor/acceptor-like mono-, di- and trisaccharides: D-glucose, D-cellobiose, D-laminaribiose and D-celotriose. The STD NMR experiments on D-glucose do not show any STD signal (Figure 3.9), indicating inefficiency of binding in the STD timescale.



**Figure 3.9:** STD NMR off-resonance and difference (off-resonance – on-resonance) spectra of D-glucose in [D<sub>11</sub>]Tris buffer 25 mM, pH 7.4 at 4 seconds saturation time. The protein:ligand ratio for the collected spectra was 1:20. The STD Difference spectra do not show any peaks, indicating

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

lack of contacts between D-glucose and CDP. The peak labelled with the star in both the off- and on-resonance spectra belongs to the residual Tris-CH<sub>2</sub> signal and gives an STD% of 1.25% at saturation time 4 seconds.

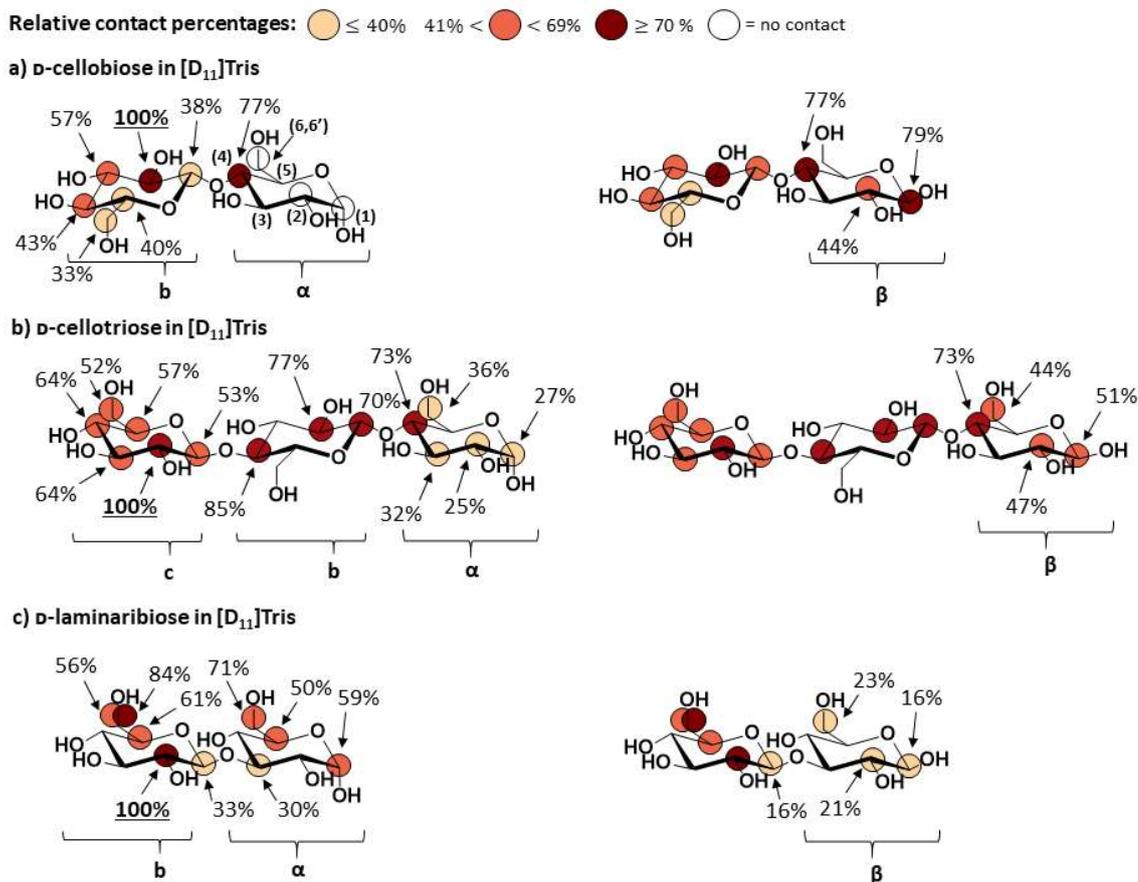
The experimental group epitope mappings of D-cellobiose, D-celotriose and D-laminaribiose for their interactions with CDP are shown in Figure 3.10. As explained in Chapter 1 (subsection 1.2.2), the phenomenon of mutarotation of the saccharides reducing ring determines the coexistence of the  $\alpha$ -isomer and  $\beta$ -isomer in solution. As these two species are diastereomeric, they present different chemical environments, and they also have different resonances in the NMR spectra. Importantly, characteristic resonances are commonly observed for each proton of a reducing monosaccharide correlated to a specific anomeric configuration. In addition, at high magnetic field the effect of the anomeric centre can spread up to the resonances of the non-reducing residues.<sup>153</sup> We were able to integrate isolated resonances for the  $\alpha$ - and  $\beta$ -anomeric spin systems of the reducing ring sugars (H1 $\alpha$ , H1 $\beta$ , H2 $\alpha$ , H2 $\beta$  and H6 $\alpha$  for D-cellobiose and D-celotriose, H1 $\alpha$ , H1 $\beta$ , H2 $\beta$ , H3 $\alpha$ , H5 $\alpha$ , H6 $\alpha$  and H6' $\alpha$  for D-laminaribiose reducing ring) as well as H1b/ $\alpha$ , H1b/ $\beta$  for the non-reducing glucose ring of D-laminaribiose.

In order to quantify the STD intensities of ligands with both anomers present in solution accurately, we took into account the different concentrations of  $\alpha$ - and  $\beta$ -anomers in equilibrium in solution, by application of a correction factor to STD intensity of each  $\alpha$ - and  $\beta$ - anomer. The correction factor was calculated as the ratio of integrals of the peaks due to the  $\alpha$ - and  $\beta$ - anomeric protons relative to the integral of the H1 signal of the non-reducing ring for D-cellobiose, and the H2 signal of the non-reducing ring for D-celotriose, and D-laminaribiose (Table 3.8).

**Table 3.8:** Ratios of integrals of the peaks of  $\alpha$ - and  $\beta$ - anomeric protons

	$\alpha$	$\beta$
<i>D-cellobiose</i>	0.15	0.85
<i>D-celotriose</i>	0.4	0.6
<i>D-laminaribiose</i>	0.45	0.55

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

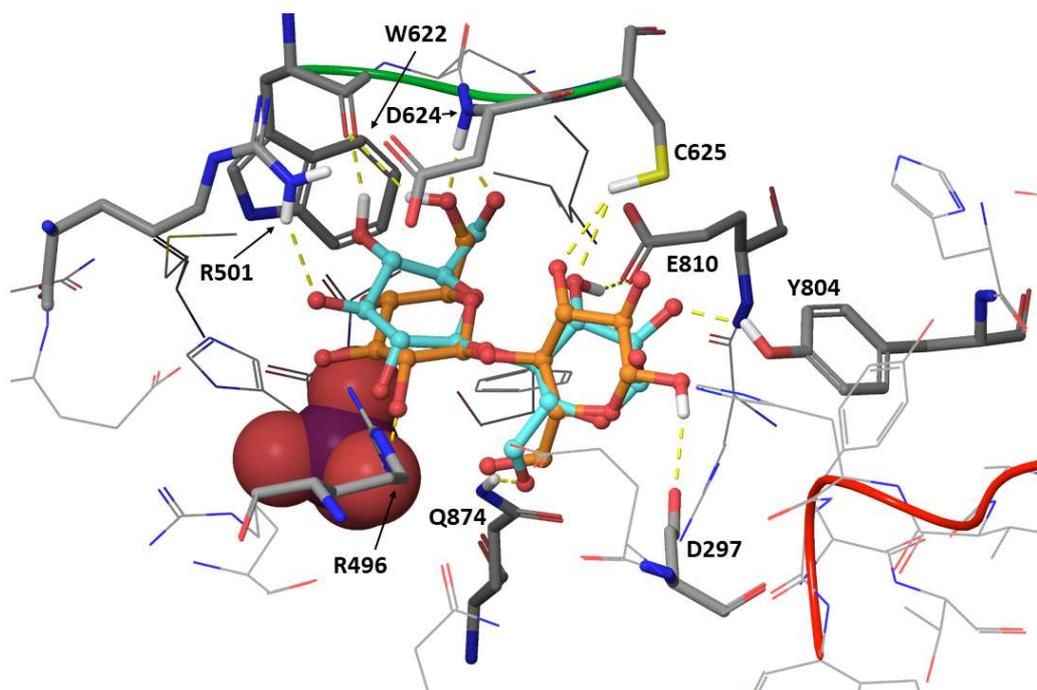


**Figure 3.10:** Group epitope mappings from STD NMR for the interactions with CDP of a) D-cellobiose, b) D-cellobiose and c) D-laminaribiose. Numbering of the glucopyranose atom positions are reported in D-cellobiose reducing ring inside brackets and in bold, while the curly brackets indicate the ring labels. Relative percentages report on the proximity of different regions of the ligands to the surface of the enzyme in the bound state (higher values corresponding to shorter ligand-enzyme distances). Except for all glucose reducing rings and H1 of the non-reducing ring of D-laminaribiose, the STD values reported are average values for both  $\alpha$ - and  $\beta$ -anomeric forms of the oligosaccharides. For accuracy, we considered only well resolved and isolated NMR resonances (for example, the reducing ring anomeric protons). The contacts of overlapping signals are not reported. Isolated resonances for which no STD signal was detected are represented within empty circles. The COSY, TOCSY, NOESY and HSQC spectra recorded for peaks assignment (Figures A.11 and A.13), the STD build-up curves with the monoexponential fit (Figures A.10, A.12 and A.14) and the derived  $STD_{max}$ ,  $k_{sat}$ ,  $STD_0$  and normalised percentages (Tables A.5, A.6 and A.7) are reported in the Appendix.

### 3.2.3.2 *Molecular recognition of the cognate acceptor substrate D-cellobiose*

D-Cellobiose interacts with CDP making closer contacts at the non-reducing glucose ring. The group epitope mapping is in excellent agreement with the non-reducing ring being the key acceptor element for the direct phosphorylase reaction (and the residue to be cleaved). The closest contact of D-cellobiose with the enzyme is at H2 at that non-reducing ring (Figure 3.10, a), while reduced contacts are observed progressively from H3 to H6, with the hydroxymethyl group showing the lowest saturation transfer of the whole disaccharide. These results agree with the inability of CDP to tolerate modifications at C2 position of the non-reducing sugar ring of the acceptor, as it shows a lack of turnover for mannotriose.<sup>126</sup> Further, in contrast to the observations at the glucose ring on Glc-1-P, reduced contacts are observed for D-cellobiose at the exocyclic hydroxymethyl group at C5, explaining the ability of CDP to recognise xylose-derivatives as acceptors.<sup>130</sup> At the (1→4)- $\beta$ -glycosidic linkage, significantly larger saturation transfer was observed for H4 at the reducing glucose ring in comparison to the non-reducing H1. Interestingly, the  $\beta$ -anomer of the disaccharide received significantly larger amount of saturation at the reducing D-glucose ring (Figure 3.10, a), with a group epitope mapping revealing a more intimate contact to CDP in comparison to the  $\alpha$ -anomer. Indeed, H1 $\beta$  presented the second closest contact of D-cellobiose with CDP, whereas most of the protons of the  $\alpha$ -anomer did not show any contacts at all (Figure 3.10, a).

To rationalise the observed differences between the  $\alpha$ - and  $\beta$ -spin anomers of D-cellobiose in structural terms, we ran molecular docking calculations in the presence of inorganic phosphate (Figure 3.11). The best scored poses were in agreement with the STD NMR data, predicting closer contacts at the non-reducing glucose ring. Comparing the poses for  $\alpha$ - and  $\beta$ -anomers indeed indicated differences in the  $\psi$  torsional angle at the inter-glycosidic linkage (Table 3.9) that explained very well the NMR experimental observations. Only in the case of the  $\beta$ -anomer of D-cellobiose the anomeric hydroxyl group establishes a hydrogen-bond with the side chain of Asp297 (Figure 3.11), which drives the H1 $\beta$  proton closer to the surface of the enzyme. This is in excellent agreement with the reported saturation transfer to H1 $\beta$  proton, and the negligible saturation transfer to H1 $\alpha$  proton.



**Figure 3.11:** Representation of the best scored poses (induced fit docking calculations, IFDs) for  $\alpha$  (light blue) and  $\beta$  (orange) anomers of D-cellobiose in the CDP binding pocket. The amino acid residues establishing interactions with the ligands are represented as thick tubes and labelled, the other are represented as wire. The two D-cellobiose are represented as ball-and-stick. Inorganic phosphate is painted in pink and represented as CPK.

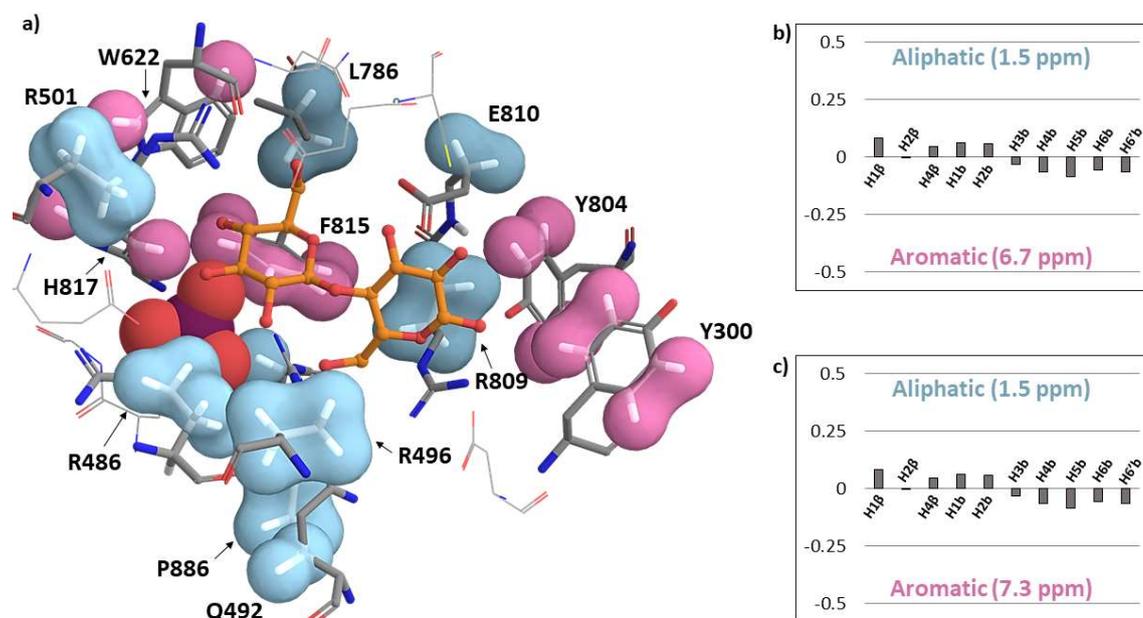
**Table 3.9:**  $\phi$  and  $\psi$  torsional angles for D-cellobiose anomers, indicating that the two ligands have a different inter-glycosidic conformation, as reported by the different  $\psi$  angles. Nonetheless, both the conformation were in the allowed region of the energy map, as reported by GlycoMapsDB (Glycosciences.de).<sup>154</sup>

	$\phi$	$\psi$
$\alpha$ -anomer	44.5	-27.8
$\beta$ -anomer	45.3	-3.7

We attempted to obtain further experimental details of the orientation of D-cellobiose in CDP binding pocket by application of Differential Epitope mapping (DEEP) STD NMR methodology.<sup>155</sup> This approach allows to derive information of the orientation of the bound ligand by reporting on the nature of the protein residues (aliphatic, polar, apolar, aromatic) in close contact with the

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

ligand.<sup>155</sup> The predicted protein protons chemical shifts were obtained by using ShiftX2 (<http://shifftx2.ca>)<sup>98</sup> and the available protein 3D structure (PDB: 5NZ8).<sup>126</sup> Based on the chemical shift reported for residues within 5 Å from the ligand (Appendix, Table A.9 and Table A.10), we selected three frequencies (1.5, 6.7 and 7.3 ppm) in order to selectively irradiate aliphatic and aromatic residues, highlighted in light blue (residues that resonate at 1.5 ppm) and pink (residues that resonate at 6.7 and 7.3 ppm), respectively. To prevent the effects of spin diffusion and reduction in the accumulation of magnetisation, a saturation time of 0.5 seconds was used. For a better visualisation of substrate-residues contacts, Schrodinger Maestro 11 version 20164 was employed for the molecular graphics generation. The protonated residues were selected within 5 Å from  $\beta$ -D-cellobiose structured obtained by the IFD calculations (Figure 3.12, a).



**Figure 3.12:** a) Representation of aliphatic (light blue) and aromatic (pink) residues in the -1, +1 and +2 subsites of CDP binding pocket, selectively irradiated at 1.5, 6.7 and 7.3 ppm, respectively (the representation covers residues up to 5 Å apart from the substrate). The substrate used for the representation is  $\beta$ -D-cellobiose from the IFD calculations, represented as ball-and-stick in orange. The irradiated protons are highlighted thanks to the construction of a molecular surface (Maestro), whereas non-irradiated residues are shown as wire. The cofactor (inorganic phosphate) is represented as CPK in purple; b)  $\Delta$ STD factor at 1.5/6.7 ppm and c)  $\Delta$ STD factor at 1.5/7.3 ppm calculated in  $[D_{11}]$ Tris buffer.

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

It was expected to obtain positive and negative  $\Delta$ STD factors reporting on STD increases when irradiating aliphatic and aromatic residues, respectively. Unfortunately, the very large molecular weight of the protein, 114.364 kDa per monomer, 228.728 kDa for the dimer (active form of CDP in solution)<sup>126</sup> resulted in a loss in the selectivity of the irradiation due to fast intramolecular magnetisation spreading among close residues driven by spin diffusion. This prevented us to obtain information about any contact with aliphatic and aromatic “patches”, as demonstrated by the very low  $\Delta$ STD factors measured for each isolated proton of D-cellobiose in [D<sub>11</sub>]Tris buffer, 25 mM for the selected irradiation frequencies (Figure 3.12, b and c; see also Appendix, Table A.11 and Table A.12).

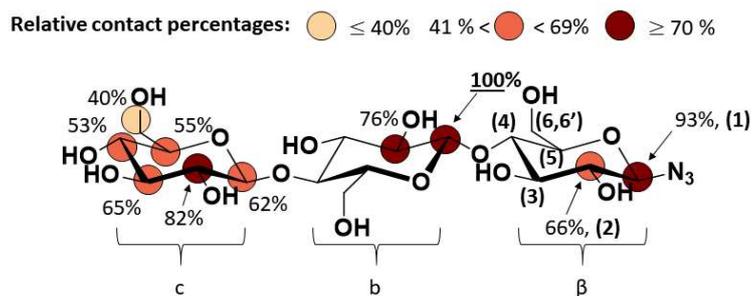
#### 3.2.3.3 *Molecular recognition of cognate acceptor substrate D-celotriose and the non-cognate acceptor-like substrate D-celotriosyl-azide*

Next, we studied the binding of a longer CDP acceptor-like substrate, the trisaccharide D-celotriose, to investigate interactions on the +2 subsite, further from the inner side of the catalytic cleft (-1 and +1 subsites). Its group epitope mapping (Figure 3.10, b) showed some similarities with that of D-cellobiose  $\beta$ -anomeric spin system, with the closest contact on H2 of the non-reducing ring and lower saturation transfer to the other protons of the ring. The central ring overall showed intimate contacts with the enzyme surface, with H4 displaying the second closest contact in the whole ligand, with higher saturation transfer than H1 at the non-reducing ring, as in the case of D-cellobiose. Different from D-cellobiose, however, D-celotriose showed contacts for both the reducing  $\alpha$ - and  $\beta$ -glucose rings, with the  $\beta$ -anomer showing a more intimate contact with CDP in the bound state.

To prove the CDP selectivity towards the  $\beta$ -anomeric configuration, we synthesised the D-celotriose derivative D-celotriosyl-azide, in which the introduction of an azide group in position C1 via an intermolecular nucleophilic attack mediated by 2-chloro-1,3-dimethylimidazolinium chloride (DMC) in aqueous solution fixes the anomeric carbon in the  $\beta$ -configuration. The choice of these acceptor-like substrate was driven by the previously described ability of CDP to recognise and promote elongation of  $\beta$ -glucosyl azide primers.<sup>132</sup> The conversion of D-celotriose into celotriosyl-azide was performed following the one-step conversion described by Tanaka *et al.*,<sup>156</sup> the progression of the reaction was monitored via TLC, and the formation of the product was

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

confirmed by NMR spectroscopy (Appendix A.15). The group epitope mapping is in excellent agreement with the ones described above for D-cellobiose and D-celotriose (Figure 3.13). The closest contact of D-celotriosyl-azide is at H1 of the central ring, while the  $\beta$ -anomeric proton presents the second fastest building-up curve. These results reinforce the CDP  $\beta$ -anomeric selectivity described for D-cellobiose and D-celotriose



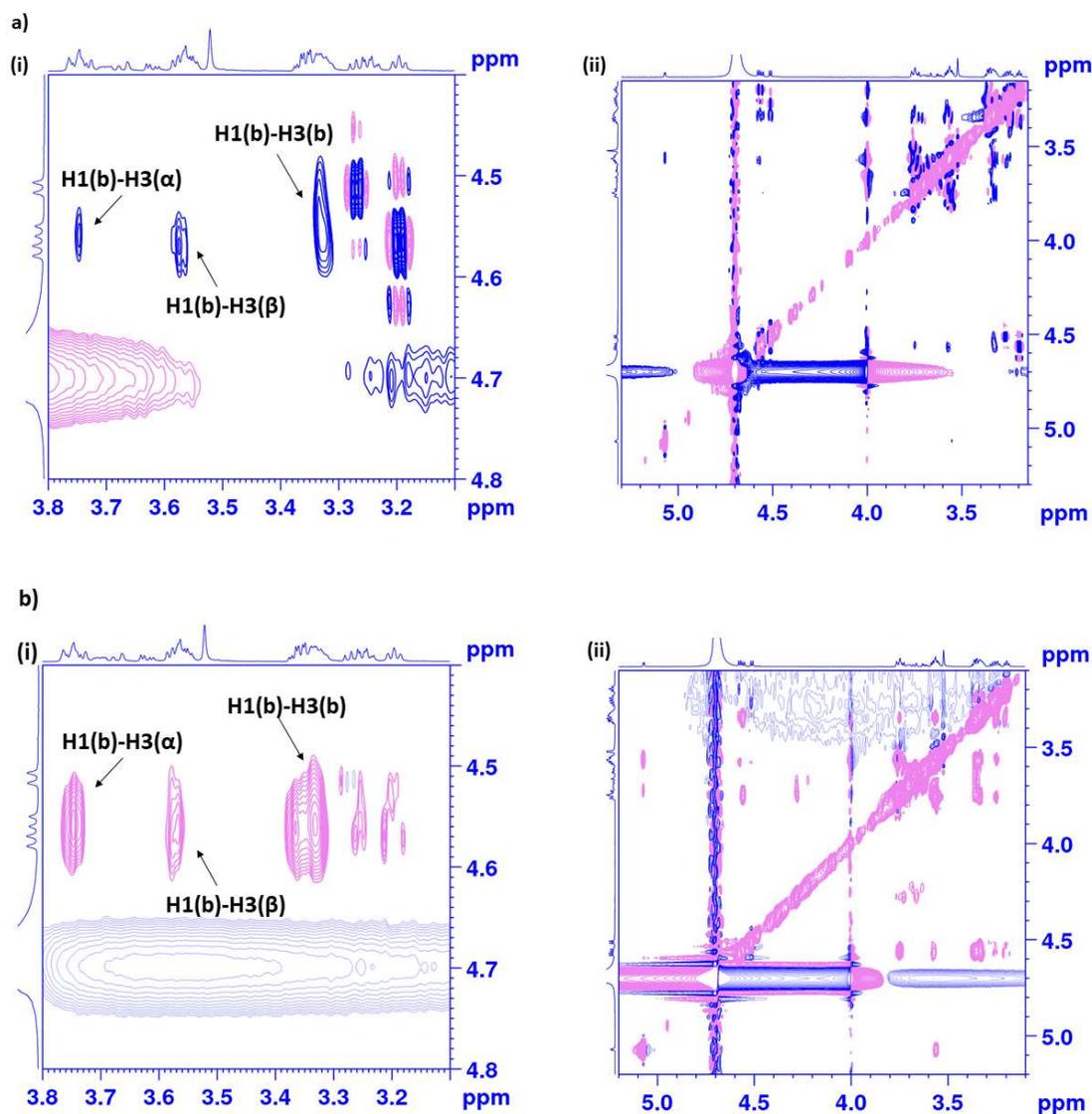
**Figure 3.13:** Binding epitope mapping of D-celotriosyl-azide in [D<sub>11</sub>]Tris buffer 25 mM, pH 7.4, employing 2 mM of ligand for 10  $\mu$ M in binding unit. The STD build-up curves with the monoexponential fit (Figure A.16) and the derived  $STD_{max}$ ,  $k_{sat}$ ,  $STD_0$  and normalised percentages (Table A.8) are reported in the Appendix.

#### 3.2.3.4 Molecular recognition of the non-cognate acceptor-like substrate: D-laminaribiose

To explore the influence of the regiochemistry of the glycosidic bond configuration on CDP binding, we compared the binding of D-cellobiose with its regioisomer D-laminaribiose (D-glucose- $\beta$ -(1,3)-D-glucose). The STD NMR results are shown in Figure 3.10, c. Similarly to D-cellobiose, H2 of the non-reducing ring received the highest saturation transfer. However, in contrast to D-cellobiose and D-celotriose, in this case, the closest contact at the reducing ring was observed for the  $\alpha$ -anomer. Additionally, for D-laminaribiose, good spectra resolution allowed us to detect also the impact of the anomeric configuration up to the non-reducing ring signals, where H1b of  $\alpha$ -D-laminaribiose showed significantly higher saturation transfer compared to H1b of  $\beta$ -D-laminaribiose. Hence, our results indicate that for the molecular recognition of regioisomers with sequence D-glucose- $\beta$ -(1-X)-D-glucose, CDP shows a preferential molecular recognition of  $\alpha$ -anomeric disaccharides showing a  $\beta$ -(1-3) inter-glycosidic regiochemistry (D-laminaribiose), whereas this preference changes towards  $\beta$ -disaccharides when the inter-glycosidic regiochemistry is  $\beta$ -(1-4) (D-cellobiose).

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

The observed differences in binding group epitope mappings depending on the reducing sugar ring configuration prompted us to characterise the bound conformations of both anomers to explore whether those epitope differences are also concomitant with conformational differences upon binding to CDP. To that aim, we carried out Transfer-NOESY experiments (*tr*-NOESY) on a sample containing a 1:10 CDP: $\alpha$ -laminaribiose ratio (Figure 3.14).  $\alpha$ -laminaribiose bound NOEs were compared to those in the free state (in the absence of the protein) to probe conformational changes due to binding. The focus was on the inter-glycosidic NOEs, and quantitative analysis of H1b-H3 $_{\alpha}$  and H1b-H3 $_{\beta}$  NOEs was performed and key  $^1\text{H}$ - $^1\text{H}$  distances of the disaccharides in the bound state were derived (Table 3.10). As a result, no significant changes in the  $^1\text{H}$ - $^1\text{H}$  inter-glycosidic distances were observed, indicating that the observed differences in binding group epitope mappings for  $\alpha$ - and  $\beta$ -anomers are not correlated with a conformational change upon binding to CDP. This result supports that, in contrast to  $\alpha$ -cellobiose, its  $\beta$ -(1-3) regioisomer,  $\alpha$ -laminaribiose, brings closer the surface of the protein to the reducing glucose ring in the case of the  $\alpha$ -anomer. The absence of any significant perturbation of the inter-glycosidic linkage conformation for  $\alpha$ -laminaribiose indicates a distinct orientation of the reducing sugar ring, imposed by the differences in inter-glycosidic linkage stereochemistry.



**Figure 3.14:** (i) Expansion and (ii) full spectra of the transferred 2D-NOESY spectra of a) free D-laminaribiose and b) D-laminaribiose bound to CDP (1:10 protein to ligand ratio) in  $[D_{11}]$ Tris buffer 25 mM pH 7.4, NaCl 100 mM registered at 300 ms mixing time and 290 K in a 800 MHz spectrometer.

**Table 3.10:** Inter-glycosidic  $^1\text{H}$ - $^1\text{H}$  distances ( $\text{\AA}$ ) of  $\text{D}$ -laminaribiose determined from *tr*-NOESY experiments considering the Isolated Spin Pair Approximation<sup>110</sup>; the cross-relaxation rates ( $\sigma_{NOE}$ ) were approximated by the ratio of the normalised NOE volume and the mixing time.

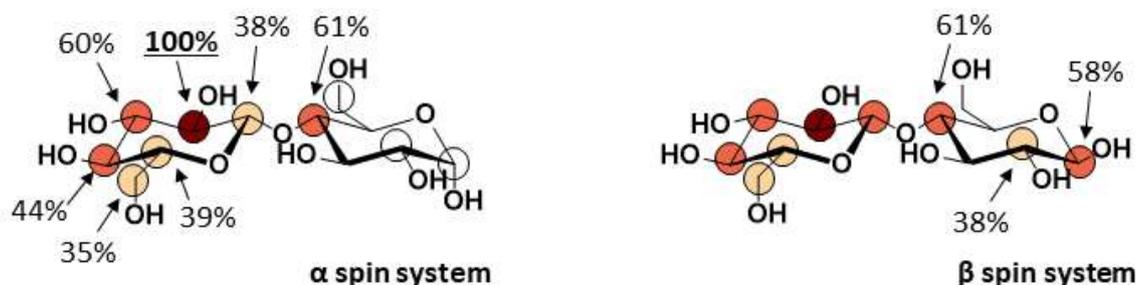
<i>Proton pairs</i>	Free $\text{D}$ -laminaribiose		Bound $\text{D}$ -laminaribiose	
	<i>distance</i> ( $\text{\AA}$ )	$\sigma_{NOE}$	<i>distance</i> ( $\text{\AA}$ )	$\sigma_{NOE}$
<b>H1(b)-H3(<math>\alpha</math>)</b>	2.94	0.02	2.92	-0.08
<b>H1(b)-H3(<math>\beta</math>)</b>	3.20	0.01	3.26	-0.04
<b>H1(b)-H3(b)</b>	2.66	0.04	2.66	-0.13

### 3.2.4 Impact of phosphate on the binding of acceptors

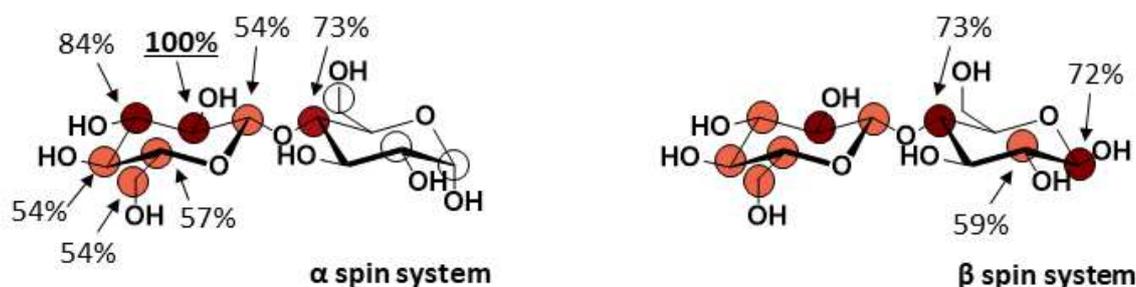
Following on our effort to understand substrate recognition by CDP in detail, we explored the effect of inorganic phosphate on the binding of the acceptor ligands, as this anion is a key player in the phosphorylase reaction. We first carried out STD NMR experiments on CDP/ $\text{D}$ -cellobiose samples after addition of 100  $\mu\text{M}$  phosphate ( $\text{K}_3\text{PO}_4$ , Figure 3.15, a). Furthermore, STD NMR experiments were also carried out with a phosphate excess large enough to ensure saturation of the CDP binding pocket, using 25 mM PBS pH 7.4 buffer (Figure 3.15, b). As previously mentioned in section 3.3.2, a correction factor was applied to the STD signal of each  $\alpha$ - and  $\beta$ -associated peaks to account on the  $\alpha$ - and  $\beta$ -anomers difference in concentration. The correction factor was calculated as ratio of area of the  $\alpha$ - and  $\beta$ -anomeric protons relative to the area of H1 of the non-reducing ring for  $\text{D}$ -cellobiose for each concentration of inorganic phosphate in the reaction mixture (Table 3.11).

Relative contact percentages:   $\leq 40\%$    $41\% <$    $< 69\%$    $\geq 70\%$   = no contact

a)  $\beta$ -D-cellobiose in  $[D_{11}]$ Tris +  $K_3PO_4$  100  $\mu$ M



b)  $\beta$ -D-cellobiose in PBS



**Figure 3.15:** Effect of phosphate on the binding epitope of  $\beta$ -D-cellobiose for its interaction with CDP. Group epitope mappings derived from the initial slope approach for each isolated proton.<sup>81</sup> a) sample in 25 mM  $[D_{11}]$ Tris pH 7.4, 100 mM NaCl in the presence of 100  $\mu$ M  $K_3PO_4$ , and b) sample in 25 mM PBS pH 7.4, 100 mM NaCl. The maximum  $STD_0$  was observed for H2 in the non-reducing ring, to which an arbitrary value of 100% was assigned. The  $STD$  build-up curves with the monoexponential fit (Figure A.17 and A.18) and The derived  $STD_{max}$ ,  $k_{sat}$ ,  $STD_0$  and normalised percentages (Table A.13 and A.14) are reported in the Appendix.

**Table 3.11:** Population of the  $\alpha$ - and  $\beta$ - anomeric spin system relative to the H1 of the nonreducing ring (integrated 1) calculated for each experimental condition.

	$\alpha$	$\beta$
<i>D-cellobiose</i>	0.15	0.85
<i>D-cellobiose + <math>K_3PO_4</math> 100 <math>\mu</math>M</i>	0.4	0.6
<i>D-cellobiose in PBS 25 mM</i>	0.3	0.7

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

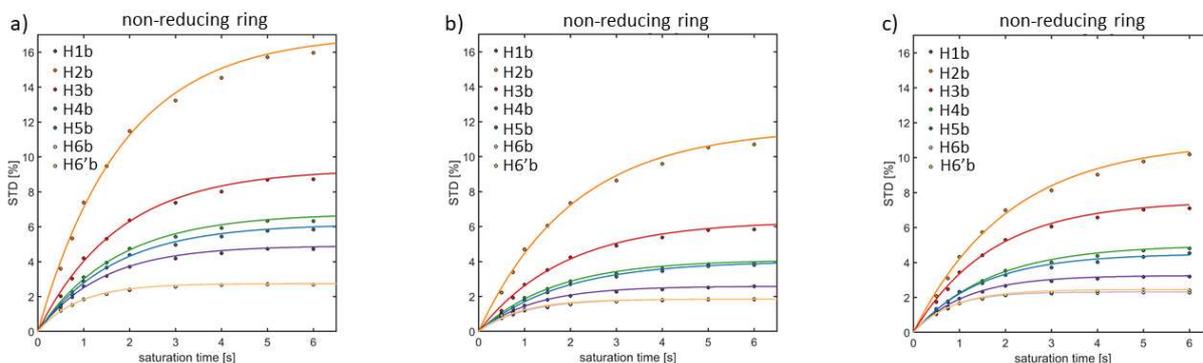
Interestingly, addition of phosphate led to some changes in the group epitope mapping at the non-reducing ring. The STD increases at protons H3, H4 and H5, particularly at high phosphate excess. This is also in excellent agreement with the 3D models for the binding of D-cellobiose anomers to CDP (Figure 3.11), generated in the presence of phosphate. In these models, hydroxyl groups at C2, C3 and C4 sit on top of the negatively charged anion, which brings H3, H4 and H5 closer to the CDP binding surface, explaining their increase in relative STD values. Globally, the group epitope mapping of D-cellobiose did not change significantly upon phosphate titration (cf. Figures 3.10 and 3.15) indicating that the presence of phosphate does not affect the acceptor substrate binding mode further than getting the C2-C3-C4 of the non-reducing ring a bit closer to the surface of the binding pocket.

Notably, addition of phosphate led to a significant decrease in absolute STD NMR intensities of D-cellobiose (Figure 3.16) which seemed to equilibrate upon saturation of CDP binding site with phosphate in the 25 mM PBS pH 7.4 sample. This flattening effect could provide information on the kinetics of binding. Indeed, having used the same irradiation frequency for the different reaction mixtures, and having demonstrated no changes in the substrate binding mode, we can consider the substrate residence time in the bound state as a variable factor. Hence, when [D<sub>11</sub>]Tris buffer is used, the fast kinetics of exchange and therefore the short bound residence time prevent the magnetisation to spread evenly from H2, the proton in very close contact with the protein, and the ligand's protons further from the protein surface. Hence, the faster the kinetics of interaction, the larger the gap between the STD value of H2 and H3 (the second strongest STD signal in the non-reducing ring spin system). On the contrary, when PBS buffer was used, the gap between the STD values of those two protons was reduced considerably.

Analogy with thermal diffusion process can help us explain how this gap relates to the kinetics of the interaction. Let us imagine a metal bar which is brought close to a source of heat. The part of the bar closer to the heat source will be the first part to get hot, and the hottest one at any point. However, the longer the bar will remain close to the heat source, the further the heat will diffuse, and the rest of the bar will gradually get hotter. If we now imagine the magnetisation to be the heat, the protein surface to be the source of heat, and the ligand to be bar-like, we can translate the thermal diffusion analogy to explain the gap in STD NMR intensities between close and far protons. We are therefore witnessing a flattening effect of the binding epitope due to longer

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

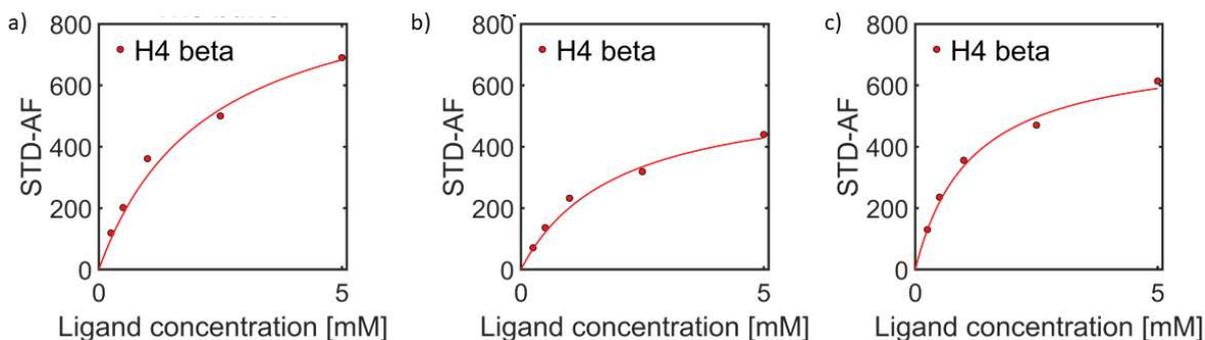
permanence of the ligand into the binding pocket. In addition, this increase on ligand affinity is reflected by the increased relative STD of the protons at the non-reducing ring.



**Figure 3.16:** Effect of phosphate on absolute STD NMR intensities for d-cellobiose in a)  $[D_{11}]$ Tris, 25 mM pH 7.4, NaCl 100 mM, b)  $[D_{11}]$ Tris, 25 mM pH 7.4, NaCl 100 mM with 100  $\mu$ M  $K_3PO_4$  and c) PBS 25 mM, pH 7.4. The STD build-up curves were recorded for each experimental condition at increasing saturation times (from 0.5 to 6 seconds).

To investigate this, we carried out titration experiments with D-cellobiose in the absence and in the presence of phosphate, and determined the apparent dissociation constant ( $K_D^{app}$ ) for D-cellobiose binding to CDP, under the three following experimental conditions: i) absence of inorganic phosphate ( $[D_{11}]$ Tris buffer 25 mM, pH 7.4, NaCl 100 mM) ; ii) 10-fold excess of inorganic phosphate to binding sites, and iii) very large excess of inorganic phosphate (PBS 25 mM pH 7.4, isotonic). As CDP is able to slowly phosphorylate the acceptor in the time scale of tens of hours, we could not follow a full initial slope analysis,<sup>79</sup> so quantification of  $K_D$  was carried out by reducing the time scale of the titration experiments by running the STD NMR spectra at one short saturation time (1 s). On top of that,  $K_D^{app}$  values were determined from the proton on D-cellobiose simultaneously showing: i) the lowest STD absolute intensity, to avoid the impact of differential relaxation times on the  $K_D^{app}$  determination<sup>79</sup>, and ii) the best mathematical fitting to a Langmuir isotherm. Thus, we monitored the titration *via* the STD-AF of H4 $\beta$  (Figure 3.17). Table 3.12 shows the different  $K_D^{app}$  values upon phosphate addition, proving that the co-factor plays a thermodynamics contribution in substrate recognition producing a slight increase in ligand binding affinity. These experiments highlight the high sensitivity of STD NMR intensities to changes in affinities under the conditions tested.

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*



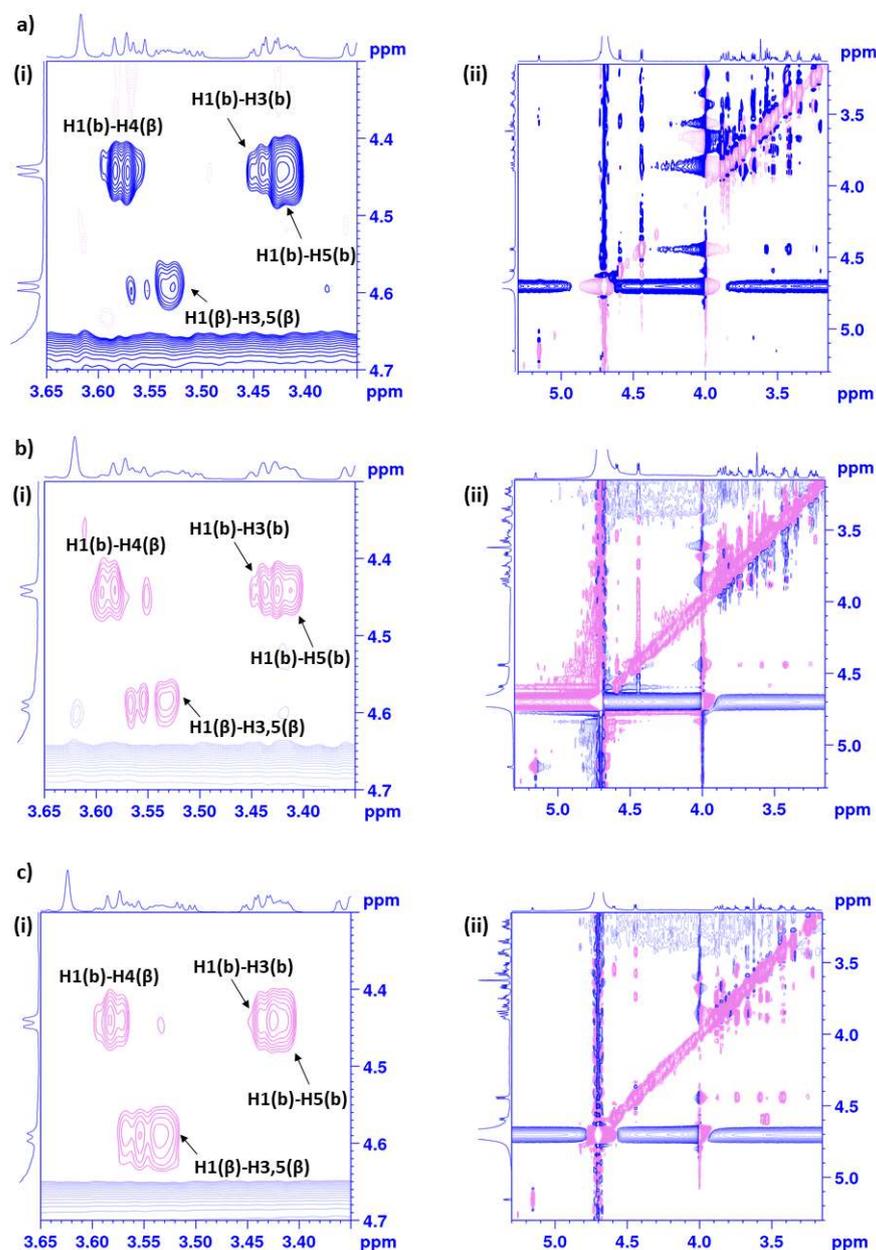
**Figure 3.17:** Binding isotherm fitted to Langmuir equation for D-cellobiose in a) [D<sub>11</sub>]Tris buffer 25 mM pH 7.4, NaCl 100 mM, b) [D<sub>11</sub>]Tris buffer 25 mM pH 7.4, NaCl 100 mM with K<sub>3</sub>PO<sub>4</sub> 100 μM and c) PBS 25 mM, pH 7.4, isotonic.

**Table 3.12:** Apparent equilibrium dissociation constant  $K_D$  [mM] calculated from STD NMR spectroscopic titration as a function of the saturation time ( $t_{sat}$ ). The STD signals from proton H4β was followed for D-cellobiose bound to CDP in [D<sub>11</sub>]Tris buffer 25 mM pH 7.4 NaCl 100 mM, [D<sub>11</sub>]Tris buffer 25 mM pH 7.4 NaCl 100 mM in presence of K<sub>3</sub>PO<sub>4</sub> and PBS buffer 25 mM pH 7.4, isotonic.

	[D <sub>11</sub> ]Tris			[D <sub>11</sub> ]Tris with K <sub>3</sub> PO <sub>4</sub>			PBS		
	K <sub>D</sub> [mM]	95% conf.	R <sup>2</sup>	K <sub>D</sub> [mM]	95% conf.	R <sup>2</sup>	K <sub>D</sub> [mM]	95% conf.	R <sup>2</sup>
<b>H4β</b>	2.23	(±0.71)	0.9972	1.961	(±1.06)	0.9911	1.157	(±0.70)	0.9828

As the inorganic anion did not affect the binding mode of the acceptor ligand yet had a slight effect on the affinity, we decided to investigate whether the presence of phosphate might affect the bioactive conformation of the disaccharide. To that aim, tr-NOESY spectra were measured to characterise the conformation of D-cellobiose around the β-(1-4) linkage in the bound state. Cross-relaxation rates ( $\sigma_{NOE}$ ) were approximated by the ratio of the normalised NOE volume and the mixing time. The Isolated Spin Pair Approximation<sup>110</sup> was used to calculate distances. NOEs were recorded for (i) the free ligand, (ii) the bound ligand (1:20 protein to ligand ratio) and (iii) the bound ligand in the presence of inorganic phosphate (1:20 protein to ligand ratio, 1:5 protein to cofactor ratio) (Figure 3.18). The results (Table 3.13) demonstrate that the inter-glycosidic H1(b)-H4(β) distance, reporting on the conformation around the β-(1-4) linkage,

did not change significantly neither exchanging from the free to the CDP-bound state, nor after addition of inorganic phosphate in the bound state.



**Figure 3.18:** (i) Expansion and (ii) full spectra of the transferred 2D-NOESY NMR spectra for a) free  $D$ -cellobiose, b)  $D$ -cellobiose bound to CDP (1:20 protein to ligand ratio) and c)  $D$ -cellobiose bound to CDP in the presence of inorganic phosphate in  $[D_{11}]$ Tris buffer 25 mM, pH 7.4 NaCl 100 mM recorded at 160 ms mixing time and 298 K in a 800 MHz spectrometer.

**Table 3.13:**  $^1\text{H}$ - $^1\text{H}$  distances (Å) measured *via* tr-NOESY experiments for D-cellobiose free ligand, bound to CDP (20:1 ligand-to-enzyme ratio) and bound to CDP in the presence of 5-fold per enzymatic cleft unit of inorganic phosphate; the cross-relaxation rates ( $\sigma_{\text{NOE}}$ ) were approximated by the ratio of the normalised NOE volume and the mixing time.

Proton pairs	Free D-cellobiose		Bound D-cellobiose		Bound D-cellobiose with phosphate	
	distance (Å)	$\sigma_{\text{NOE}}$	distance (Å)	$\sigma_{\text{NOE}}$	distance (Å)	$\sigma_{\text{NOE}}$
<b>H1(b)-H5(b)</b>	2.38	0.03	2.38	-0.10	2.38	-0.04
<b>H1(b)-H3(b)</b>	3.07	0.01	2.91	-0.03	2.83	-0.01
<b>H1(b)-H4(β)</b>	2.54	0.02	2.44	-0.09	2.38	-0.04
<b>H4(b)-H6(b)</b>	2.92	0.01	3.11	-0.02	n.r	n.r
<b>H4(b)-H6'(b)</b>	3.53	0.003	3.06	-0.02	2.95	-0.01
<b>H2β-H4β</b>	2.52	0.02	2.21	-0.16	2.05	-0.09
<b>H1(b)-H6(β)</b>	n.r.*	n.r.*	2.46	-0.08	2.66	-0.02
<b>H1(b)-H6'(β)</b>	n.r.*	n.r.*	2.62	-0.06	2.46	-0.03

\*n.r.= not recorded

### 3.2.5 Structural details of the CDP-donor-acceptor ternary complex interactions

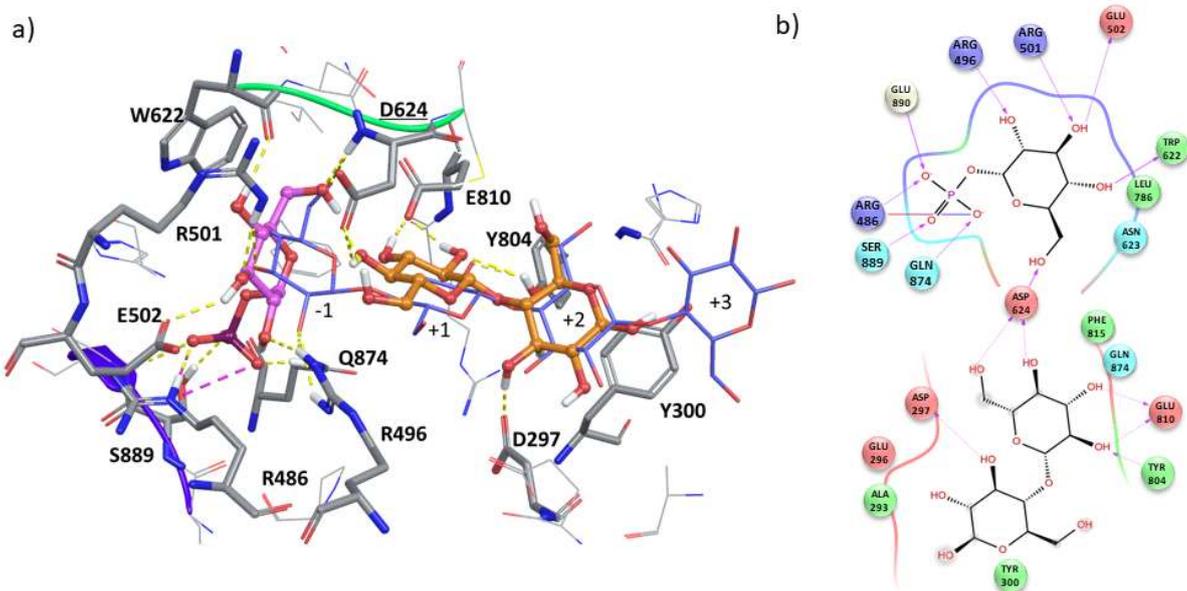
To deepen our understanding of the molecular recognition of substrates taking place during the reverse phosphorylase reaction, we generated a 3D model of the CDP/Glc-1-P/D-cellobiose ternary complex by molecular modelling (protein-ligand docking SP) and analysed the resulting structure on the bases of its correlation with the published CDP-bound D-cellobiose structure.<sup>126</sup> The calculations were performed by docking D-cellobiose to the previously obtained CDP-bound Glc-1-P structure, and the most populated solution is shown in Figure 3.19. To help visualisation of the location and orientation of the donor Glc-1-P and acceptor D-cellobiose substrates within the binding site, we show a superimposition of the published CDP-bound D-cellobiose structure.<sup>126</sup>

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

The main contacts established by the acceptor substrate and the residues in the catalytic cleft are reported in Table 3.14.

The docking model shows that D-cellobiose enters the +1 subsite of the binding pocket with the non-reducing ring, orientation compatible with the reverse phosphorylase mechanism. Figure 3.19, a shows that the residue Asp624 has a bridge function between donor and acceptor substrates. Specifically, the hydroxymethyl group of Glc-1-P acts as H-bond acceptor with the backbone NH of Asp624, while the hydroxyl groups at C4 and C6 act as H-bond donors with Asp624 side chain (Figure 3.19, b).

The hydroxyl groups at C2 and C3 of D-cellobiose non-reducing ring act as H-bond donor with the side chain of Glu810, while the hydroxyl group at C2 acts as H-bond acceptor with the side chain of Tyr804. Additionally, D-cellobiose reducing ring establishes a CH- $\pi$  stacking interaction with Tyr300 in the +2 subsite. The establishment of this CH- $\pi$  stacking interaction is in very good agreement with the STD binding epitope mapping of D-cellobiose, where a more intimate contact of the  $\beta$ -anomer in comparison with the  $\alpha$ -anomer was reported for the reducing ring occupying the +2 subsite. It is expected that a configurational change of the anomeric proton from  $\beta$ - to  $\alpha$ - will determine a disruption of the CH- $\pi$  stacking interaction, causing a reduction in the enthalpic contribution for the binding event. Furthermore, the position of this interaction at the +2 subsite might explain the lower  $K_M^{app}$  of D-cellobiose in comparison with D-cellobiose, as well as the CDP inability to perform the phosphorylase reaction on substrates shorter than D-cellobiose.



**Figure 3.19:** a) Representation of Glc-1-P / D-cellobiose / CDP ternary complex obtained by Docking SP calculations run for all the obtained poses of D-cellobiose conformational search (MacroModel), a 4-fold enhanced conformational sampling and a grid box of 10 Å inner box and 23 Å outer box. The substrates pose is compared with D-cellotetraose structure from O'Neill *et al.*, 2017.<sup>126</sup> The main amino acids residues establishing interactions with the substrates are represented as thick tubes; Glc-1-P and D-cellobiose are represented as ball-and-stick in fuchsia and orange, respectively. b) Ligand interaction diagram reported to facilitate the visualisation of the set of interactions established by Glc-1-P and D-cellobiose with the binding pocket residues in the 1, +1 and +2 subsites. The arrows indicate H-bond going from donor to receiver, dash lines represent H-bond with residues sidechain, whereas the solid line with residues backbone. The solid line shaded from red to blue represents a salt bridge.

**Table 3.14:** Interactions established between D-cellobiose and the catalytic cleft residues in the ternary complex CDP/Glc-1-P/D-cellobiose. The established interactions are specified in brackets nearby the involved group.

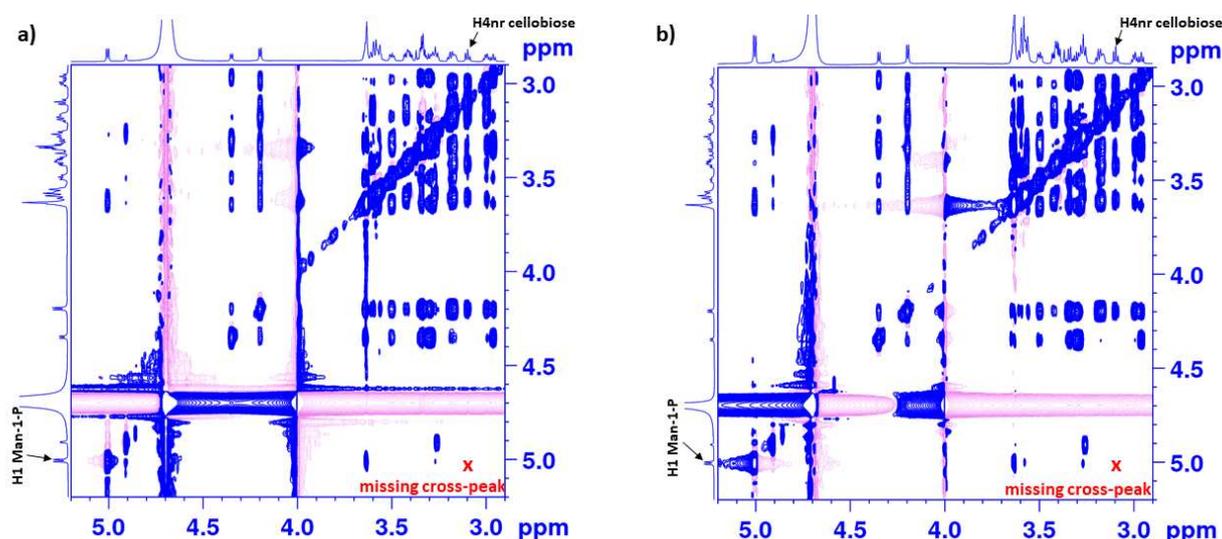
D-cellobiose non-reducing ring	CDP interacting residues
-OH2 (H-bond acceptor)	Tyr804
-OH2 (H-bond donor)	Glu810
-OH3 (H-bond donor)	Glu810
-OH4 (H-bond donor)	Asp624
-CH <sub>2</sub> OH (H-bond donor)	Asp624
<b>D-cellobiose reducing ring</b>	
-OH3 (H-bond donor)	Asp297
-CH (1,3,5) (CH- $\pi$ stacking)	Tyr300

Important structural information can be obtained comparing the 3D docking model of the ternary complex with the available literature on cellobiose phosphorylase from *Cellovibrio gilvus* (CBP) and on CDP from *Clostridium stercoarium*. In the first case, CBP crystal structures (PDB: 3QG0)<sup>157</sup> shows that the enzyme misses a residue able to establish a CH- $\pi$  stacking interaction in the +2 subsite. This observation might explain the ability of CBP to synthesize only disaccharides and no longer oligosaccharides chains, supporting the hypothesis that this residue is pivotal for the productive binding of D-cellobiose and longer saccharides. In the second case, previous studies on CDP from *Clostridium stercoarium*<sup>158</sup> reported on the presence of a TRP residue in the +2 binding subsite level. The construct of a mutant with an ALA residue in place of the TRP resulted in retention of 50% activity on recognising D-cellobiose as acceptor, indicating the contribution of other residues at the +2 subsite in the substrate recognition.<sup>158</sup>

To experimentally validate our 3D docking model of the ternary complex, we resort to Inter-Ligand NOE (ILOE) NMR experiments to probe donor and acceptor proximity by detection of intermolecular NOE between the two CDP-bound substrates.<sup>159</sup> We selected Man-1-P as donor-like

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

substrate as the very quick turnover of D-cellobiose to D-cellobiose in the presence of Glc-1-P (the appearance of peaks belonging to D-cellobiose happens in a matter of seconds) prevented us from using CDP cognate donor substrate in our investigation. Importantly, the previously described binding epitope of Man-1-P obtained with the STD NMR experiments demonstrated that Man-1-P binds CDP in a similar way of Glc-1-P, even though no catalytic efficiency was reported for this donor-like molecule.<sup>126</sup> We performed ILOE experiments for the ternary complex CDP/Man-1-P/D-cellobiose with a fixed enzyme-to-acceptor ratio of 1:50 and two different acceptor-to-donor ratio (1:1 and 1:2, respectively) at 278 K (Figure 3.20). Based on our CDP/Glc-1-P/D-cellobiose ternary complex docking model, we expected to observe a cross-peak between the proton at C1 of Glc-1-P and proton at C4 of D-cellobiose non-reducing ring, for which a 3.3 Å distance was measured. Unfortunately, no inter-ligand cross-peak was detected.



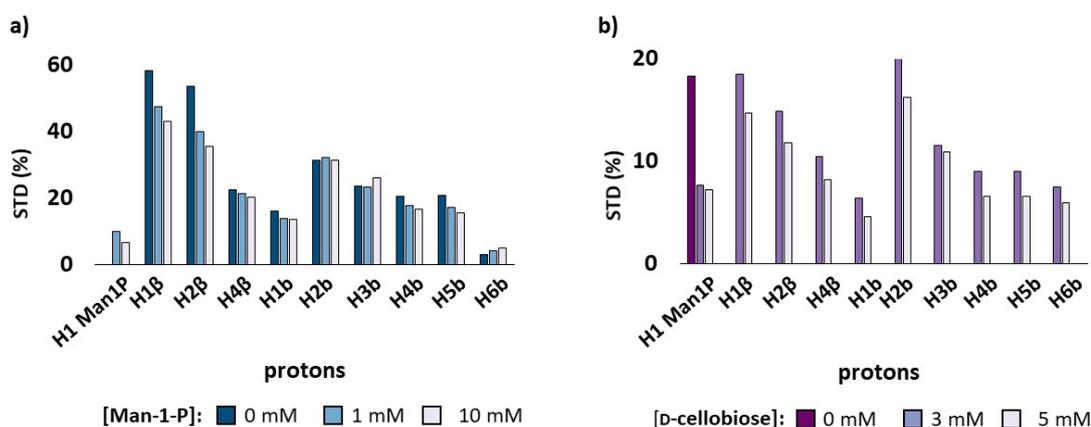
**Figure 3.20:** ILOE NMR spectra of a) D-cellobiose 5 mM, Man-1-P 5 mM, CDP 100 μM and b) D-cellobiose 5 mM, Man-1-P 10 mM, CDP 100 μM in [D<sub>11</sub>]Tris 25 mM, pH 7.4 NaCl 100 mM at 278 K. The experiments were performed using a mixing time of 1.5 s.

The STD NMR competition experiments helped us to explain why the expected cross-peak could not be detected in the ILOE spectra. In the competition experiments, STD NMR spectra are recorded at a single saturation time, and the changes in absolute STD intensity of one species are monitored upon introduction of a competitive binder. At first, we recorded the STD NMR spectra for a fixed concentration of D-cellobiose, and we gradually added Man-1-P on the sample

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

(Figure 3.21, a). The experiment shows that the introduction of Man-1-P does not change the absolute STD intensity recorded for D-cellobiose protons, indicating that D-cellobiose is a stronger binder than Man-1-P and it is not displaced by the presence of the donor molecule.

We then recorded the STD NMR spectra for a fixed Man-1-P concentration, to which we added D-cellobiose on top (Figure 3.21, b). In this case, a drastic reduction of the intensity of H1 of Man-1-P was recorded, even when D-cellobiose was present at lower concentration. Then, increasing D-cellobiose concentration, no significant changes in D-cellobiose protons absolute STD intensity was observed. The data indicate that, upon introduction of D-cellobiose, Man-1-P is fast and completely displaced from the binding pocket.



**Figure 3.21:** a) Results of STD NMR experiments recorded at 2 seconds saturation time for a sample containing D-cellobiose 1 mM to which Man-1-P was gradually added. The experiments were recorded with 100  $\mu$ M CDP binding unit in in  $[D_{11}]$ Tris 25 mM, pH 7.4 NaCl 100 mM at 278 K.; b) Results of STD NMR experiments recorded at 4 seconds saturation time for a sample containing Man-1-P 5 mM to which D-cellobiose was gradually added. The experiments were recorded with 50  $\mu$ M CDP binding unit in  $[D_{11}]$ Tris 25 mM, pH 7.4 NaCl 100 mM at 278 K.

### 3.3 Final discussion and future prospective

CDP is one of the most studied and used biocatalyst belonging to GPs family. Understanding the nature of interactions involved in both natural and non-cognate substrates recognition is important to achieve background knowledge for carbohydrate synthesis, with potential applications in both academic and industrial environments. CDP broad specificity towards

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

non-cognate donor-like and acceptor-like substrates has already been discussed in subsection 3.1.3.1. Nonetheless, the use of non-cognate substrates is not optimised yet, as demonstrated by the substrates reduced enzyme catalytic efficiency and turnover.

The work presented in this chapter is a thorough structural investigation on the molecular recognition processes of donor and acceptor substrates by CDP, conducted by STD NMR spectroscopy and computational tools. The study covers four main points, herein summarised:

- i) CDP molecular recognition of cognate and non-cognate donor substrates (Glc-1-P, Man-1-P, Gal-1-P, GIN-1-P and 6F-Glc-1-P);
- ii) CDP molecular recognition of cognate and non-cognate acceptor substrates (D-glucose, D-cellobiose, D-cellotriase, D-celotriosyl-azide and D-laminaibiose);
- iii) the role played by inorganic phosphate in acceptor substrates recognition;
- iv) the structural characterisation of the ternary complex CDP/donor/acceptor.

Studying donor and donor-like substrates is important for the synthesis of cellodextrins with functional groups introduced along the chains in a regiospecific manner. For example, the synthesis of multiply 6-fluoro-6-deoxy-cellodextrins, decorated with fluorine atom at C6 position, has been recently achieved in our research group, indicating the possibility of using CDP as enzymatic route to synthesise novel and tailored carbohydrate structures.<sup>152</sup> The close contacts between Glc-1-P and CDP, as indicated by the STD NMR group epitope mapping, suggests a restricted specificity of this enzyme towards the donor substrates. Our results indicate that the hydroxymethyl group plays a pivotal role in substrate recognition. Hence, we can foresee that the introduction of a bulky group on C6 would impair an important contact for substrate recognition, leading to a molecule with poor donor ability. The close contact for H6s in 6F-Glc-1-P proves CDP tolerance for a group in position 6 isosteric to OH acting as an H-bond acceptor. In addition, the lower contact demonstrated for the protons in position C2 and C4 might indicate CDP ability to accept functional modification at these positions.

The study of acceptors and acceptor-like substrates binding is interesting for the preparation of various cellodextrins substituted at their reducing end. STD NMR group epitope mapping showed an intimate contact between the sugars non-reducing ring and CDP in all the investigated acceptor substrates and revealed a pivotal role played by position C2 in acceptor substrates recognition.

## Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

In addition, the study revealed an anomeric selectivity of CDP towards the  $\alpha$ - and  $\beta$ -anomeric configuration for  $\beta$ -(1-3) and  $\beta$ -(1-4)-oligosaccharides acceptors, respectively.

The performance of CDP reaction is dictated by the presence of inorganic phosphate in the binding pocket. The co-factor binds to a lobe adjacent to the -1 subsite and it is ionised at physiological pH.<sup>126</sup> Nonetheless, really few information has been gathered so far about the impact that the inorganic phosphate on the substrate binding properties. Our study indicates that the inorganic phosphate plays a thermodynamics contribution in acceptor substrate binding, as demonstrated by the reported reduction of  $K_D$  upon phosphate introduction.

Our study complements the already published apo- and bound- CDP crystal structures and the CDP specificity information available both for donor and acceptor substrates. The structural details at the atomic level achieved for those complexes reveal the molecular basis of the observed kinetics and specificity of CDP towards both cognate and non-cognate donor and acceptor substrates. Together, these studies provide structural details on CDP specific recognition of both single and multiple substrates. These studies are valuable starting point towards the design and engineering of CDP for tolerance towards different non-cognate donors and acceptors, hence expanding GPs applications in carbohydrate synthesis.

### 3.4 Material and methods

#### 3.4.1 CDP expression and purification

A recombinant plasmid (pET15b) containing the CDP gene from *Ruminiclostridium thermocellum* (YM4 strain) was transformed into *E. coli* BL21 (DE3) cells and grown as described previously.<sup>126</sup> Briefly, 1 L of LB medium containing the transformant and carbenicillin (100  $\mu$ g/mL) was incubated at 37 °C with shaking (200 rpm) until OD<sub>600</sub> around 0.6. Heterologous protein expression was induced by adding isopropyl  $\beta$ D1thiogalactopyranoside (IPTG) to a final concentration of 1 mM and incubating for 4 hours at 30 °C with shaking (180 rpm). The cells were harvested by centrifugation (4,000  $\times$  g, 20 min), resuspended in lysis buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, EDTAfree protease inhibitor cocktail tablet, 0.02 mg/mL DNaseI), lysed by cell disruption (30 Kpsi, constant flow) and the supernatant containing the recombinant

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

proteins was separated from cell debris by centrifugation (20,000 × g, 30 min). Proteins were purified at 4 °C using an ÄKTA pure FPLC system (GE Healthcare). The supernatant was loaded to a 5 mL His-Trap™ HP column (GE healthcare) pre-equilibrated with buffer A (50 mM Tris-HCl, pH 8, 50 mM glycine, 5% glycerol, 500 mM NaCl, 20 mM imidazole). The column was washed with buffer A to remove unbound proteins followed by elution of bound proteins with buffer B (50 mM Tris-HCl, pH 8, 50 mM glycine, 5% glycerol, 500 mM NaCl, 500 mM imidazole). Further purification was carried out by gel filtration chromatography (Superdex S200 16/600 column, GE Healthcare) with 20 mM HEPES, pH 7.5, 150 mM NaCl, 1 mL/min. Fractions containing CDP were pooled and concentrated using Amicon Ultra15 Centrifugal Filter (30,000 MW cut off) and the enzyme concentration (5.7 mg/mL) was determined by NanoDrop™ spectrophotometer (Thermo Fisher Scientific, UK). The Histag CDP was stored in aliquots at 80 °C until required.

The exchange of the protein with deuterated solvent ([D<sub>11</sub>]Tris, pH 7.4, 100 mM NaCl) was performed via 50K MW filter, centrifuge 4000 rpm, 4 degree. 5 cycles of 20 minutes each. The final concentration of the protein was measured using a Thermo Scientific NanoDrop UV-Vis spectrophotometer with a A280 method and an extinction coefficient  $\epsilon$  of 117.635 set up. The obtained concentrated protein was then diluted to the desired concentration for NMR analysis using [D<sub>11</sub>]Tris buffer solution.

#### 3.4.2 Nuclear Magnetic Resonance

All the experiments were carried out using an Avance 800.23 MHz on a Bruker Avance III spectrometer equipped with a 5-mmD probe TXI 800 MHz H-C/N-D-05 Z BTO. <sup>1</sup>H and <sup>13</sup>C resonance assignment for the ligands was performed on the bases of 1D <sup>1</sup>H, 2D DQF-COSY, HSQC and NOESY experiments run on the free ligands in [D<sub>11</sub>]Tris, pH 7.4 NaCl 100 mM at a temperature of 5°C.

##### 3.4.2.1 Saturation Transfer Difference (STD) NMR

**Donors:** The Glc-1P sample consisted of 3 mM ligand and 15  $\mu$ M in binding site (CDP is a homodimer) in [D<sub>11</sub>]Tris buffer (25 mM, pH 7.4, NaCl 100 mM), for an enzyme-to-ligand ratio of 1:200. A decreasing amount of number of scans was collected (from 128 to 32) based on the applied saturation time. For the non-cognate donors (Gal-1P, Man-1P, GIN-1P and 6F-Glc1P) we

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

increased the enzyme-to-ligand ratio, hence increasing the fraction of bound ligand ( $f_{LB}$ ) and in turn the STD signal, using 5 mM of ligand and 50  $\mu$ M in binding site. A reverse amount of number of scans was collected (from 256 to 16 according to the saturation time) and a recycle delay of 6.1 s was used.

**Acceptors:** All the samples were prepared using a 200-fold excess of ligand over catalytic unit (3 mM ligand, 15  $\mu$ M enzyme monomer) in  $[D_{11}]$ Tris buffer (25 mM, pH 7.4, NaCl 100 mM). 128 number of scans were acquired for each saturation time, and a relaxation delay of 6.1 seconds was employed. The impact of inorganic phosphate on D-cellobiose was tested by adding on top of the same sample  $K_3PO_4$  solution in  $[D_{11}]$ Tris buffer, for a final concentration of 100  $\mu$ M  $K_3PO_4$ . In addition, STD NMR experiments in PBS (25 mM, pH 7.4, isotonic) were run for D-cellobiose with the same experimental conditions. Finally, STD NMR experiments for D-celotriosyl-azide were run employing 2 mM of ligand for 10  $\mu$ M in binding unit. A reverse amount of number of scans was collected (from 128 to 32) based on the applied saturation time.

STD NMR experiments were acquired at different saturation times (0.5, 0.75, 1, 1.5, 2, 3, 4, 5 and 6 seconds) to determine the initial slope of the STD build-up of each proton involved in binding to CDP. The employed irradiation frequencies were 0.3 ppm and 50 ppm for the *on-resonance* and the *off-resonance* spectra respectively. Cascades of 50 ms Gaussian-shaped pulses at a field strength of 50 Hz were employed, with a delay of 1 ms between successive pulses. The broad protein signals were removed using a 40 ms spinlock (T1 $\rho$ ) filter (stddiff.3).<sup>81</sup>

Following, the obtained curves were fitted to a mono-exponential function

$$STD(t_{sat}) = STD_{max}(1 - e^{-k_{sat} \cdot t_{sat}})$$

From these STD build-up curves, we mapped out the main contacts of the ligands to CDP in the bound state by determining the initial slopes ( $STD_0$ ) of the curves, obtained as the product of the  $STD_{max}$  and  $k_{sat}$  coefficients, and thereafter normalising all the  $STD_0$  values within a given ligand by the highest one, to which an arbitrary value of 100% was assigned. It should be noted that the use of STD initial slopes increases the accuracy of the STD methodology for binding epitope mapping by avoiding the detrimental effects of different relaxation properties of the ligand protons on the determination of the binding epitope. Different mapping scales were used for the binding epitopes of donor and acceptor molecules, as the donor is buried inside the binding pocket and

therefore receives higher saturation transfer. Therefore, we have considered > 90% and >70% of relative STD as close contact for donors and acceptors, respectively.

For accuracy, in the present study we considered only well resolved and isolated NMR resonances for each investigated ligand (for example, H1 $\beta$  of the reducing ring). The contacts of non-isolated protons (overlapping signals) are not reported. On the other hand, the lack of contact of well-resolved protons, which did not show any STD signal, is represented within empty circles (Figure 3.10).

#### 3.4.2.2 Differential Epitope mapping (DEEP-STD)

DEEP-STD NMR experiments were carried out for D-cellobiose in [D<sub>11</sub>]Tris buffer (25 mM, pH 7.4, NaCl 100 mM). The predicted protein protons chemical shifts were obtained by using ShiftX2 (<http://shiftx2.ca>)<sup>84</sup> and the available protein 3D structure (PDB: 5NZ8).<sup>126</sup> Based on the chemical shift reported for residues within 5 Å from the ligand (Table A.X), we selected three irradiation frequencies (1.5 ppm, 6.7 ppm and 7.3 ppm) in order to selectively irradiate aliphatic and aromatic residues, highlighted in light blue and pink, respectively. A total of 512 scans and 8 dummy scans were collected for each experiment, with a saturation time of 0.5 seconds and a relaxation delay of 6 seconds. The absolute STD intensities was measured for each isolated proton of D-cellobiose, and the  $\Delta$ STF factors for each pair of irradiation frequencies (1.5/6.7 ppm and 1.5/7.3 ppm).<sup>83</sup>

$$\Delta_{\text{DEEP-STD}_i} = \frac{STD_{\text{exp1},i}}{STD_{\text{exp2},i}} - \frac{1}{n} \sum_i^n \left( \frac{STD_{\text{ex},i}}{STD_{\text{ex},i}} \right)$$

Schrodinger Maestro 11 version 2016-4 was employed for all the molecular graphics generation. The protonated residues were selected within 5 Å from D-cellobiose structure obtained from docking calculations.

#### 3.4.2.3 Transferred-NOESY (tr-NOESY)

Tr-NOESY experiments were carried out using a phase sensitive pulse programme with gradient pulses in the mixing time and a relaxation delay of 1.5 seconds. D-cellobiose was analysed using [D<sub>11</sub>]Tris buffer 25 mM pH 7.4 NaCl 100 mM and a protein-to-ligand ratio of 1:20 at 298 K. Experiments at different mixing times (40 and 160 ms) were collected for the free and bound state. Finally, the same experiments were collected in the presence of a 5-fold excess of inorganic

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

phosphate, in comparison with a CDP single binding unit. Tr-NOESY experiments were also carried out to probe conformational rearrangement upon binding on D-laminaribiose at 290 K. In this case, a protein-to-ligand ratio of 1:10 was employed. Experiments at different mixing times (40 and 300 ms) were collected in the free and bound state. The experimental distances of the ligand in the bound state were derived using the isolated spin pair approximation (ISPA)<sup>110</sup> approach. First, each cross peak was divided by its corresponding diagonal peak at 40 ms, thus obtaining the normalized NOE volume. Each volume was then divided by the mixing time to get a good approximation of the cross-relaxation rate ( $\sigma_{NOE}$ ). Finally, using the fixed H1-H5 and H1-H3 distances of the non-reducing ring terminal (2.38 Å and 2.66 Å) for D-cellobiose and D-laminaribiose, respectively, the key inter-glycosidic proton-proton distances were calculated according to the expression

$$d_x = d_{ref}(\sigma_{ref}/\sigma_x)^{1/6}$$

where  $d_x$  is the unknown distance to be determined,  $d_{ref}$  is the distance used as reference, and  $\sigma_{ref}$  and  $\sigma_x$  are the cross-relaxation rates of the reference and unknown distances, respectively.

#### 3.4.2.4 $K_D$ measurement

The measurement of the ligand binding affinity was performed by application of STD NMR experiments.<sup>79, 82</sup> 1H STD NMR spectra of D-cellobiose were acquired at different ligand concentration (0.25, 0.5, 1, 2.5, and 5 mM) with saturation times of 1 second. *On-resonance* and *off-resonance* frequencies were 0.3 and 50 ppm, respectively. A total of 128 number of scans were collected for each experiment, and a relaxation delay of 6 seconds was employed. Three different experimental conditions were investigated: 1) total absence of inorganic phosphate in the sample, with [D<sub>11</sub>]Tris 25 mM, pH 7.4 NaCl 100 mM, 2) concentration of inorganic phosphate 10-fold per binding unit, 3) large excess of inorganic phosphate, with PBS 25 mM. To obtain the  $K_D$  values, the obtained Langmuir isotherm was fitted to<sup>79</sup>

$$STD - AF ([L]) = \frac{\alpha_{STD}[L]}{K_D + [L]}$$

### 3.4.2.5 ILOE<sup>159</sup>

ILOE experiments were performed for the ternary complex CDP/Man-1-P/D-cellobiose for two different acceptor-to-donor ratios (1:1 and 1:2, respectively) at 278 K in [D<sub>11</sub>]Tris 25 mM, pH 7.4 NaCl 100 mM. An enzyme-to-acceptor ratio of 1:50 for a final concentration of 100 μM binding unit was used in both experiments. Phase-sensitive <sup>1</sup>H-<sup>1</sup>H NOESY experiments with water suppression with gradients were carried out at 1.5 seconds mixing time with 128 increments in F1.

### 3.4.3 Molecular Docking:

All the reported model calculations were performed within the Schrödinger molecular modelling suite (MAESTRO).<sup>149</sup> The crystal structure of CDP complexed to D-cellobiose (PDB code 5NZ8)<sup>126</sup> was processed using the protein preparation wizard tool. Conformational sampling of α-D-glucose-1-phosphate (Glc-1-P) and D-cellobiose was performed by a conformational search (MacroModel) based on Monte Carlo Multiple Minimum method in order to enhance the samples conformations. In the case of Glc-1-P, the obtained poses (13 in total) were used to run docking SP (Glide) with a receptor grid of 10 Å inner box and 20 Å outer box, a 4-fold enhanced conformational sampling and OPLS3 as force field. In addition, the sampling of ring conformations was forbidden, and the calculation was run in the absence of inorganic phosphate inside the binding pocket. The obtained conformers were clustered by atomic RMSD (RMSD in place) and the most energetically favourable pose of the most populated cluster (which showed a good overlapping with the non-reducing ring of D-cellobiose complexed with CDP) was selected for analysis and further docking calculations. On the other hand, for D-cellobiose a separate conformational search was performed for the α- and β-anomeric configuration. In both cases, the obtained poses were clustered (RMSD in place) and a representative of the most populated cluster was selected for further calculations. Flexible induced fit docking was performed with a 0.8 Å tolerance constrain on the interglycosidic-linkage position referenced to D-cellobiose complexed with CDP. This constraint was introduced as previous attempts reported a distortion of the interglycosidic-linkage conformation, with φ and ψ angles values not allowed in the β-(1-4)- Carbohydrate Ramachandran Plot (CARP) reported by GlycoMapsDB (Glycosciences.de).<sup>154</sup> The obtained poses were clustered (RMSD in place). For the α-anomeric configuration, the most

### Chapter 3 – Cellodextrin phosphorylase from *Clostridium thermocellum*

energetically favourable pose of the most populated cluster was saved, whereas in the case of the  $\beta$ -anomeric configuration we had to discard the first two most populated clusters as the first presented an inverse orientation of the ligand in the binding pocket and the second presented forbidden values of  $\phi$  and  $\psi$  angles. Finally, the selected Glc-1-P (obtained from the first docking stage) was introduced in the CDP structure to dock a second D-cellobiose molecule, allowing us to obtain a ternary CDP/D-cellobiose/Glc-1-P complex. The selection of D-cellobiose as second docked substrate was led by its ability act as acceptor in the reverse phosphorolysis reaction (indeed, D-cellobiose is the natural disaccharide acceptor). In addition, the closer contact demonstrated for the  $\beta$ -anomer by STD NMR experiments, as well as the additional H-bonding reported from docking calculations, has driven us to select this configuration for our studies. Docking SP was run for all the obtained poses of D-cellobiose conformational search, a 4-fold enhanced conformational sampling and a grid box of 10 Å inner box and 23 Å outer box.

## Chapter 4

### Deoxyfluorinated cellodextrin derivatives: structural characterisation

## 4. Deoxyfluorinated cellodextrin derivatives: structural characterisation

Parts of the finding from this Chapter are published as:

de Andrade Peterson, Muñoz-García Juan C., Pergolizzi Giulia, Gabrielli Valeria, Nepogodiev Sergey, Iuga Dinu, Fábíán László, Nigmatullin Rinat, Johns Marcus A., Harniman Robert, Eichhorn Stephen J., Angulo Jesús, Khimyak Yaroslav Z., Field Robert Robert A. - Chemoenzymatic synthesis of fluorinated cellodextrins identifies a new allomorph for cellulose-like materials. *Chemistry - A European Journal*, **2020**. DOI: 10.002/chem.202003604

### 4.1 Introduction

#### 4.1.1 Principal aims of the work

Cellodextrin phosphorylase (CDP) has been demonstrated a powerful biological tool for the synthesis of both cellodextrin and cellodextrin-derivatives (Table 3.2 in *Chapter 3* reports a full list of recognised donor-like and acceptor-like substrates). Enzymatically synthesised cellodextrins are known to have limited water solubility beyond DP9 and to self-assemble into particles resembling the antiparallel cellulose II crystalline packing.<sup>128</sup> The expression and purification of CDP and the mechanistic details of the catalysed reaction have been already discussed in *Chapter 3*. Herein, have prepared and characterised a series of site selectively fluorinated cellodextrins of different degrees of fluorination and substitution patterns by chemoenzymatic synthesis: 2-deoxy-2-fluoro cellodextrin (2F-EpC, 1), 3-deoxy-3-fluoro cellodextrin (3F-EpC, 2) and 6-deoxy-6-fluoro cellodextrin (6F-EpC, 3) and the multifluorinated multiply 6-deoxy-6-fluoro cellodextrin (multi-6F-EpC, 4).

2F-, 3F- and 6F-EpCs represent a series of single site selectively fluorinated cellodextrins, synthesised to assess the structural impact upon addition of one fluorine atom per chain. On the contrary, multi-6F-EpCs presents decoration of fluorine atoms along the whole cellodextrin chain, except for the reducing cellobiose-unit. Scope of multi-6F-EpCs was to probe the potential disruption of the hydrogen bond network of cellulose II.

Thanks to the scientific consortium established within the John Innes Centre, the University of East Anglia and the University of Bristol, we combined microscopy (TEM and AFM), Raman spectroscopy, PXRD and advanced NMR spectroscopy to obtain structural characterisation at different length scales (morphological, long-range and short-range structural information).

The introduction of a single fluorine atom per cellodextrin (2F-EpC, 3F-EpC and 6F-EpC) determines minor changes on the material morphology and crystalline arrangement in comparison with unfunctionalized cellodextrin chains. On the contrary, multi-6-EpC chains showed unprecedented features for cellulose-like materials. In addition, we highlight the importance of combining solution and solid-state NMR approaches to achieve a full assignment of material characterised by different ordered and disordered domains.

### 4.1.2 On the importance of fluorine functionalised cellulose

#### 4.1.2.1 *Effect of fluorination to tailor material properties*

The introduction of a fluorine atom in a molecular structure has been proven to profoundly modify its chemical and physical properties. Fluorinated materials possess higher stability, attributed to the stronger C-C bond in fluoropolymers, which might achieve extremely high values such as 112 kcal mol<sup>-1</sup> in CF-CF species.<sup>160</sup> In addition, C-F chemical bond is the strongest single bond known (*ca.* 450 kJ mol<sup>-1</sup>), and even though highly dipolar, it is marked by a relative non-polarizability (hardness).<sup>161</sup>

The presence of fluorine increases the molecular hydrophobicity and allows tuning of hydrophilic-hydrophobic balance (HLB) of a material. Extensive fluorination determines the creation of the so-called “fluorous phase”, which does not mix with either polar or non-polar hydrogenated phases.<sup>162</sup> As an example, perfluoroalkanes, perfluorodialkyl ethers, and perfluorotrialkyl amines presents limited miscibility with common organic solvents such as toluene, THF, and acetone.<sup>163</sup> Finally, fluorinated compounds exhibit high CO<sub>2</sub> solubility due to the favourable quadrupole-dipole interactions established among the CO<sub>2</sub> quadrupolar moment deriving from its highly electronegative oxygen atoms and the polar fluorocarbons, which facilitates a close physical association (less than 4.3 Å between the centres of mass) between CO<sub>2</sub> and fluorinated groups.<sup>164</sup>

## Chapter 4 – Deoxyfluorinated cellodextrin derivatives

In conclusion, fluorine-organic compounds present a variety of new functional properties such as enhanced hydrophobicity, high lipophobicity in per-fluorinated substances, high thermal and oxidative stability, chemical and biological inertness, high gas dissolving capacity, weak intermolecular interactions and low surface energy and surface tension.<sup>165</sup>

Despite its high natural abundance, fluorine rarely occurs in natural organic molecules.<sup>166</sup> The reason of this scarcity of fluorometabolites in natural products derives from the low abundance and hence low bioavailability of fluoride ions in the oceans in comparison with chloride and bromide. In addition, fluoride presents the highest heat of hydration (*ca.* 120 kcal mol<sup>-1</sup>). Therefore, enzymes had to evolve a desolvation strategy to being able to perform nucleophilic catalysis with fluorine atoms. Furthermore, the high electronegativity of fluorine (in the Pauling scale, F = 4.0) goes against its oxidation and the generation of the fluoride ion ( $F^-$ ), limiting the evolution of fluorine biochemistry.<sup>167</sup>

### 4.1.2.2 Fluorine-functionalised cellulose derivatives

Fluorination has been used to control interfacial properties of cellulose, aiming to reduce its surface hydrophilicity and produce a superhydrophobic polymer. Indeed, cellulose fluorination is expected to increase water resistance, improve thermal and anti-oxidative stability, and enhance compatibility with synthetic polymers via low surface energy. To date, etherification and esterification of cellulose surface with fluorine derivatives have been reported.<sup>168-170</sup> In addition, Kasuya et al. reported the regioselective substitution of the hydroxyl group in C6 and formation of the deoxy-fluorine cellulose derivative obtained by reaction of cellulose 2,3-dibenzoate and cellulose 2,3-diacetate with diethylaminosulphur trifluoride.<sup>171</sup>

### 4.1.2.3 Biocatalysis of organo-fluorine compounds

To promote selective fluorination and reduce costs of laborious chemical synthesis, many research groups are investigating biosynthetic pathways. Biocatalysis of organo-fluorine compounds can be performed i) by extending the scope of existing enzymes for the formation of C-F bond or ii) by expanding the biosynthetic pathways of different enzymes to accept fluorinated precursors.

In the first instance, fluorinases (FLAs) are so far the only class of enzymes known to directly introduce fluorine into organic molecules (creation of a C-F bond). In 2002 O'Hagan *et al.*<sup>172</sup> isolated

## Chapter 4 – Deoxyfluorinated cellodextrin derivatives

the first-ever natural FLA from *Streptomyces cattleya*, while from 2014 to 2016 four additional FLAs have been identified from *Streptomyces* sp. MA37, *Nocardia brasiliensis*, *Actinoplanes* sp.<sup>173</sup> and *Streptomyces xinghaiensis*.<sup>174, 175</sup>

In the second instance, a variety of enzymes have been proven to accept fluorine-functionalised substrates. Within the framework of this thesis, it is important to highlight the existence of kinases capable of recognising fluorinated monosaccharides and producing  $\alpha$ -D-sugars-1-phosphates.<sup>176</sup> Notably, sugars-1-phosphate do not just have a pivotal role in sugars metabolism and utilisation, but they act as donors in the reverse phosphorolysis reaction catalysed by phosphorylases, central point of this dissertation. Several phosphorylases showed specificity towards various fluorinated donors and/or acceptors substrates. As examples,  $\alpha$ -1,4-d-Glucan phosphorylases and  $\alpha$ -1,4-Glucan:maltose-1-phosphate maltosyltransferase demonstrated to accept fluorinated derivatives as donor,<sup>22</sup> while D-galactosyl- $\beta$ 1–3-N-acetyl-D-hexosamine phosphorylase accepted both fluorinated donor and acceptor derivatives.<sup>177,178, 179</sup>

### 4.1.2.4 Fluorine in NMR spectroscopy

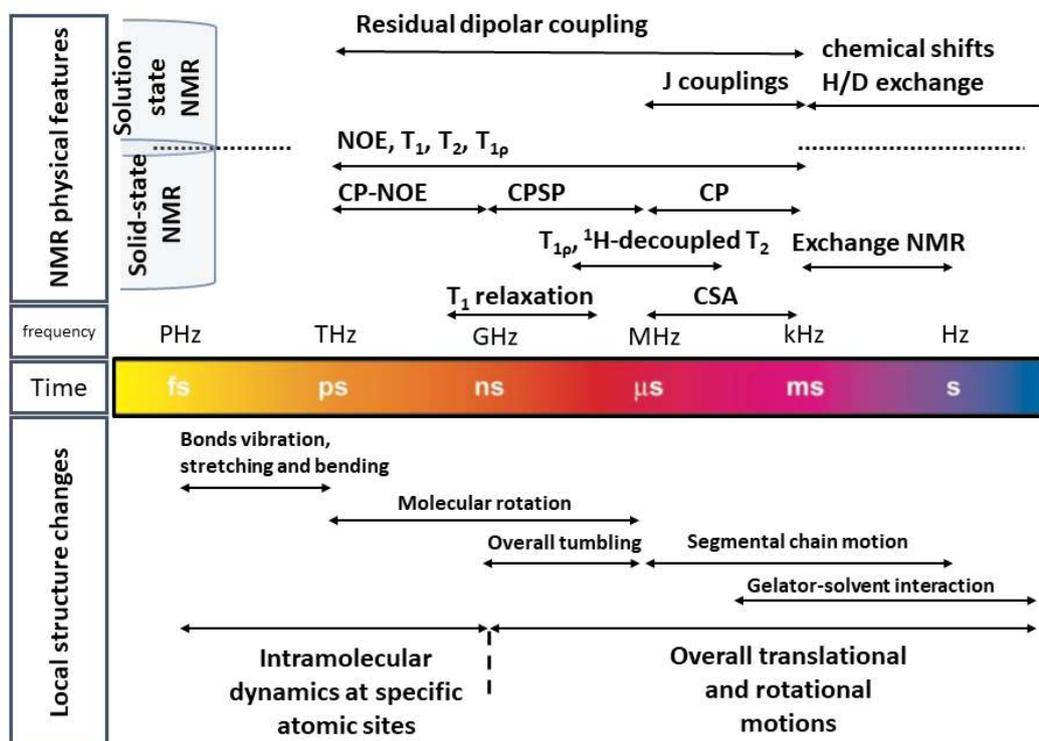
Fluorine is a promising sensitive NMR probe with high intrinsic sensitivity and 100% natural abundance of the NMR active isotope  $^{19}\text{F}$ . Indeed, application of fluorine in combination with NMR screening has been recognised as a powerful tool in the identification and optimisation phases of drug discovery projects. Methods based on protein-, substrate/cofactor-, and ligand-based  $^{19}\text{F}$  NMR screening have been developed for the identification of drugs binding site, measurements of the dissociation binding constant ( $K_D$ ) and binding assays in direct or competition mode. The great sensitivity of  $^{19}\text{F}$  to changes in chemical environment (i.e. protein binding or particle-particle interactions) derives from its great chemical shift dispersion which reduces signals overlap in chemical mixtures. This large dispersion originates mostly from the paramagnetic contribution of energetically low-lying p atomic orbitals.<sup>180</sup>

It is important to note that, although the use of fluorinated organic derivatives had recently found great applicability in biological systems, its use in biomaterials is still mostly under explored.

### 4.1.3 Combining solution and solid-state NMR approaches for material characterization

Macromolecular systems, either in solution or solid phase, present two kinds of molecular motion: i) local motions – in the sub nanometre length scale and the ps – ns timescale (corresponding to PHz to GHz range in frequency scale)<sup>181</sup>, and ii) larger scale motions – within the a length scale greater than nanometre (from GHz to Hz range). The larger scale motions are further divided into translational (diffusion, flow or advection)<sup>182</sup> and rotational motions.

In the case of polymers, such as cellulose or cellodextrin chains which are of interest for this thesis, mobility is determined by the physical state (solid form, colloidal dispersion or dissolved solution), the short and long-range (crystallinity) order, the intramolecular and intermolecular interactions (e.g. H-bond, lipophilic, electrostatic and steric interactions).<sup>183</sup> Macromolecules might present domains of different mobility regimes;<sup>184</sup> for instance, the core of tightly packed polymeric fibers presents lower mobility in comparison with those parts in the solid/air or solid/liquid interface.<sup>92</sup> In addition, a crystalline and ordered organisation would generally be more rigid than an amorphous and disordered form. A schematic representation of commonly employed NMR parameters and their relationship with the scales of molecular motion is shown in Figure 4.1:



**Figure 4.1:** Correlation between different mobility regimes in a macromolecule and corresponding NMR parameters. We represent the principal features of solution and solid-state NMR spectroscopy which reports to different molecular mobility regimes, highlighting how different experiments can be used to obtain information on different timescales.

#### 4.1.3.1 Sources of broadening in NMR spectra

In NMR, careful analysis of peaks broadening can provide valuable information on intermolecular interaction and structural disorder. Notably, in solution and solid-state NMR line widening is affected by different phenomena.

##### Resonance broadening in solution state NMR

In solution state NMR,  $T_2$  relaxation contribution (i.e. the rate of disappearance of net transverse magnetisation) is predominant, and the relationship between  $T_2$  relaxation and molecular tumbling is well established (*Chapter 2*,  $T_1$  and  $T_2$  curves vs molecular tumbling rate, Figure 2.3) For colloidal dispersion, two dynamically different regions can be distinguished; a surface on the colloid-solvent interface, and a core with established both particles-particles intra- and inter-molecular

## Chapter 4 – Deoxyfluorinated cellodextrin derivatives

interactions. Different dynamics regimes have already been reported for polymers constituted both from proteins and carbohydrates.<sup>92, 185</sup>

Historically, peaks broadening has been the first revelation of faster transverse relaxation time. Measurements of NMR relaxation  $T_1$  and  $T_2$  time constants have been used to assess the mobility of the polymeric chain components and hence differentiate crystalline/ordered (more rigid) from non-crystalline/disordered (more mobile) regions. In many cases the difference in mobility of these two regions determines whether or not the signal is observed in the NMR spectra, the more flexible being NMR visible, the more rigid being NMR invisible (hidden in the spectra baseline). Examples of NMR spectra obtained from the residual mobility of a small subunit covalently linked to a larger object are microtubulin-associated proteins in intact microtubules,<sup>185</sup> or solvent-exposed waxy rice starch granules.<sup>92</sup>

On the contrary, the dissolution of polymeric and particulate networks (e.g. dissolution of cellulose at pH 14) is due to the disruption of the intermolecular interactions that hold together the packing of polymer chains. This disruption determines a general decrease of the correlation time of each polymeric chain and, thus, narrower peaks. Material in the gel state would be expected to be more mobile than the solid state, but more restricted than in solution.

### *Resonance broadening in solid-state NMR*

In solid-state NMR the strong dipolar coupling and chemical shift anisotropy (CSA) contribute to line widening in molecules resonances as they are not averaged out by the free isotropic tumbling. Combination of high-power proton decoupling and magic angle sample spinning (MAS) yields solid-state  $^{13}\text{C}$  spectra with resonances appearing at their isotropic chemical shift positions.

Organic solids can be distinguished between crystalline/ordered/rigid systems and noncrystalline/disordered/mobile systems (i.e. glassy polymers below the glass transition temperature  $T_g$ ) or a mixture of them. In polymers presenting crystalline and noncrystalline domains, crystalline regions will give narrow resonances, while noncrystalline regions will give broader resonances. In the latter, the variability in conformations and molecular packing contributes to the chemical shift dispersion by: (1) modification of the bond angles; (2) variation of short-range interactions; (3) modification of the hydrogen bond connections and (4) distortion of the orientation source of CSA. For those systems, the relaxation broadening mechanism is

## Chapter 4 – Deoxyfluorinated cellodextrin derivatives

usually not dominant, although not neglectable. In specific, the higher molecular mobility in correlation with the rotor spinning causes a faster loss in transverse magnetisation coherence.<sup>186</sup>

Generally, solid-state NMR experiments often rely on Cross-Polarization (CP) - a phenomenon explained in the *Chapter 2*, section 2.2.8 - which favour the detection of rigid, strongly dipolar coupled components rather than mobile. In this case, system mobility could contribute to the broadening of <sup>13</sup>C resonances due to motional modulation of the C-H coupling.<sup>187</sup>

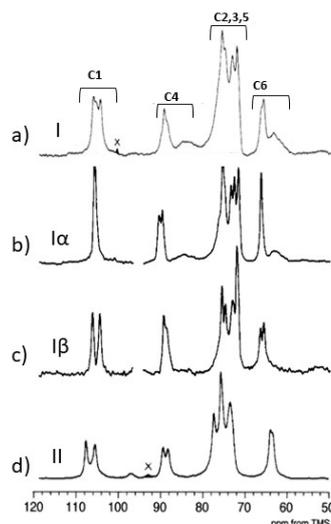
Importantly, in solid-state NMR carbons in different environments (i.e. surface or interior and/or ordered and disordered regions or any inter- or intramolecular inequivalences) might present chemical shift differences.<sup>188</sup>

Given the above we suggest that, in the studies of heterogeneous systems in which domains with different order and mobility coexist, a full assignment of the NMR spectra can be achieved only by a combination of specific solution and solid-state experiments targeting different domains. Within the combination of the aforementioned approaches, we were able to assign resonances belonging to surface (mobile) and core (rigid) regions, visible only in solution and solid-state NMR respectively (see the experimental section for further details).

### 4.1.4 Determination of structure of cellulose by NMR spectroscopy

As previously referred in *Chapter 1*, cellulose can be assembled in different crystal forms (I, II, III and IV).<sup>36</sup> In this thesis chapter, previous structural studies on cellulose allomorphs are used as reference for the novel deoxyfluorinated cellodextrin synthesised by us.

Characterisation of cellulose structure has been performed via solid-state NMR spectroscopy since the early 1980s, with VanderHart and Gast pioneering the work.<sup>189, 190</sup> The assignment of the peaks in the <sup>13</sup>C spectra started with a direct comparison of the peaks obtained for glucose, cellobiose and short cello-oligosaccharides in solution state NMR.<sup>189</sup> Herein we report comparison between <sup>1</sup>H-<sup>13</sup>C CP/MAS NMR spectra of cellulose type I (I $\alpha$  and I $\beta$ ) and II (Figure 4.2) and a comparison of <sup>13</sup>C chemical shifts for cellulose I $\alpha$ , I $\beta$ , II, IIIi, IIIii, IVi and IVii (Table 4.1). The <sup>1</sup>H-<sup>13</sup>C CP/MAS spectra reported in the literature for the allomorphs IIIi, IIIii and IVi, IVii are reported in the corresponding subchapters.



**Figure 4.2:**  $^1\text{H}$ - $^{13}\text{C}$  CP/MAS NMR spectra of a) cotton linters (cellulose type I $\alpha$  and I $\beta$  mixture); b) cellulose allomorph I $\alpha$ , from *Acetobacter*; c) cellulose allomorph I $\beta$ , from low-DP cellulose I; d) cellulose type II allomorph, from a low-DP cellulose II sample. *Source:* The figure was adapted from Atalla and VanderHart, 1984.<sup>188</sup>

**Table 4.1:**  $^{13}\text{C}$  chemical shifts reported in the literature for all the known cellulose allomorphs.

<sup>b</sup> Broad amorphous contribution; <sup>c</sup> values taken from deconvolution data

<i>Cellulose allomorph</i>	<i>Chemical shift (ppm)</i>		
	<b>C1</b>	<b>C4</b>	<b>C6</b>
cellulose I $\alpha$ (Glaucocystis) <sup>44</sup>	105	89.7, 88.8	65.3
cellulose I $\beta$ (tunicin) <sup>44</sup>	105.7, 103.9	88.7, 88.0 <sup>b</sup>	65.5, 64.9 <sup>b</sup>
cellulose II <sup>43</sup>	107.0, 104.7	88.6, 87.4	62.9, 62.2
cellulose III <sub>I</sub> <sup>44</sup>	104.8	87.8	62.3
cellulose III <sub>II</sub> <sup>c, 43</sup>	106.5, 105.7, 106.2 <sup>b</sup>	88.5, 87.3, 85.2 <sup>b</sup>	62.5, 62.1, 61.5 <sup>b</sup>
cellulose IV <sub>I</sub> <sup>39, 191</sup>	105.7, 103.9	88.7, 88.0, 84.8 <sup>b</sup>	65.3, 64.0 <sup>b</sup>
cellulose IV <sub>II</sub> <sup>191</sup>	105.5	83.6, 84.4	63.7

### 4.1.4.1 Cellulose type I ( $\alpha$ and $\beta$ )

The  $^1\text{H}$ - $^{13}\text{C}$  CP/MAS solid-state spectra of microcrystalline cellulose (composite of  $\alpha$  and  $\beta$  allomorphs in different quantities, depending on cellulose source)<sup>192</sup> presented isolated peaks for C1, C4 and C6, while peak overlapping made the unambiguous assignment of C2, C3 and C5 sites impossible. Interestingly, C4 and C6 showed two different components, a sharp one and a broader one. The assignment of those two components has not been trivial, but finally the broad component was assigned to the surface and amorphous regions, while the narrower one to the internal core and crystalline regions of the cellulose elementary fibril. The broadening effect has been attributed to differences in mobility,<sup>193, 194</sup> to higher disorder in packing<sup>187, 189, 195</sup> and to different conformational arrangements.<sup>196</sup> Importantly, it was noted that glucose units on the surface of the cellulose elementary fibril presents looser packing and lower H-bonding constrain in comparison with the anhydroglucose in the centre.<sup>187</sup>

The multiplicity of C1 carbon peak and the narrow peaks of C4 and C6 sites depends on the coexistence of  $\alpha$  and  $\beta$  allomorphs within the cellulose fibril unit cell.<sup>188, 197</sup> Microcrystalline celluloses from different sources exhibit different multiplicities, hence different ratio between the  $\alpha$  and  $\beta$  allomorph components.<sup>188</sup> The  $\alpha$  crystalline form shows singlets for C1 and C6, and a doublet for C4, hence indicating the presence of one (C1 and C6) and two (C4) chemical environments per unit cell. On the other hand, the  $\beta$  crystalline form has doublets at C1, C4 and C6, ascribed to different conformations of the inter-glycosidic linkages and of the hydroxymethyl group in non-equivalent chains (see Figure 4.2 to compare carbons multiplicities).<sup>194</sup>

### 4.1.4.2 Cellulose type II

Differences in peaks splitting can be highlighted for cellulose type II, in which peaks of C1, C4 and C6 appear as doublets with 1:1 intensity. This was ascribed to the inequivalent lattice position of the antiparallel chains of cellulose type II<sup>198</sup> (in contrast with the parallel polarity of cellulose type I) and to conformational differences among the two chains (the center chain is more puckered than the origin chain and presents a different torsion of the glycosidic linkage).<sup>41</sup> In addition, the peaks of C4 and C6 present different chemical shifts for the two allomorphs (reported in Table 4.1).<sup>188, 191</sup> For C6, this is attributed to different conformations arounds of the hydroxymethyl group,

with cellulose type I and cellulose type II adopting a *trans-gauche* (*tg*) and *gauche-trans* (*gt*) conformation, respectively.<sup>196</sup>

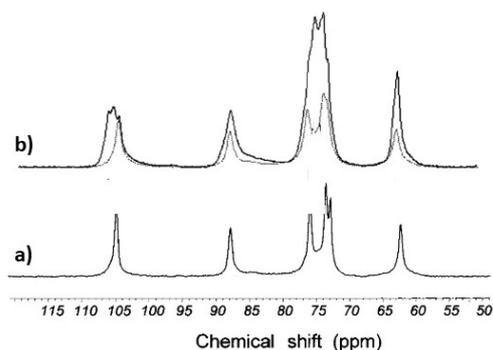
Hence, these differences in multiplicity and chemical shifts can be used as markers to distinguish cellulose I and II in <sup>13</sup>C solid-state NMR spectra (Figure 4.2 and Table 4.1).

### 4.1.4.3 Cellulose type III (IIIi and IIIii)

Cellulose type IIIi and IIIii can be obtained by treatment of cellulose I and II in liquid ammonia or organic amines, respectively.<sup>191</sup> Cellulose IIIi derived from both cellulose I $\alpha$  and I $\beta$  allomorphs presents identical chemical shift patterns. Cellulose IIIi <sup>1</sup>H-<sup>13</sup>C CP/MAS spectrum shows six sharp peaks (singlets), corresponding to the six carbon atoms of the glucosyl ring (Figure 4.3, a; the chemical shifts of C1, C4 and C6 are also reported in Table 4.1). Cellulose IIIi appears to have some features of cellulose II, i.e. matching chemical shifts for C1 and C6 with cellulose type II C1 and C6 upfield peak, and C4 chemical shift correspond exactly to the mean position of cellulose type II C4 doublet.<sup>44</sup> In addition, a change in the chemical shift of the hydroxymethyl carbon was reported (see Table 4.1)<sup>191</sup> and associated to a conformational change from a *trans-gauche* arrangement in cellulose I (64.9 – 65.3 ppm) to *gauche-trans* in IIIi (62.3 ppm) allomorph.<sup>36</sup>

Cellulose IIIii differs from IIIi in the value of C1, C4 and C6 chemical shift and multiplicity (see Figure 4.3, b and Table 4.1, C1, C4 and C6 appear as singlets for cellulose IIIi, triplets for cellulose IIIii) as well as in C2, C3 and C4 patterns.<sup>191</sup> Cellulose IIIii presents a higher chemical shift of the C1 resonance and a similar peak C6 resonance pattern and chemical shift (ascribed to the *gt* conformation of the hydroxymethyl groups).

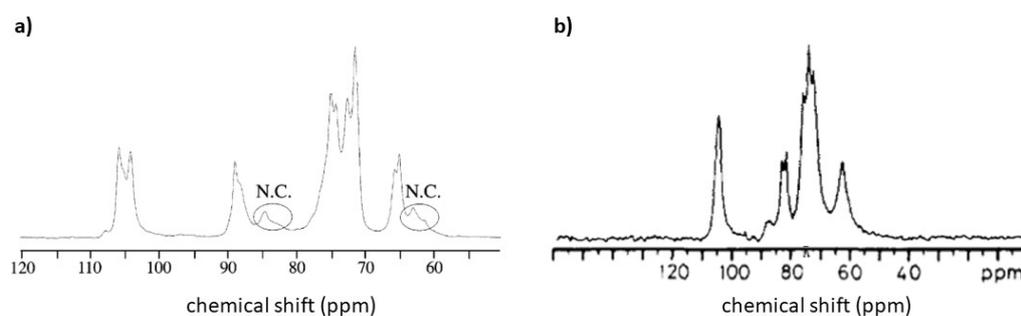
Cellulose IIIii <sup>1</sup>H-<sup>13</sup>C CP/MAS spectrum complexity reveal the presence of two distinct types of cellulose (one type similar to cellulose IIIi, with each carbon atom giving rise to a singlet, the another type giving rise to a series of doublets) and three distinct molecular conformations. In addition, subtraction of cellulose IIIi to cellulose IIIii spectra and peak deconvolution supported the doublet feature for C1, C4 and C6 sites in cellulose IIIii.<sup>43</sup>



**Figure 4.3:**  $^1\text{H}$ - $^{13}\text{C}$  CP/MAS NMR spectra of a) cellulose III<sub>i</sub> obtained from oriented film of microcrystalline cellulose in supercritical ammonia for 1h at 140 °C and b) cellulose III<sub>ii</sub> obtained from mercerized ramie treated with ammonia for 1h at 180 °C (solid line) and cellulose III<sub>i</sub> (dashed line). *Source:* The figure was adapted from a) Wada *et al.*, 2001<sup>44</sup> and b) Wada *et al.*, 2009.<sup>43</sup>

#### 4.1.4.4 Cellulose type IV (IV<sub>i</sub> and IV<sub>ii</sub>)

Cellulose IV, also called “high-temperature cellulose”, is prepared from cellulose II or III by treatment in glycerol at 260 °C, while cellulose I cannot be transformed directly into cellulose IV.<sup>39, 199</sup> The allomorph cellulose IV<sub>i</sub> is exclusively formed from cellulose III<sub>ii</sub> treatment, while the allomorph cellulose IV<sub>ii</sub> is obtained from both cellulose II and III<sub>ii</sub> as source material. Figure 4.4 presents the  $^1\text{H}$ - $^{13}\text{C}$  CP/MAS NMR spectra recorded for cellulose IV<sub>i</sub> and IV<sub>ii</sub>.<sup>39, 191</sup>



**Figure 4.4:** a) cellulose IV<sub>i</sub> obtained by treatment of *Cladophora* cellulose first with super-critical ammonia and heating at 140° C for 1h (to obtain cellulose type III<sub>ii</sub>) and then with glycerol at 260° C for 30 minutes; N.C. = Non Crystalline b) cellulose IV<sub>ii</sub> prepared from the cellulose II sample by heating in water at 190 °C for 2 h. *Source:* a) Wada *et al.*, 2004<sup>39</sup> and b) Isogai *et al.*, 1989.<sup>191</sup>

## Chapter 4 – Deoxyfluorinated cellodextrin derivatives

The spectra of cellulose IVi (Figure 4.4, a) presents similarities with cellulose I $\beta$  spectra in carbons C1, C4 and C6 chemical shifts and multiplicities (Table 4.1), as well as in the pattern of C2,3,4. Wada *et al.*,<sup>39</sup> recently suggested that cellulose IVi is not a distinct polymorph, but a finely-divided cellulose I $\beta$  with higher lateral disorder introduced during the heat treatment. Furthermore, Newman *et al.*,<sup>200</sup> reported a good fit in simulating cellulose IVi diffractogram data based on cellulose I $\beta$  unit cell.

On the contrary, the existence of cellulose IVii as a distinct polymorph is supported by several characteristic features in the <sup>1</sup>H-<sup>13</sup>C CP/MAS NMR spectrum (Figure 4.4, b). C4 appears as a sharp doublet resonance at 83.6 - 84.4 ppm assigned to the crystalline component, while almost the same position was assigned to the C4 surface/amorphous component in cellulose I allomorphs (suggesting a similar molecular conformation of the crystal-interior chains in cellulose IVii and the crystal-surface chains in cellulose I).<sup>200</sup> C1 and C6 appear as single resonances at 105.5 and 63.7 ppm, respectively. C6 resonance appears in the chemical shift range of the gt conformation for the hydroxymethyl group.<sup>196</sup> C2,3,5 resonances display maxima at 73.8, 74.8 and 77.0 ppm, respectively, and show a triplet pattern similar to cellulose type II allomorph.<sup>191</sup>

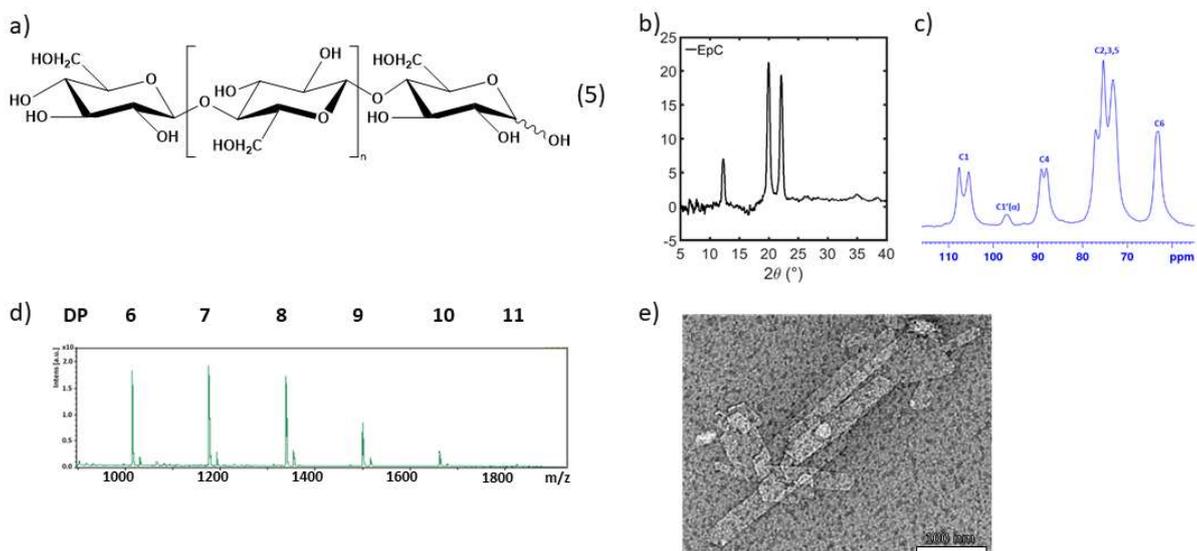
## 4.2 Experimental section

### 4.2.1 Background on the synthesis of cellodextrin and cellodextrin fluorinated derivatives

#### 4.2.1.1 Enzymatic synthesis of cellodextrin (EpC)

The synthesis of short enzymatically produced cellodextrin chains (EpC, 5) was performed using glucose-1-phosphate (6) as donor and D-cellobiose (7) as acceptor. The complete characterisation of EpC (5) (chemical structure reported in Figure 4.5, a) by PXRD, solid-state <sup>1</sup>H-<sup>13</sup>C CP/MAS NMR, Maldi-ToF and TEM methods (Figure 4.5, b, c, d and e) was carried out and used as reference for the deoxy-fluorinated cellodextrin chains described below. Solid-state NMR <sup>1</sup>H-<sup>13</sup>C CP/MAS and the PXRD confirm the organisation the chain into cellulose type II allomorph.<sup>128</sup> NMR spectrum of EpC shows multiplicity in C1 and C4, which is consistent with the anti-parallel chain orientation of cellodextrins, determined by the relative inter-chain polarity of the reducing and nonreducing

rings.<sup>201</sup> In addition, the PXRD pattern is in agreement with previously described cellulose type II.<sup>128</sup> The Maldi-Tof confirmed polydispersity of the synthesised chains within a DP between 6 and 11 anhydroglucose units.<sup>128</sup> TEM showed a platelet-like (sheets) shapes for the aggregated EpC chains.



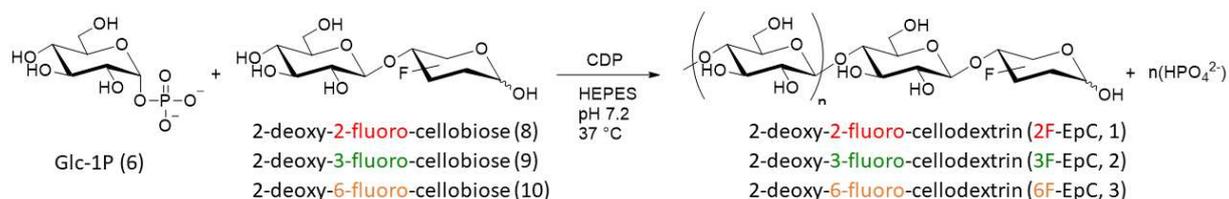
**Figure 4.5:** Enzymatically produced cellodextrin (EpC, 5) a) chemical structure; b) PXRD patterns (Cu  $\lambda$  1.54 Å) ; c)  $^1\text{H}$ - $^{13}\text{C}$  CP/MAS NMR spectra of EpC powder recorded using a 400 MHz spectrometer (12 kHz spinning rate, contact time of 1 ms and recycling delay of 5 s); d) MALDI-ToF (matrix assisted laser desorption/ionization–time-of-flight mass spectrometry (MALDI-ToF MS) and e) TEM image of EpC produced from G1P 200 mM, Clb 25 mM in HEPES buffer 500 mM at pH 7 and T 37 °C.

#### 4.2.1.2 Enzymatic synthesis of fluorinated cellulose-like derivatives:

Substitution of the acceptor substrate D-cellobiose (7) with the deoxy-fluoro-derivatives 2-deoxy-2-fluoro-cellobiose (8), 3-deoxy-3-fluoro-cellobiose (9) and 6-deoxy-6-fluoro cellobiose (10) led to the formation of cellodextrin chains containing a fluorine atom at the reducing ring in position 2, 3 and 6, respectively. The structures of the acceptor molecules and the 2-deoxy-2-fluoro-cellodextrin (1), 3-deoxy-3-fluoro-cellodextrin (2) and 6-deoxy-6-fluoro cellodextrin (3) are reported in Figure 4.6, together with the reaction scheme. Description of the reaction conditions utilised for the enzymatic synthesis of the monofluorinated derivatives 8, 9 and 10 and the subsequent CDP catalysed polymerization reaction using

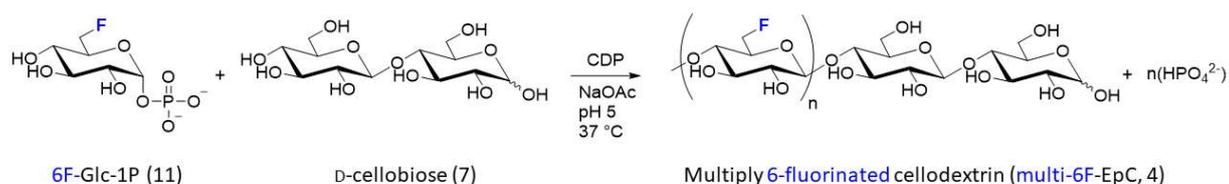
## Chapter 4 – Deoxyfluorinated cellodextrin derivatives

$\alpha$ -D-glucose-1-phosphate (6) is reported in the material and methods (subsection 4.4.2.1 and 4.4.2.2).



**Figure 4.6:** (Left) Schematic representation of the reverse phosphorolysis reaction carried out by CDP in the presence of 2-, 3- or 6-fluoro-substituted acceptor molecules (8, 9 and 10), and (Right) chemical structure of the single 2-, 3- and 6-fluorinated cellodextrin product 1, 2 and 3. The synthesis of the compounds 1, 2, 3, 8, 9 and 10 was performed by Peterson De Andrade (John Innes Centre).

The substitution of the CDP natural donor  $\alpha$ -D-glucose-1-phosphate (6) with its derivative 6-deoxy-6-fluoro-glucose-1-phosphate (11) led to the production of cellodextrin chains with fluorine atoms regio-selectively introduced in position 6 all along the chain (multiply 6-fluorinated cellodextrin or multi-6F-EpC, 4), with the exception of the cellobiose moiety at the reducing terminal (from the acceptor; Figure 4.7). The reaction conditions are reported in the material and methods subsection 4.4.2.3 and 4.4.2.4.



**Figure 4.7:** (Left) Schematic representation of the reverse phosphorolysis reaction carried out by CDP in the presence of 6-deoxy-6-fluoro glucose-1-phosphate (donor molecule, 11) and (Right) chemical structure of the obtained product multiply 6-fluorinated cellodextrin (4). The synthesis of the compound 11 and 4 was carried out by Peterson De Andrade (John Innes Centre).

### 4.2.2 Deoxyfluorinated cellodextrin derivative multi-scale characterisation

The complete characterisation of the novel materials was achieved throughout the joint effort of a scientific consortium established between the University of East Anglia, the John Innes Centre and the University of Bristol. A combination of Transmission Electron Microscopy (TEM), Atomic Force Microscopy (AFM), Raman spectroscopy, Powder X-ray diffraction and advanced Nuclear Magnetic Resonance techniques has been employed for the thorough structural characterisation of compounds 1, 2, 3 and 4. In the following sections the discussion will focus on X-ray and NMR characterisations, as my main contribution to this work. The conclusions derived from TEM, AFM and Raman spectroscopy are important to be taken into account for the complete structural characterisation of the materials and they will be briefly discussed in the following session.

#### 4.2.2.1 *Morphological characterisation*

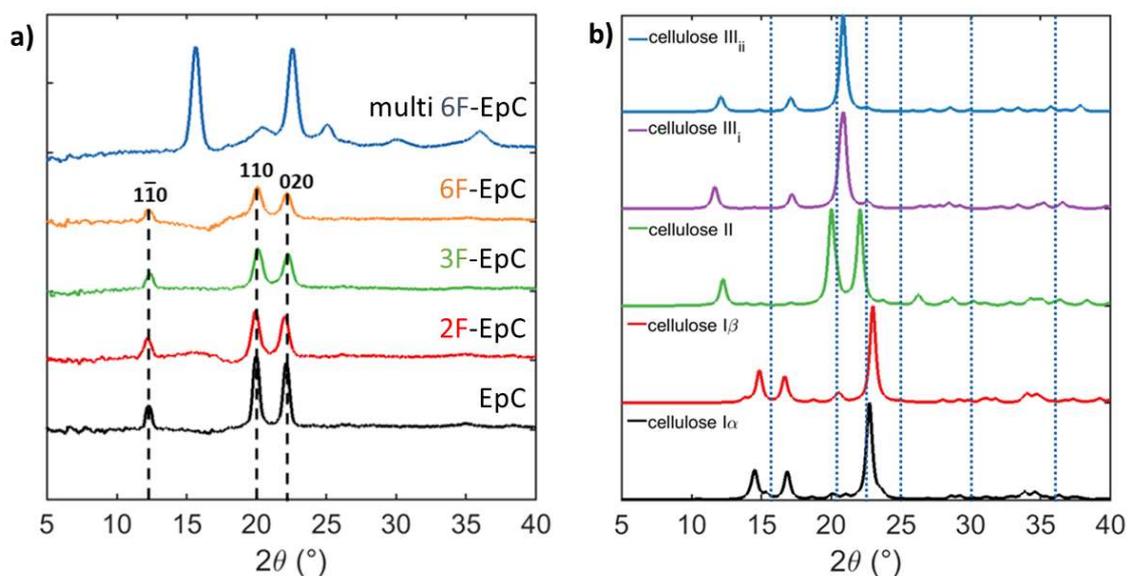
TEM images (Appendix, Figure A.19) show that the introduction of a single fluorine in the reducing end does not modify the sheet-like morphology previously reported for both enzymatically synthesised unfunctionalized<sup>128, 133</sup> and functionalised (such as acrylated cellulose<sup>142</sup> and cellulose conjugated with oligo(ethylene glycol)<sup>139</sup>) cello-oligosaccharides. On the other hand, multi-6F-EpC (4) formed shorter platelets (< 100 nm length). AFM imaging (Appendix, Figure A.20) confirmed the highlighted differences showing the formation of few long platelets in multi-6F-EpC (4) but still in a smaller fraction in comparison with EpC (5).

Raman spectra (Appendix, Figure A.21) did not show any significant changes between EpC (5), 2F-EpC (1), 3F-EpC (2), and 6F-EpC (3), but new bands corresponding to CH<sub>2</sub>F groups and a shift of the band associated with the glycosidic-linkage were observed for multi-6F-EpC (4). The latter indicates that the crystalline arrangement of multi-6F-EpC (4) does not correspond to either cellulose type I or II. In this regard, quantum mechanics calculations for fluorinated cellobiose showed that the substitution of all hydroxyl groups by fluorine atoms does not affect either the ring puckering of glucose (<sup>4</sup>C<sub>1</sub>) or the conformation of the glycosidic bond.<sup>202</sup> Hence, the shift in the Raman spectra was attributed to differences in packing of the oligosaccharide chains, rather than to conformational changes.

### 4.2.2.2 *Long-range structural characterisation*

X-ray powder diffraction (PXRD) is a rapid and effective analytical technique for the characterisation of the long-range structural order of materials (see Chapter 2, section 2.3 for further details). Importantly, each crystalline material (different crystallites or even allomorph of the same crystallite) presents a characteristic X-ray diffraction patterns, which works as fingerprint for the specific crystal organisation.

2F-EpC (1), 3F-EpC (2) and 6F-EpC (3) show diffraction patterns very similar to that of EpC (5), indicating that the single fluorinated cellodextrins assemble into cellulose type II crystalline organisation. This allomorph is characterised by three intense and sharp diffraction peaks at  $2\theta$  of  $12.24^\circ$ ,  $20.02^\circ$  and  $22.08^\circ$  (d-spacings of 0.74, 0.44 and 0.40 nm, respectively), which represent the  $(1\bar{1}0)$ ,  $(110)$  and  $(020)$  planes.<sup>128</sup> On the contrary, multi-6F-EpC (4) produced a very different diffraction pattern, with two well defined peaks at  $2\theta$  equal to  $15.6^\circ$  and  $22.5^\circ$  (d-spacings of 0.57 and 0.39 nm, respectively), and four different broad components at  $2\theta$  of  $21.6^\circ$ ,  $25^\circ$ ,  $30.2^\circ$  and  $36^\circ$  (d-spacings of 0.41, 0.36, 0.30 and 0.25 nm, respectively). The collected diffraction patterns are shown in Figure 4.8, a. Interestingly, the comparison with the PXRD patterns of all cellulose allomorphs reported so far ( $I\alpha$ ,  $I\beta$ , II, III<sub>i</sub> and III<sub>ii</sub>) demonstrated that multi-6F-EpC (4) ensemble into a crystalline organisation that is unprecedented for a cellulose-like material (Figure 4.8, b and Table 4.2).



**Figure 4.8:** a) Experimental PXR patterns for EpC (5, black), 2F-EpC (1, red), 3F-EpC (2, green), 6F-EpC (3, orange) and multi-6F-EpC (4, blue). All the samples were analysed in powder form; b) Predicted diffraction patterns for cellulose types I $\alpha$  (black)<sup>40</sup>, I $\beta$  (red)<sup>38</sup>, II (green)<sup>41</sup>, III<sub>I</sub> (purple)<sup>42</sup> and III<sub>II</sub> (blue)<sup>43</sup> from previously published synchrotron data. The patterns were generated using Mercury with FWHM set to 0.5. The position of the experimental 2 $\theta$  values for multi-6F-EpC (4) are shown as blue dashed lines.

**Table 4.2:** Predicted 2 $\theta$  and *d-spacing* values of cellulose I $\alpha$ , I $\beta$ , II, III<sub>I</sub> and III<sub>II</sub>. Peak indices refer to cellulose II PXR pattern (Figure 4.9, a).

	PEAK 1 (1 $\bar{1}$ 0)		PEAK 2 (110)		PEAK 3 (020)	
	2 $\theta$	<i>d</i> (nm)	2 $\theta$	<i>d</i> (nm)	2 $\theta$	<i>d</i> (nm)
I $\alpha$	14.56	0.610	16.86	0.530	22.74	0.390
I $\beta$	14.86	0.600	16.66	0.530	22.98	0.390
II	12.24	0.722	20.02	0.443	22.08	0.402
III <sub>I</sub>	11.66	0.758	17.18	0.516	20.88	0.425
III <sub>II</sub>	12.1	0.731	17.1	0.518	20.86	0.425

**Table 4.3:** Experimental  $2\theta$  and  $d$ -spacing values of multi-6F-EpC (4).

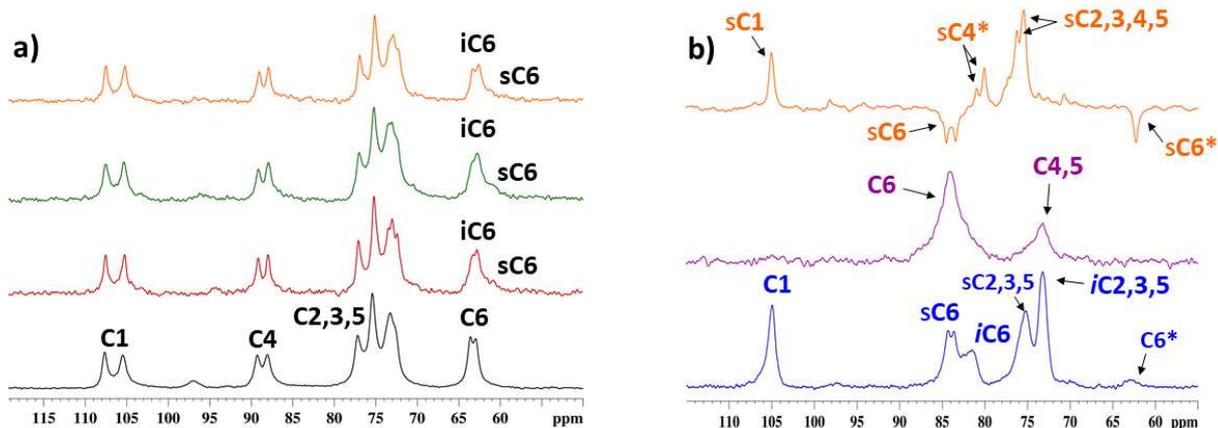
	PEAK 1		PEAK 2		PEAK 3		PEAK 4		PEAK 5		PEAK 6	
	$2\theta$	$d$ (nm)										
4	15.6	0.570	21.6	0.410	22.5	0.390	25.0	0.360	30.2	0.300	36.0	0.250

#### 4.2.2.3 Molecular characterisation by solution and solid-state NMR

The  $^1\text{H}$ - $^{13}\text{C}$  CP/MAS NMR experiment is the most widely used solid-state NMR method to characterise solid or solid-like samples, and it can be used as molecular fingerprint of the material.<sup>203, 204</sup>

Figure 4.9, a shows the  $^1\text{H}$ - $^{13}\text{C}$  CP/MAS NMR spectra of the monofluorinated EpCs (1, 2, and 3) in comparison with non-modified EpC (5). The collected spectra show the characteristic cellulose type II fingerprint. Interestingly, the spectra of monofunctionalised EpCs show a broad peak at *ca.* 61 ppm that corresponds to the surface/disordered population of C6 site. The C6 chemical environment has been previously reported for bacterial cellulose<sup>205</sup>, plant cell walls,<sup>206</sup> nanocrystalline cellulose<sup>203</sup> and TEMPO-oxidised cellulose nanofibrils.<sup>207</sup>

In contrast, the  $^1\text{H}$ - $^{13}\text{C}$  CP/MAS NMR spectrum of multi-6F-EpC (4) (Figure 4.9, b) presented a totally different molecular fingerprint, indicating a different chain packing in comparison to EpC (5). The different multiplicity of C1 (doublet in EpC and singlet in multi-6F-EpC) indicates the presence of two and one non-equivalent anomeric carbons per unit cell in EpC and multi-6F-EpC, respectively. As previously described, the multiplicity in C1 of EpC reflects the anti-parallel orientation of the cellodextrin chains.<sup>201</sup>



**Figure 4.9:** a)  $^1\text{H}$ - $^{13}\text{C}$  CP/MAS NMR spectra of EpC (black) in the powder form acquired using MAS rate of 10 kHz and the single fluorinated 2F-EpC, 3F-EpC, 6F-EpC 10 wt% dispersions acquired using MAS rate of 6 kHz and 5 °C; b) Comparison of the  $^1\text{H}$ - $^{13}\text{C}$  CP (blue) and  $^{19}\text{F}\{^1\text{H}\}$ - $^{13}\text{C}$  CP (purple) spectra of powdered multi-6F-EpC (4, 15 kHz MAS rate and room temperature) and  $^{13}\text{C}$  DEPT-135 (with  $^1\text{H}$  decoupling) spectrum (orange) of multi-6F-EpC (4) 0.5 wt% dispersion in  $\text{D}_2\text{O}$ . The \* symbole indicates the resonances of the cellobiose reducing terminal, i and s indicate internal and surface resonances, respectively.

Notably, the  $^1\text{H}$ - $^{13}\text{C}$  CP/MAS spectrum of multi-6F-EpC does not correspond to any of the cellulose  $^1\text{H}$ - $^{13}\text{C}$  CP/MAS NMR chemical shifts reported so far in the literature, as showed in Table 4.1 and Table 4.3. Only some similarities were observed between multi-6F-EpC and cellulose III<sub>I</sub> (comparable chemical shifts and multiplicity for C1 and C6); however, the large differences between their diffraction patterns (Figure 4.8, Table 4.2) clearly indicates a different crystalline organisation for multi-6F-EpC.

The full assignment of multi-6F-EpC (4) could be achieved only by combining solution and solid-state NMR experiments (Figure 4.10, b and Figure 4.11), approach already discussed in the introduction of this chapter (subsection 4.1.3).

**Table 4.3:**  $^{13}\text{C}$  chemical shifts of peaks of multi-6F-EpC (4) acquired in solid-state (top) and solution state (bottom) NMR.

Chemical shifts (ppm)

<i>Solid-state NMR</i>					
C1	sC6	iC6c	sC2,3,4,5	iC2,3,4,5	C6*
104.8	84.2, 83.5	81.9	75.2	73.1	62.6

<i>Solution state NMR</i>					
sC1	sC6	sC4*	sC2,3,4,5	sC6*	
104.8	84.3, 83.3	79.9	76.0, 75.3	62.1	

$\{^1\text{H} - ^{19}\text{F}\}$ -decoupled  $^{19}\text{F}$ - $^{13}\text{C}$  CP/MAS spectra (Figure 4.9, b) enabled the assignment of C6, C5 and C4 peaks of the fluorinated residues in multi-6F-EpC (4). The highest intensity peak at lower field (83.8 ppm) was assigned to C6 as the closest carbon atom to fluorine (1.3 Å), and hence more efficient cross-polarisation. The peak at 73.1 ppm was assigned to C5 and C4 (second shortest distance from fluorine), while C3 and C2 were too far away from fluorine to cross-polarise effectively. On the other hand, the solution state  $^{13}\text{C}$  DEPT-135 spectra allowed us to assign the methylene peaks corresponding to 6-monofluorinated (C6) and non-fluorinated (C6\*) residues (in antiphase with respect to the CH carbons), and confirmed the assignment of C6 in the  $^1\text{H}$ - $^{13}\text{C}$  CP/MAS NMR spectrum (Figure 4.9, b).

Further, 2D solution NMR spectroscopy was used to complete the characterisation of both EpC (5) and multi-6F-EpC (4) (Figure 4.10). COSY and HSQC experiments of EpC (Figure 4.10, a) enabled the detailed assignment of the  $\alpha$  and  $\beta$  anomeric spin systems and distinguish between internal and reducing end peaks for H1, H2 and H6, H6'. In addition, the non-reducing ring chemical shifts were assigned (*nr* labels, Figure 4.10, a). With regard to multi-6F-EpC (4) the spectra presented a higher level of complexity due to the presence of a larger number of possible chemical environments (Figure 4.10, b). Nonetheless, it was still possible to differentiate between the cellobiose unit at the reducing end (labelled with a star - \*), the fluorinated residues and the non-reducing terminal environment (labelled as *nr*).



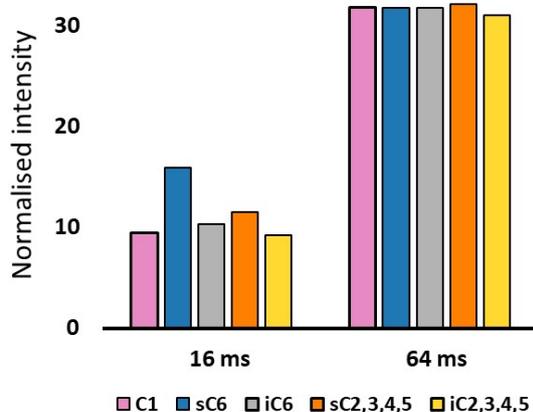
## Chapter 4 – Deoxyfluorinated cellodextrin derivatives

are not observed due to large peak broadening. Hence, the combination of  $^{13}\text{C}$ -detected solution and solid-state NMR experiments allow for the differentiation of chemical environments from surface and interior domains in gels. In this regard, it should be noted that the  $^1\text{H}$ - $^{13}\text{C}$  CP/MAS peaks of multi-6F-EpC (4) at 81.9 and 73.1 ppm (Figure 4.9, b) are not observed in either the  $^{13}\text{C}$  DEPT-135 or the  $^1\text{H}$ - $^{13}\text{C}$  HSQC solution NMR experiments (Figure 4.9, b and Figure 4.10, b). Thus, these CP-observed peaks were assigned to the interior (immobile) domains (*i*C6 and *i*C2,3,5, Figure 4.9, b) whereas the C6 and C2,3,4,5 chemical environments observed only in the  $^{13}\text{C}$  DEPT-135 solution NMR experiments were assigned to surface (mobile) domains (*s*C6 and *s*C2,3,4,5, respectively, Figure 4.9, b).

The assignment of *s*C6 and *i*C6 was further validated by water polarisation transfer CP (WPT-CP) solid-state NMR experiments (Figure 4.11).<sup>100</sup> In these experiments, polarization transfer is mediated by three mechanisms: i) chemical exchange – uniformly active in our samples due to the presence of OH groups with labile protons, ii) nuclear Overhauser effect (NOE) – expected to be less statistically relevant at moderate MAS frequency of 6 kHz, and iii) spin diffusion mediated by coherent  $^1\text{H}$ - $^1\text{H}$  dipolar coupling, which depends on both internuclear distances and mobilities. Therefore, in our experiments the water–polysaccharide polarization transfer is driven by chemical exchange followed by spin diffusion and the build-up curves of the recorded peaks intensities depends on i) the number of interacting water molecules at a particular site and ii) the distance and relative mobility of bound water at the particle surface. As peaks corresponding to surface domains are more water-exposed, it is expected for the a faster WPT growth at short mixing times in comparison with interior domains. This has been recently observed in a study conducted by our research group in bacterial cellulose (BC), where we have been able to characterize to which extent individual Plant Cell Wall (PCW) polysaccharides affect the hydration of BC.<sup>205</sup>

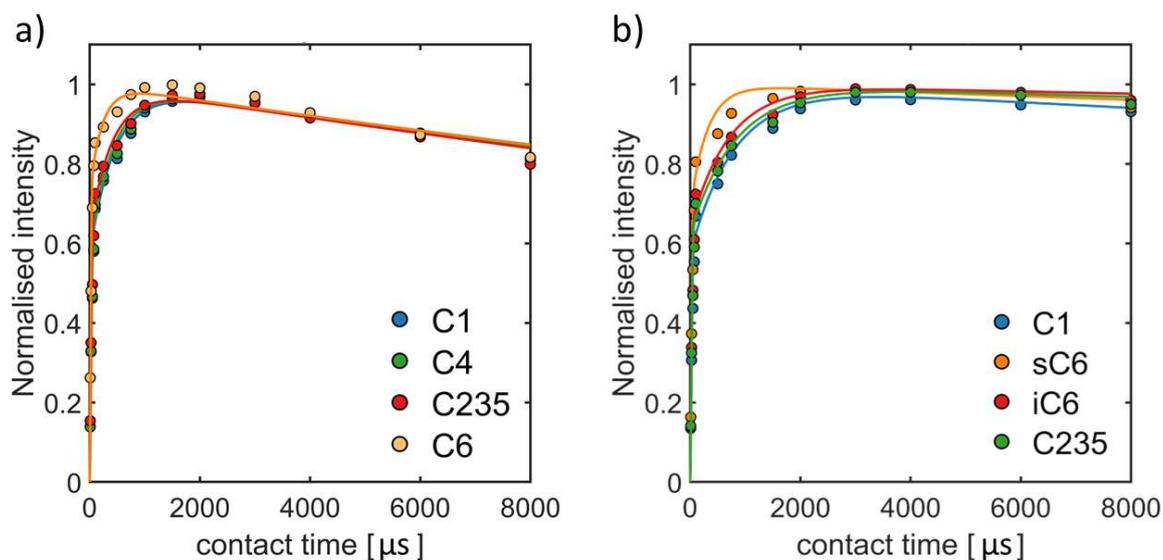
At sufficiently long mixing times, WPT become homogeneous for both surface and interior domains due to the efficient spin diffusion. Figure 4.11 shows the WPT factors for a 25 wt% dispersion of multi-6F-EpC (4) at 16 ms and 64 ms mixing time, point at which homogenisation of surface-interior water polarisation transfer is achieved. At short mixing time, a much higher WPT factor was observed for the *s*C6 (83.8 ppm) compared to the *i*C6 peak (81.8 ppm), confirming the assignment of *s*C6 and *i*C6 peaks to surface and interior domains, respectively. Also, *s*C2,3,4,5 showed higher

WPT compared to  $iC_{2,3,4,5}$ , in agreement with solution NMR data where the  $sC_{2,3,4,5}$  and  $iC_{2,3,4,5}$  peaks are visible and invisible, respectively.



**Figure 4.11:** a) Bar graph showing the water polarisation transfer (WPT) factors determined for each carbon peak of multi-6F-EpC 25 wt% hydrogel using a mixing time of 16 ms and 64 ms. At 16 ms mixing time, the higher WPT factor observed for  $sC_6$  and  $sC_{2,3,4,5}$  compared to  $iC_6$  and  $iC_{2,3,4,5}$  demonstrates the increased solvation of the formers, hence being assigned to surface domains. At 64 ms mixing time, the WPT is homogenised by the spin diffusion process.

To characterise the internal dynamics of these materials,  $^1H$ - $^{13}C$  CP/MAS NMR kinetics experiments were carried out for both EpC (5) and multi-6F-EpC (4) (Figure 4.12). The CP build-up curves were simulated using the equation already presented in Chapter 2, section 2.1.2.10, Equation 2.48 (simplified I-I\*-S CP kinetics model), as reported by Kolodziejewski *et al.*<sup>99</sup> The obtained parameter  $T_{1\rho}^I$ ,  $T_{df}$ ,  $\lambda$  and  $T_2$  represent, respectively, the  $I$  spin lattice relaxation time in the rotating frame, the  $^1H$  spin-diffusion time constant describing the strength of the homonuclear dipolar interactions, the homogeneity of the  $I$  spin pool (hence, the number  $n$  of  $I$  spins attached to the  $S$  spin under study), and is the spin-spin relaxation time (Table 4.4 and 4.5). The  $^1H$ - $^{13}C$  CP kinetics analysis showed a very similar behaviour for both EpC (5) and multi-6F-EpC (4), indicating a similar degree of rigidity at the molecular level for both materials



**Figure 4.12:**  $^1\text{H}$ - $^{13}\text{C}$  CP/MAS kinetics curves of EpC (a) and multi-6F-EpC (b) acquired with a 12 kHz MAS rate, a  $^{13}\text{C}$  frequency of 100.6 MHz at 298 K. The curves are fitted to the I-I\*-S CP kinetics model.

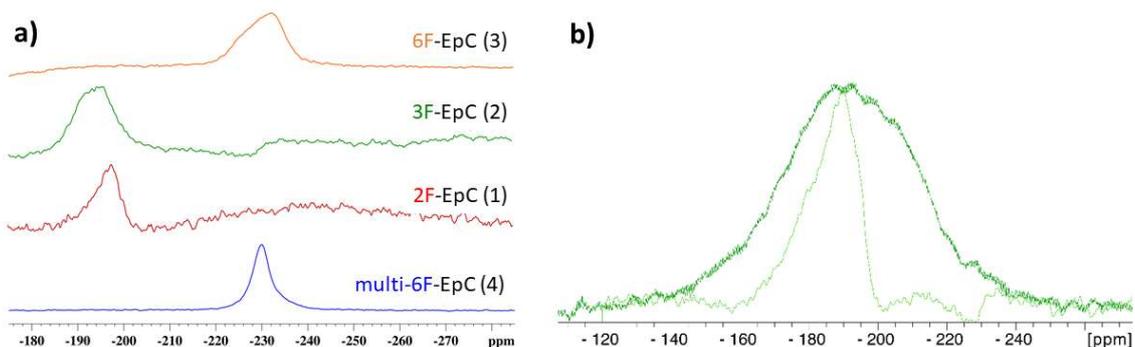
**Table 4.4:** Parameters obtained from the fit of the  $^1\text{H}$ - $^{13}\text{C}$  CP/MAS build-up curves of EpC (Figure 4.13, a) to Equation 2.48. The error indicates 95% confidence.

	C1		C4		C235		C6	
$\lambda$	0.4278	$\pm 0.07$	0.4228	$\pm 0.07$	0.3769	$\pm 0.09$	0.272	$\pm 0.12$
$T_{1\rho}$ (ms)	> 100		> 100		> 100		> 100	
$T_2$ ( $\mu\text{s}$ )	23.07	$\pm 4.83$	23.57	$\pm 4.77$	23.72	$\pm 5.39$	18.17	$\pm 3.94$
$T_{df}$ ( $\mu\text{s}$ )	456	$\pm 162.6$	419.6	$\pm 150.4$	408.3	$\pm 200.6$	223.7	$\pm 195.8$
$R^2$	0.987		0.987		0.981		0.976	

**Table 4.5:** Parameters obtained from the fit of the  $^1\text{H}$ - $^{13}\text{C}$  CP/MAS build-up curve of multi-6F-EpC (Figure 4.13, b) to Equation 2.48. The error indicates 95% confidence.

	C1		sC6		iC6		C2345	
$\lambda$	0.4259	$\pm 0.07$	0.3132	$\pm 0.1$	0.3621	$\pm 0.07$	0.3746	$\pm 0.07$
$T_{1\rho}$ (ms)	> 100		> 100		> 100		> 100	
$T_2$ ( $\mu\text{s}$ )	24.05	$\pm 5.14$	25.94	$\pm 5.54$	23.75	$\pm 4.44$	24.36	$\pm 5.10$
$T_{df}$ ( $\mu\text{s}$ )	810.4	$\pm 306.4$	308.5	$\pm 225.5$	702.8	$\pm 308.2$	819.3	$\pm 374$
$R^2$	0.9887		0.9876		0.9859		0.9959	

In addition, direct polarisation  $^{19}\text{F}$  NMR experiments were carried out for 2F-EpC (1), 3F-EpC (2), 6F-EpC (3) and multi-6F-EpC (4) (Figure 4.13, a). The monofluorinated cellodextrins presented a broad and asymmetric peak centred at -190 (2F-EpC, 1), -197 (3F-EpC, 2) and -232 ppm (6F-EpC, 3), respectively, while multi-6F-EpC (4) showed a sharper Lorentzian-shaped peak centred at -230 ppm. The line widths at half height of each derivative are reported in Table 4.6. The recorded chemical shifts are in good agreement with reported solution state NMR  $^{19}\text{F}$  data (Figure 4.14).

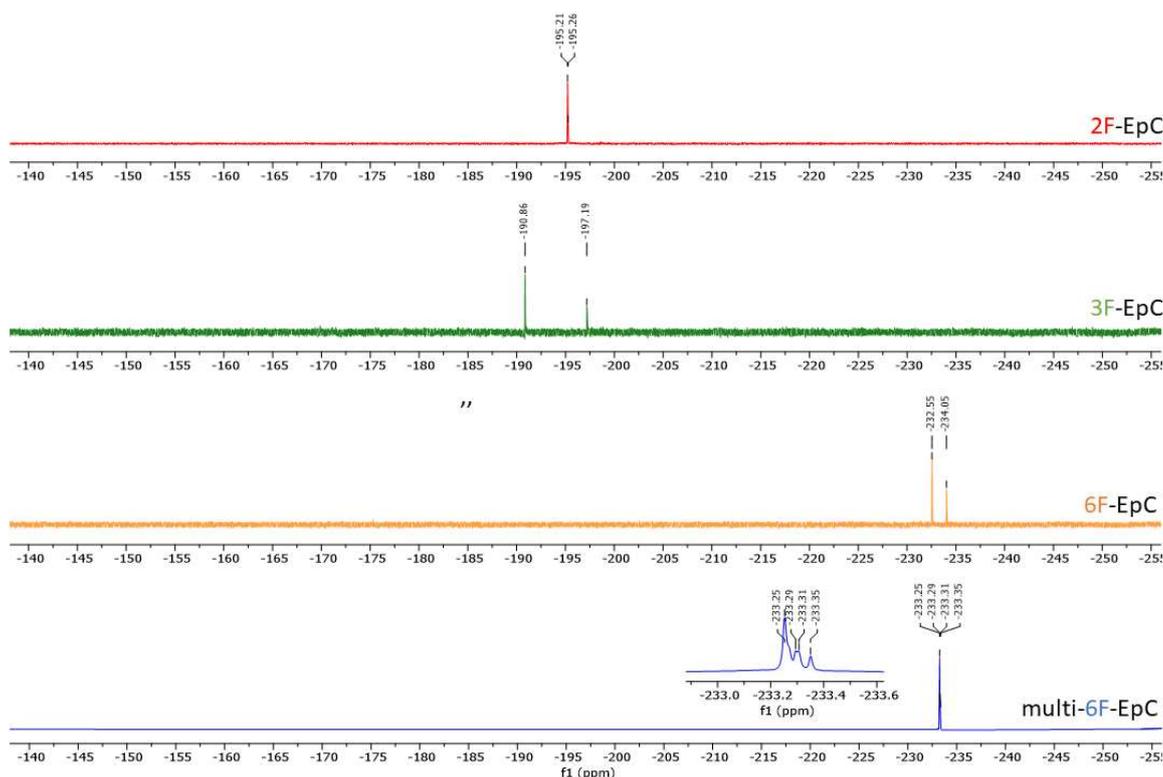


**Figure 4.13:** a)  $^{19}\text{F}$  MAS NMR spectra of multi-6F-EpC (4, blue) and 2F- (1, red), 3F- (2, green) and 6F-EpC (3, orange) powders, acquired at 60 kHz MAS rate and 800 MHz  $^{19}\text{F}$  frequency (20 T magnetic field) without  $^1\text{H}$  decoupling; b)  $^{19}\text{F}$  NMR spectrum of 3F-EpC powder acquired with (dark green) and without (light green)  $^1\text{H}$ - $^{19}\text{F}$  decoupling at the MAS rate of 15 kHz and 60 kHz, respectively, and 800 MHz  $^{19}\text{F}$  frequency.

## Chapter 4 – Deoxyfluorinated cellodextrin derivatives

**Table 4.6:** Full width at half maximum (in kHz) measured without  $^1\text{H}$ - $^{19}\text{F}$  decoupling during acquisition at 60 kHz MAS rate for 2F-EpC, 3F-EpC, 6F-EpC and multi-6F-EpC and with  $^1\text{H}$ - $^{19}\text{F}$  decoupling during acquisition at 15 kHz MAS rate for 3F-EpC.

	FWHM	
	Without decoupling (kHz)	With decoupling (kHz)
<b>2F-EpC</b>	3.7	
<b>3F-EpC</b>	11.9	33.0
<b>6F-EpC</b>	9.4	
<b>Multi-6F-EpC</b>	3.8	



**Figure 4.14:**  $^{19}\text{F}\{^1\text{H}\}$  NMR spectra (1 M NaOD) of 2-, 3- and 6-monofluorinated cellodextrins (2F-EpC, 1; 3F-EpC, 2 and 6F-EpC, 3) and multiply 6-fluorinated cellodextrin (multi-6F-EpC, 4).

Interestingly, it was noted that spectra recorded using a slower MAS rate (15 KHz instead of 60 KHz) and with the application of proton decoupling resulted in a broader  $^{19}\text{F}$  peak, as represented

for 3F-EpC in Figure 4.13, b. This observation indicated that, under certain conditions, fast MAS can decouple more efficiently than radiofrequency pulses. This is possible due to the probe setup employed to perform  $^1\text{H}$ -decoupled  $^{19}\text{F}$  experiments at low MAS conditions, where both  $^{19}\text{F}$  and  $^1\text{H}$  had to be tuned simultaneously leading to an imperfect tuning for both nuclei. Excluding a strong contribution of homonuclear and heteronuclear dipolar couplings in the spectra recorded at fast MAS, peak broadening was attributed to the large heterogeneity of  $^{19}\text{F}$  chemical environments. Indeed, in the investigated fluorinated-cellodextrins, the  $^{19}\text{F}$  atoms can be found potentially in different environments in the nanofibrils (i.e. surface, core, far from or nearby other fluorinated residues) resulting in a distribution of chemical shifts.

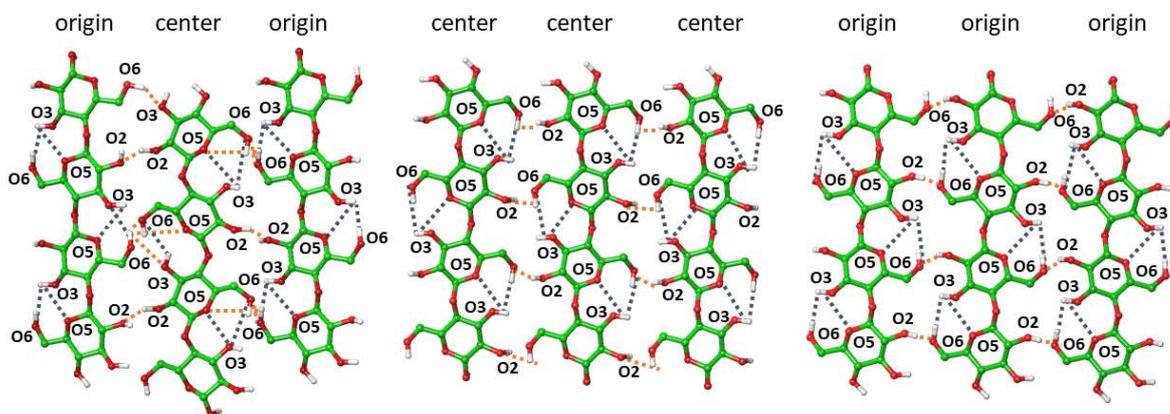
### 4.3 Discussion and future prospective

Our multiscale approach combining microscopy, Raman, PXRD and NMR demonstrates that the insertion of a single fluorine atom per cellodextrin chain does not exert a substantial impact on the morphology and long and short-range order structure of the material. On the contrary, the introduction of multiple 6-deoxyfluoro substituents along the cellodextrin chain yielded to a crystalline allomorph never reported before for a cellulose-like material.

To rationalise the potential impact of deoxyfluorination on the cellodextrin structure, it is key to analyse the hydrogen bonds (HB) network that defines the cellulose type II arrangement. Langan *et al.*,<sup>58</sup> reported three different HBs patterns for cellulose type II allomorph, established between two different types (origin, o, and centre, c) cellulose chains, as shown in Figure 4.15.<sup>41, 208</sup> HBs were identified within groups less than 2.8 Å apart and with an angle D-H-A greater than 110° (where D = HB donor and A = HB acceptor). Table 4.5 reports the intra- and inter-chain HBs. In specific, intermolecular HBs between origin (o-o-o), centre chains (c-c-c) or centre and origin chains (o-c-o) were identified as O2-H...O6 (o-o-o), O6-H...O2 (c-c-c) and O6-H...O6 and O2-H...O2 (o-c-o), respectively. Also, intramolecular HBs O3-H...O5 with a minor component involving O3-H...O6 are present in each arrangement, while O6-H...O3 is present only in o-c-o sheets.

In addition, Dalvit *et al.*, reported on the ability of fluorine to form intermolecular HB only either in the presence of a strong HB donors or at HB donor high concentration.<sup>209</sup> Thus, the presence of

other HB acceptor in the investigated samples (all the other hydroxyl groups) will compete with the formation of HBs with the fluorine atom.



**Figure 4.15:** Representation of cellulose type II hydrogen bonds network differentiating between origin and centre chain. Intra- and inter-chain hydrogen bonds are represented as blue and orange dashed lines, respectively. *Source:* Langan *et al.*, 1999.<sup>208</sup>

**Table 4.5:** Summary of all the intra- and inter-chain hydrogen bonds for each different chain arrangement in cellulose type II allomorph.

O-C-O		O-O-O		C-C-C	
intrachain	Interchain	intrachain	interchain	intrachain	interchain
O3(D)···O6(A)	O6(D)···O3(A)	O3(D)···O6(A)	O6(D)···O2(A)	O3(D)···O6(A)	O2(D)···O6(A)
O3(D)···O5(A)	O2(D)···O2(A)	O3(D)···O5(A)		O3(D)···O5(A)	
O6(D)···O3(A)	O6(D)···O6(A)				
	O6(D)···O5(A)				
	O6(D)···O3(A)				

The site-specific hydroxyl-group substitution with a fluorine atom will prevent the HB donor action of that specific position. In specific, the substitution of the 3-hydroxyl group of glucose with a fluorine atom would affect the formation of the intra-chain hydrogen bond with O6, while the formation of the O2···O2 and O6···O6 inter-sheet hydrogen bonds of cellulose II would be affected in 2F- and 6F-EpC, respectively. Thus, the introduction of a single fluorine atom in the reducing ring unit would affect both the intra-chain and inter-chain HB network, leading to increased disorder.

## Chapter 4 – Deoxyfluorinated cellodextrin derivatives

Nonetheless, the most significant effect was expected for the multiple 6-deoxy-fluoro substitution (i.e. multi-6F-EpC), as the 6-hydroxyl group is highly involved in inter-chain packing. The different organisation of multi-6F-EpC is also reflected in the  $^1\text{H}$ - $^{13}\text{C}$  CP/MAS NMR spectra; indeed, this material exhibits a totally different CP fingerprint in comparison with cellulose type II. In addition, the different multiplicity reported for C1 between EpC and multi-6F-EpC indicated the loss of the antiparallel chain arrangement typical of cellulose type II, and the presence of only one non-equivalent anomeric carbon per unit cell in multi-6F-EpC. These data are supported by an unprecedented PXRD pattern collected for multi-6F-EpC. Additionally, both ordered and disordered domains are seen for multi-6F-EpC by PXRD (presence of broad components) and NMR (where two very different  $^{13}\text{C}$  chemical environments observed for C6).

In conclusion, we have demonstrated the enzymatic incorporation of single and multiple fluorinated glucose residues into cellodextrin chains. The OH to F substitution is tolerated by the cellodextrin phosphorylase, albeit at low efficiency. Nonetheless, we were able to produce selectively fluorinated cellodextrins averaging ca DP 9 in size, which self-assemble into crystalline materials. Singly fluorinated cellodextrins display structural features reminiscent of cellulose II, as judged by solid-state NMR, powder X-ray diffraction and Raman spectroscopy. In contrast, multiply 6-fluorinated cellodextrin gave rise to a new allomorph, not previously reported for either native celluloses or cellulose-like materials. Advanced solid-state NMR methods have enabled the detailed characterisation of these novel materials, deciphering the water exposed and interior chemical environments for different carbon sites. Our findings highlight the considerable potential of chemoenzymatic synthesis for generating novel glycomaterials of controlled molecular structure and morphology.

## 4.4 Material and methods

### 4.4.1 General materials and methods

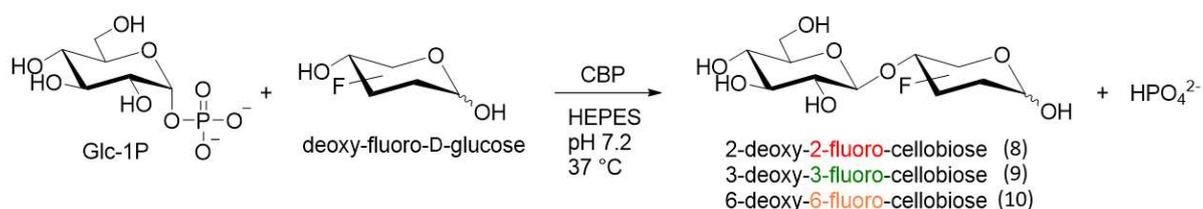
Chemicals were commercially obtained as reagent grade and used without any purification. Deoxy-fluoro-D-glucoses (2F-, 3F- and 6F-Glc) and  $\alpha$ -D-glucose 1-phosphate disodium salt hydrate (Glc-1P) were purchased from Toronto Research Chemicals (Canada) and Sigma-Aldrich (UK),

## Chapter 4 – Deoxyfluorinated cellodextrin derivatives

respectively. Cellobiose phosphorylase (CBP) (PRO-GH94-004) was kindly provided by Prozomix Limited (UK) and Milli-Q (MQ) H<sub>2</sub>O was used to prepare all buffers. Thin-layer chromatography (TLC) was performed on pre-coated silica gel 60 F<sub>254</sub> plates (Merck) and compounds were visualised by UV irradiation ( $\lambda$  254 nm) and/or by spraying TLC with staining solution (2% orcinol w/v in EtOH/H<sub>2</sub>O/H<sub>2</sub>SO<sub>4</sub> 15:1:2 v/v/v) followed by heating. Biotage SP4 flash chromatography system was used for purification of protected monosaccharides using normal phase (pre-packed SNAP cartridges) and the monofluorinated cellobiose analogues were purified by HPLC (Thermo Scientific Dionex Ultimate 3000) on a Luna OH column (5  $\mu$ m HILIC 200 Å, 250  $\times$  10 mm, Phenomenex) using 5 mM ammonium formate buffer (5%) and acetonitrile (95%) at 5 mL/min in isocratic elution over 25 min. Detection was performed by charged aerosol detector (CAD) with power function 1.00, data collection rate 10 Hz and nebulizer temperature 25 °C. Products were lyophilised using a Labconco FreeZone Benchtop freeze dryer. <sup>1</sup>H, <sup>13</sup>C, <sup>31</sup>P and <sup>19</sup>F NMR spectra were recorded on a Bruker Avance III 400 MHz and/or Bruker Avance Neo 600 MHz spectrometers at 298 K. Chemical shifts recorded in D<sub>2</sub>O are reported with respect to the solvent residual peak at 4.79 ppm in <sup>1</sup>H NMR. High resolution mass spectra were acquired in a Synapt G2-Si mass spectrometer (Waters, UK) using electrospray ionisation (positive or negative mode). Optical rotations were measured at 20 °C using a Perkin-Elmer Model 341 polarimeter.

### 4.4.2 Synthesis of deoxyfluorinated-cellodextrin derivatives

#### 4.4.2.1 Enzymatic synthesis of 2-, 3-, and 6-monofluorinated cellobioses 8, 9 and 10



To a solution of deoxy-fluoro-glucose (F-Glc) (18.2 mg, 100 mM stock solution in Milli-Q (MQ) water, 1 mL, 1 eq., 13 mM final concentration) and Glc-1P (disodium salt hydrate, 30.4 mg, 100 mM stock solution in MQ water, 1 mL, 1 eq., 13 mM final concentration) in HEPES buffer (80 mM stock solution in MQ water, 5 mL, pH 7.2), was added cellobiose phosphorylase (CBP, 1 mg/mL

## Chapter 4 – Deoxyfluorinated cellodextrin derivatives

stock solution, 1 mL, 1 mg) and the reaction was incubated at 37 °C shaking for 16 h. The reaction mixture was centrifuged, and the supernatant was passed through HisTrap HP (1 mL) and High Q (5 mL) Bio-Rad columns to remove CBP and remaining Glc-1P, respectively. The pre-purified mixture was freeze dried, dissolved in 2 mL MQ water and purified by gel filtration chromatography (Toyopearl TSK HW40S, 1.6 × 90 cm, MQ water, 0.5 mL/min) followed by HPLC (Luna OH 5 μm HILIC 200 Å, 250 × 10 mm, Phenomenex, 5 mL/min in isocratic elution over 25 min, 5% 5 mM ammonium formate buffer and 95% acetonitrile) to afford the title compounds 8-10:

### 2-Deoxy-2-fluoro-cellobiose (8).

Yield: 26% (9 mg; 0.029 mmol); Rf = 0.40 (isopropanol:NH<sub>4</sub>OH:H<sub>2</sub>O, 6:3:1); [α]<sub>D</sub><sup>20</sup> +32.5 (c 1, CH<sub>3</sub>OH). <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O) δ 5.47 (0.45 H, d, *J*<sub>1,2</sub> 3.9 Hz, H1α), 4.95 (0.5 H, dd, *J*<sub>1,2</sub> 7.9 Hz, *J*<sub>1,F</sub> 2.4 Hz, H1β), 4.54 (1 H, two d, *J*<sub>1',2'</sub> 8.0 Hz, H1'), 4.48 (0.45 H, ddd, *J*<sub>2,F</sub> 49.5 Hz, *J*<sub>2,3</sub> 9.5 Hz, *J*<sub>1,2</sub> 3.9 Hz, H2α), 4.22-4.07 (1 H, m, H2β, H3α), 4.02-3.87 (3.5 H, m, H5α, H3β, H6<sub>a</sub>β, H6<sub>a</sub>α, H6'<sub>a</sub>, H6<sub>b</sub>α), 3.83 (0.5 H, dd, *J*<sub>6a,6b</sub> 12.4 Hz, *J*<sub>5,6</sub> 5.0 Hz, H6<sub>b</sub>β), 3.78-3.72 (2 H, m, H4α, H4β, H6'<sub>b</sub>), 3.67 (0.5 H, ddd, *J*<sub>4,5</sub> 9.9 Hz, *J*<sub>5,6a</sub> 5.0 Hz, *J*<sub>5,6b</sub> 2.2 Hz, H5β), 3.56-3.50 (2 H, m, H3', H5'), 3.47-3.42 (1 H, m, H4'), 3.37-3.32 (1 H, m, H2'). <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O): δ 102.48 (C1'), 93.44 (C1<sub>β</sub>), 93.21 (C2<sub>β</sub>), 91.99 (C2<sub>β</sub>), 90.55 (C2<sub>α</sub>), 89.44 (C1<sub>α</sub>), 89.32 (C2<sub>α</sub>), 78.02 (C4<sub>α</sub>), 77.97 (C4<sub>β</sub>), 75.97 (C5'), 75.46 (C3'), 74.89 (C5<sub>β</sub>), 73.11 (C2'), 72.76 (C3<sub>β</sub>), 69.94 (C5<sub>α</sub>), 69.92 (C3<sub>α</sub>), 69.48 (C4'), 60.64 (C6'), 59.83 (C6<sub>β</sub>), 59.68 (C6<sub>α</sub>). <sup>19</sup>F NMR (<sup>1</sup>H-decoupled, 376 MHz, D<sub>2</sub>O) δ -199.08, -199.27. HRMS (ESI): *m/z* calculated for C<sub>12</sub>H<sub>21</sub>FO<sub>10</sub>Na<sup>+</sup> [M+Na]<sup>+</sup>: 367.1011; found: 367.1013.

### 3-Deoxy-3-fluoro-cellobiose (9).

Yield: 12% (4 mg; 0.012 mmol); Rf = 0.29 (isopropanol:NH<sub>4</sub>OH:H<sub>2</sub>O, 6:3:1); [α]<sub>D</sub><sup>20</sup> +28.7 (c 1, CH<sub>3</sub>OH). <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O) δ 5.28 (0.5 H, t, *J*<sub>1,2</sub> = *J*<sub>1,F</sub> 3.8 Hz, H1α), 4.77-4.66 (1 H, m, H3α, H1β), 4.64-4.51 (1.5 H, m, H3β, H1'), 4.04-3.96 (2 H, m, H5α, H4α, H4β, H6<sub>a</sub>β), 3.94-3.90 (2 H, m, H6<sub>a</sub>α, H6'<sub>a</sub>, H6<sub>b</sub>α), 3.88-3.82 (1 H, m, H2α, H6<sub>b</sub>β), 3.72 (1 H, dd, *J*<sub>6'a,6'b</sub> 12.5 Hz, *J*<sub>5,6'b</sub> 5.9 Hz, H6'<sub>b</sub>), 3.63-3.60 (0.5 H, m, H5β), 3.57 (0.5 H, ddd, *J*<sub>2,F</sub> 14.2, *J*<sub>2,3</sub> 9.1, *J*<sub>1,2</sub> 8.0 Hz, H2β), 3.53-3.49 (1 H, m, H3'), 3.48-3.44 (1 H, m, H5'), 3.42-3.37 (1 H, m, H4'), 3.33-3.28 (1 H, m, H2'). <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O): δ 102.39 (C1'), 95.29 (C3<sub>β</sub>), 95.05 (C1<sub>β</sub>), 94.08 (C3<sub>β</sub>), 93.73 (C3<sub>α</sub>), 92.53 (C3<sub>α</sub>), 91.99 (C1<sub>α</sub>), 76.04 (C5'), 75.69 (C5<sub>α</sub>), 75.56 (C3'), 75.49 (C4<sub>α</sub>), 73.84 (C5<sub>β</sub>), 73.27 (C2'), 72.76 (C2<sub>β</sub>), 70.07 (C2<sub>α</sub>), 69.86 (C4<sub>β</sub>), 69.61 (C4'), 60.67 (C6'), 59.74 (C6<sub>β</sub>), 59.61 (C6<sub>α</sub>). <sup>19</sup>F NMR (<sup>1</sup>H-decoupled, 376 MHz, D<sub>2</sub>O)

$\delta$  -192.35, -197.01 HRMS (ESI):  $m/z$  calculated for  $C_{12}H_{21}FO_{10}Na^+$   $[M+Na]^+$ : 367.1011; found: 367.1014.

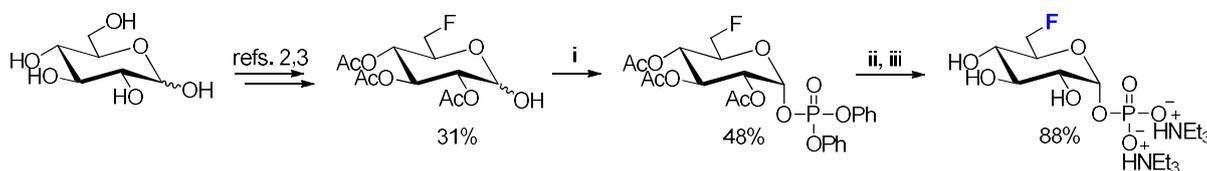
*6-Deoxy-6-fluoro-cellobiose (10).*

Yield: 23% (8 mg; 0.023 mmol);  $R_f = 0.34$  (isopropanol:NH<sub>4</sub>OH:H<sub>2</sub>O, 6:3:1);  $[\alpha]_D^{20} +17.0$  (c 1, CH<sub>3</sub>OH). <sup>1</sup>H NMR (600 MHz, D<sub>2</sub>O)  $\delta$  5.27 (0.55 H, d,  $J_{1,2}$  3.8 Hz, H1 $\alpha$ ), 4.96-4.85 (1 H, m, H6<sub>a,b</sub> $\alpha$ ), 4.76-4.65 (1.5 H, m, H1 $\beta$ , H6<sub>a,b</sub> $\beta$ ), 4.52 (1 H, d,  $J_{1',2'}$  7.9 Hz, H1'), 4.15-4.06 (0.55 H, m, H5 $\alpha$ ), 3.94 (1 H, dd,  $J_{6'a,6'b}$  12.3 Hz,  $J_{5,6'a}$  2.2 Hz, H6'a), 3.87 (0.55 H, t,  $J_{2,3} = J_{3,4}$  9.3 Hz, H3 $\alpha$ ), 3.82-3.72 (2.5 H, m, H5 $\beta$ , H4 $\alpha$ , H4 $\beta$ , H6'b), 3.69-3.65 (0.45 H, m, H3 $\beta$ ), 3.62 (0.55 H, dd,  $J_{2,3}$  9.8 Hz,  $J_{1,2}$  3.8 Hz, H2 $\alpha$ ), 3.56-3.49 (2 H, m, H3', H5'), 3.47-3.43 (1 H, m, H4'), 3.37-3.31 (1.5 H, m, H2', H2 $\beta$ ). <sup>13</sup>C NMR (151 MHz, D<sub>2</sub>O):  $\delta$  102.62 (C1'), 95.90 (C1 $\beta$ ), 91.96 (C1 $\alpha$ ), 82.39 (C6 $\alpha$ ), 82.12 (C6 $\beta$ ), 81.28 (C6 $\alpha$ ), 81.01 (C6 $\beta$ ), 77.80 (C4 $\alpha$ ), 77.56 (C4 $\beta$ ), 75.97 (C5'), 75.48 (C3'), 74.10 (C3 $\beta$ ), 73.76 (C2 $\beta$ ), 73.33 (C5 $\beta$ ), 73.15 (C2'), 71.24 (C3 $\alpha$ ), 71.13 (C2 $\alpha$ ), 69.42 (C4'), 69.02 (C5 $\alpha$ ), 60.56 (C6'). <sup>19</sup>F NMR (<sup>1</sup>H-decoupled, 376 MHz, D<sub>2</sub>O)  $\delta$  -233.68, -234.24. HRMS (ESI):  $m/z$  calculated for  $C_{12}H_{21}FO_{10}Na^+$   $[M+Na]^+$ : 367.1011; found: 367.1014.

**4.4.2.2 Enzymatic synthesis of 2-, 3-, and 6-monofluorinated cellodextrins (2F-EpC, 1, 3F-EpC, 2, and 6F-EpC, 3)**

The monofluorinated cellodextrins were synthesised in a one-pot reaction. After CBP removal, more Glc-1P (110 mg, 4 eq.) was added together with cellodextrin phosphorylase (CDP, 0.9 mg/mL stock solution, 0.7 mL, 0.63 mg) and the reaction was incubated at 50 °C while shaking for 16 h. A white precipitate was formed and isolated by centrifugation, followed by re-suspension and washing with MQ water (3 $\times$ ). To the supernatant, more Glc-1P was added (90 mg) and the reaction incubated at 50 °C shaking for 12 h. A white precipitate was formed again and isolated by centrifugation, followed by re-suspension and washing with MQ water (3 $\times$ ). The precipitates were combined to give reasonable final yields [47% 2F-EpC (**1**), 30% 3F-EpC (**2**) and 32% 6F-EpC (**3**)]. The white solid was analysed by MALDI-ToF (DP8 on average) and solution-state <sup>19</sup>F NMR (<sup>1</sup>H-decoupled, 376 MHz, 1 M NaOD)  $\delta$  -195.21 and -195.26 ppm (**1**); -190.86 and -197.19 ppm (**2**); -232.55 and -234.05 ppm (**3**).

#### 4.4.2.3 Chemical synthesis of 6-deoxy-6-fluoro- $\alpha$ -D-glucose 1-phosphate (6F-Glc-1P, 11)



**Reagent and conditions:** i) *n*-BuLi, (PhO)<sub>2</sub>POCl, THF, -78 °C to room temperature, overnight; ii) PtO<sub>2</sub>, H<sub>2(g)</sub>, EtOH, room temperature, 24 h; iii) Et<sub>3</sub>N:H<sub>2</sub>O:MeOH (1:3:7, v/v/v), room temperature, 48 h.

*n*-BuLi (2 mL of 1.6 M in hexane; 3.20 mmol; 1.2 eq.) was added dropwise to a solution of 6-deoxy-6-fluoro-2,3,4-tri-*O*-acetyl- $\alpha$ -D-glucopyranose (822 mg; 2.67 mmol; 1 eq.) (synthesised from D-glucose in five steps)<sup>2,3</sup> in anhydrous THF (40 mL) at -78 °C under N<sub>2</sub> atmosphere and stirred for 15 min. Diphenyl chlorophosphate (664  $\mu$ L; 3.20 mmol; 1.2 eq.) was added dropwise and the reaction mixture was allowed to warm to room temperature overnight. The reaction was quenched with NH<sub>4</sub>Cl saturated solution (30 mL) and partitioned with EtOAc (3 x 30 mL). The organic phase was washed with NaCl saturated solution (2 x 30 mL), dried over MgSO<sub>4</sub>, filtered, concentrated under vacuum and purified by flash chromatography [cartridge SNAP 25g; solvent: Hexane/EtOAc; gradient: 0-20%, 20-20% and 20-30% (v/v); flow: 25 mL/min] to afford the product 6-deoxy-6-fluoro-2,3,4-tri-*O*-acetyl- $\alpha$ -D-glucopyranosyl 1-diphenylphosphate<sup>4</sup> in 48% yield (690 mg; 1.28 mmol). The deprotection steps of the synthesised phosphate (600 mg; 1.11 mmol) were performed with PtO<sub>2</sub> (50 mg; 0.22 mmol; 0.2 eq.) in absolute ethanol (25 mL) at room temperature under H<sub>2</sub> atmosphere for 48 h. After catalyst removal by filtration, the crude was concentrated under vacuum, dissolved in MeOH (20 mL) followed by addition of Et<sub>3</sub>N (10 mL) and concentrated again. Lastly, the residue was dissolved in Et<sub>3</sub>N:H<sub>2</sub>O:MeOH (1:3:7, v/v/v) (45 mL) and stirred for 48 h at room temperature. After concentration under vacuum, the crude was dissolved in water and freeze dried to afford 6-deoxy-6-fluoro- $\alpha$ -D-glucose 1-phosphate triethylammonium salt<sup>4</sup> in 88% yield (457 mg; 0.98 mmol).

<sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O)  $\delta$  5.49 (1 H, dd, *J*<sub>1,P</sub> 7.0 Hz, *J*<sub>1,2</sub> 3.5 Hz, H1), 4.83-4.58 (2 H, m, H6a, H6b), 3.97 (1 H, dd, *J*<sub>5,F</sub> 30.3 Hz, *J*<sub>4,5</sub> 10.6 Hz, H5), 3.78 (1 H, t, *J*<sub>2,3</sub> = *J*<sub>3,4</sub> 9.5 Hz, H3), 3.60-3.51 (2 H, m, H4,

## Chapter 4 – Deoxyfluorinated cellodextrin derivatives

H<sub>2</sub>), 3.20 (10 H, q, *J* 7.4 Hz, 2x HN(CH<sub>2</sub>CH<sub>3</sub>)<sub>3</sub>), 1.28 (16 H, t, *J* 7.3 Hz, 2x HN(CH<sub>2</sub>CH<sub>3</sub>)<sub>3</sub>). <sup>13</sup>C NMR (101 MHz, D<sub>2</sub>O): δ 94.61 (C1), 82.80 (C6), 81.13 (C6), 72.56 (C3), 71.48 (C2), 71.36 (C5), 71.17 (C5), 68.27 (C4), 46.62 (HN(CH<sub>2</sub>CH<sub>3</sub>)<sub>3</sub>), 8.19 (HN(CH<sub>2</sub>CH<sub>3</sub>)<sub>3</sub>). <sup>19</sup>F NMR (<sup>1</sup>H-decoupled, 376 MHz, D<sub>2</sub>O): δ -236.30. <sup>31</sup>P NMR (162 MHz, D<sub>2</sub>O): δ -1.07. HRMS (ESI): *m/z* calculated for C<sub>6</sub>H<sub>11</sub>FO<sub>8</sub>P [M-H]<sup>-</sup>: 261.0181; found 261.0173.

### 4.4.2.4 Enzymatic synthesis of multiply 6-fluorinated cellodextrin (multi-6F-EpC, 4)

Cellodextrin phosphorylase (CDP, 5.7 mg/mL stock solution, 300 μL, 1.7 mg) was added to a solution of cellobiose (5.5 mg, 16.1 μmol, 1 eq.) and 6-deoxy-6-fluoro-α-D-glucose 1-phosphate (6F-Glc-1P) (triethylammonium salt, 45 mg, 96.6 μmol, 6 eq.) in 500 μL NaOAc buffer (200 mM, pH 5.0) and the reaction was incubated at 37 °C for 24 h with shaking (300 rpm). A white precipitate started forming and more CDP was added (1-fold) and the reaction continued in the same conditions for 48 h (72 h in total). The white precipitate was isolated by centrifugation, re-suspended and washed with MQ water (4×) and freeze dried to afford the multiple 6-fluorinated cellodextrin **7** in 64% yield (17 mg; 10.3 μmol). The white solid was analysed by MALDI-ToF (DP 10 on average) and solution-state <sup>19</sup>F NMR (<sup>1</sup>H-decoupled, 565 MHz, 1 M NaOD) δ -233.25, -233.29, -233.31, -233.35 ppm.

## 4.4.3 Analytical characterisation

### 4.4.3.1 Matrix-assisted laser desorption ionisation time-of-flight mass spectrometry (MALDI-ToF MS)

(DHB) matrix (10 mg/mL in 30% acetonitrile in MQ water), spotted on a target plate (Bruker MTP 384 Polished Steel TF Target) and analysed on an Autoflex<sup>TM</sup> Speed MALDI-TOF/TOF mass spectrometer (Bruker Daltonics<sup>TM</sup> GmbH, Coventry, UK). The instrument was controlled by a flexControl<sup>TM</sup> (version 3.4, Bruker) method optimised for peptide detection and calibrated using peptide standards (Bruker). All spectra were processed with flexAnalysis<sup>TM</sup> (3.4, Bruker).

### 4.4.3.2 *Electron microscopy (EM)*

Transmission Electron Microscopy (TEM) images were viewed on a Thermo Fisher Talos F200C transmission electron microscope at 200kV (Thermo Fisher UK Ltd, Cambridge, UK) using a Gatan OneView 4k X4K digital camera (Gatan, Abingdon, UK) to record DM4 files. 400 mesh EM Resolution Formvar/Carbon coated copper grids were glow discharged, the samples were suspended in 200  $\mu$ L MQ water and 5  $\mu$ L drop was pipetted onto the grid for 1 minute. The excess was then blotted off with filter paper and 5  $\mu$ L of 2% uranyl acetate was pipetted onto the grid for 30 seconds then blotted off with filter paper and allowed to dry.

### 4.4.3.3 *Powder X-ray diffraction (PXRD)*

X-ray diffraction data were collected using a single crystal diffractometer (Rigaku Synergy S, Cu X-ray tube, 50kV-1mA) with Cu K $\alpha$  radiation ( $\lambda = 0.154$  nm). Samples of enzymatically produced cellodextrin (EpC), 2F-EpC, 3F-EpC, 6F-EpC and multi-6F-EpC were placed in a 96-well plate and analysed using an XtalCheck screening plate mounted on the diffractometer. The collected diffraction images were integrated between diffraction angles ( $2\theta$ ) 5 and 40°. All the samples were analysed in powder form and after gentle grinding with mortar and pestle. The simulated powder patterns were generated using Mercury<sup>5</sup> and the published crystal structures of cellulose I $\alpha$ ,<sup>40</sup> I $\beta$ ,<sup>38</sup> II,<sup>41</sup> III $\alpha$ ,<sup>42</sup> and III $\beta$ .<sup>43</sup> All the graphs were generated using MATLAB®.

### 4.4.3.4 *Solution and solid-state NMR*

#### *Solution state NMR*

The NMR characterisation of EpC dispersed in D<sub>2</sub>O was carried out using a Bruker Avance I spectrometer equipped with a 5 mm probe operating at frequencies of 499.7 MHz (<sup>1</sup>H) and 125.7 MHz (<sup>13</sup>C). Around 600  $\mu$ L of dispersion (2 w/V%) in 99.9% D<sub>2</sub>O (Sigma-Aldrich®) was pipetted into a 5 mm NMR tube at room temperature. Both phase-sensitive <sup>1</sup>H-<sup>13</sup>C HSQC experiment with <sup>1</sup>H-<sup>13</sup>C correlation via double inept transfer and <sup>1</sup>H-<sup>1</sup>H COSY experiment with multiple quantum filter and gradient ratio for artifact suppression were acquired with 256 increments in the F1 dimension, 8 number of scans and a relaxation delay of 2 s. A <sup>13</sup>C DEPT135 experiment with <sup>1</sup>H decoupling was acquired using a pulse length of 11.25  $\mu$ s and 8k number of scans.

## Chapter 4 – Deoxyfluorinated cellodextrin derivatives

The solution NMR characterisation of a 0.5 wt% dispersion of multi-6F-EpC (7) was carried out using a Bruker Avance III spectrometer equipped with a 5 mm inverse triple-resonance probe operating at frequencies of 800.2 MHz ( $^1\text{H}$ ) and 201.2 ( $^{13}\text{C}$ ). Phase-sensitive  $^1\text{H}$ - $^{13}\text{C}$  HSQC experiment with  $^1\text{H}$ - $^{13}\text{C}$  correlation via double inept transfer was acquired at 293K with 128 increments in the F1 dimension, 8 number of scans and a relaxation delay of 3s. A  $^1\text{H}$ - $^1\text{H}$  COSY experiment with multiple quantum filter and gradients was acquired at 5 °C with 256 increments in the F1 dimension, 12 number of scans and a relaxation delay of 2 seconds. Finally, a  $^{13}\text{C}$  DEPT135 experiment with  $^1\text{H}$  decoupling was acquired using a Bruker Avance Neo spectrometer equipped with a cryoprobe operating at frequencies of 600.2 MHz ( $^1\text{H}$ ) and 150.9 MHz ( $^{13}\text{C}$ ). A relaxation delay of 1 s was used and 30k scans were registered.

### Solid-state NMR

$^1\text{H}$ - $^{13}\text{C}$  cross-polarisation solid-state NMR experiments of the single 2-, 3- and 6- monofluorinated EpC 10 wt% hydrogels and multi-6F-EpC 25 wt% hydrogels were performed at 5 °C using a Bruker Avance III spectrometer equipped with a 4 mm triple resonance probe operating at frequencies of 400.2 MHz ( $^1\text{H}$ ) and 100.6 MHz ( $^{13}\text{C}$ ). Each gel was packed into a kel-f insert, sealed using a plug and a screw, and spun at 6 kHz. The WPT-CP experiment was carried out using a  $T_2$  filter and mixing time of 16 and 64 ms, respectively. The WPT factors shown in Figure 4.11 were calculated by normalisation of the peak intensities of the spectrum acquired at 16 ms and 64 ms mixing times against a reference spectrum at 0 ms mixing time.

For the powder samples, experiments were carried out at the UK 850 MHz solid state NMR facility (Warwick). All the experiments were run at frequencies of 850.2 MHz ( $^1\text{H}$ ), 799.8 MHz ( $^{19}\text{F}$ ) and 213.8 MHz ( $^{13}\text{C}$ ).  $^1\text{H}$ -decoupled  $^1\text{H}$ - $^{19}\text{F}$  CP ( $^1\text{H}$ - $^{19}\text{F}\{^1\text{H}\}$ ),  $^1\text{H}$ - $^{13}\text{C}$  CP with  $^1\text{H}$  and  $^{19}\text{F}$  decoupling ( $^1\text{H}$ - $^{13}\text{C}\{^1\text{H}, ^{19}\text{F}\}$ ), and  $^{19}\text{F}$ - $^{13}\text{C}$  CP with  $^1\text{H}$  and  $^{19}\text{F}$  decoupling ( $^{19}\text{F}$ - $^{13}\text{C}\{^1\text{H}, ^{19}\text{F}\}$ ) experiments were run using a 2.5 mm HFX H13894 probe, at 15 kHz MAS spinning and  $12 \pm 3$  °C. A field strength of 67 or 83 kHz was employed for  $^1\text{H}$  decoupling. The CP contact time was optimised to 1 ms and a relaxation delay of 3 s was employed. A  $90^\circ$  pulse of 4 and 3  $\mu\text{s}$  was used for  $^1\text{H}$  and  $^{19}\text{F}$ , respectively.  $^1\text{H}$ -undecoupled  $^{19}\text{F}$  spectra with  $^{19}\text{F}$  background suppression ( $^{19}\text{F}$ -bs), and  $^1\text{H}$ -decoupled and  $^{19}\text{F}$ -undecoupled  $^1\text{H}$ - $^{13}\text{C}$  CP ( $^1\text{H}$ - $^{13}\text{C}\{^1\text{H}\}$ ) experiments were run using a 1.3 mm X/Y/H-F H13863 probe, at 60 kHz MAS spinning and  $27 \pm 3$  °C. The CP contact time was optimised to 3 ms

## Chapter 4 – Deoxyfluorinated cellodextrin derivatives

and a relaxation delay of 3 s was employed. A 90° pulse of 1.5  $\mu$ s was employed for  $^1\text{H}$  and  $^{19}\text{F}$ . All spectra were referenced with respect to TMS.

## Chapter 5

Spin Diffusion Transfer Difference (SDTD) NMR:  
a novel methodological tool to study solvent in 3D  
gels network

## 5. SDTD: a novel methodological tool to study solvent in 3D gels network

Parts of the finding from this Chapter are published as:

Gabrielli Valeria, Kuraite Agne, Alves da Silva Marcelo, Edler Karen J., Angulo Jesús, Nepravishta Ridvan, \* Muñoz-Garcia Juan C.,\* Khimyak Y.\* - Spin Diffusion Transfer Difference (SDTD) NMR: An Advanced Method for the Characterisation of Water Structuration Within Particle Networks. *Journal of Colloid and Interface Science* (2021) 594 217–227. DOI: 10.1016/j.jcis.2021.02.094  
Introduction

### 5.1.1 Principal aims of the work

Carbohydrate-based hydrogels are 3D networks constituted by particles or polymers self-assembled in large amount of water. Understanding the molecular details of carbohydrate-water interactions is pivotal to comprehend gelation mechanisms. Historically, NMR relaxometry and solvent self-diffusion coefficient have been used to study the dynamics of the solvent and distinguish between its free and bound states. Nonetheless, the dynamic character of the interactions between the gel network and water at solid/solution interface might lead to the observation of an averaged value which reflects a variety of motionally-averaged molecular environments.

To overcome this limitation, we resort on Saturation Transfer Difference NMR spectroscopy. This technique allows the detection of the water directly interacting with the gel network (bound water) by monitoring the saturation that accumulates on the water molecules following selective saturation of the gel network. In addition, to prevent chemical exchange processes between labile protons of the gel network and the protons of water, we formulated our gels in pure D<sub>2</sub>O, limiting in this way the pool of protons available for exchange.

We have studied a set of colloidal dispersions and hydrogels composed by oxidized cellulose nanofibril (OCNF), corn starch (CS) and enzymatically produced cellulose (EpC). The measured STD signals report on the percentage of bound water, but are dependent

on gelator and solvent concentrations, precluding comparison between different systems and concealing information on solvent structure within the network.

We propose to resort to STD build-up curves normalisation against the maximum STD factor and mathematically analyse the resulting curves by means of a 1D diffusion equation. We demonstrate that this approach cancels the effect of gelator and solvent concentrations on the resulting build-up curves. The proposed analysis of the curves provides access to parameters such as the solvent-network minimum distance ( $r$ ) and the diffusion rate at the solid-liquid interface ( $D$ ).

In conclusion, the novel protocol, called Spin Diffusion Transfer Difference (SDTD) NMR, is a powerful tool for the study and comparison of different gel systems and opens the doors to novel insights into solvation properties.

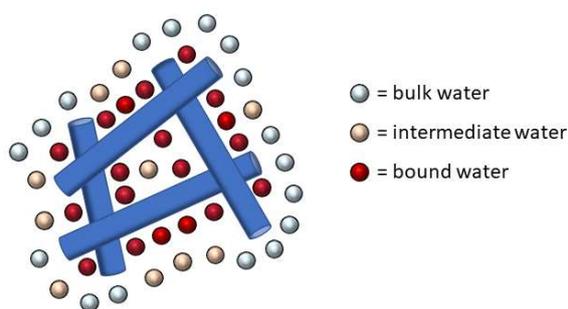
### 5.1.2 The role of water in hydrogels

As explained in *Chapter 1*, hydrogels are swollen 3D networks with water as main constituent. Hence, characterisation of the organisation of water on the surface of the gels solid component is pivotal to understand the gel physical properties and interfacial interactions. In biomaterial science biocompatibility of hydrogels is associated with the extensive presence of water,<sup>210</sup> and therefore a great deal of work has been devoted in investigating the nature of water.

In hydrogel systems a cooperative interaction exists between the water and the 3D network, the first preventing the network from collapsing, and the latter preventing the water from flowing away.<sup>211-213</sup> Water plays critical role in controlling not only hydrogels mechanical properties. It also acts as a diffusion barrier for drug release. Indeed, the rate of drug release depends on the diffusion rate of water into the gel matrix, forming a barrier gel layer, and the subsequent diffusion of water through this gel layer. Hence, distribution and mobility of water within the gel layer are important factors to be taken into account.<sup>214,215</sup>

Jhon and Andrade<sup>216</sup>, pioneers in this work, developed a three-state model according to which water can exist as 1) non-freezing, or bound water, 2) freezing interfacial, or intermediate water, and 3) free, or bulk water (Figure 5.1). Importantly, the nature and extent of the water-solid phase interaction will depend upon the gelator chemical structure. Indeed, water can have either a

plasticising or anti-plasticising effect<sup>217</sup> (namely, decrease and increase of the glass transition temperature), inducing major changes in the gelator mechanical properties. In turn, the physical-chemical properties of water (i.e. crystallization exotherms or melting endotherms) are affected by the interaction within the gel solid component. Indeed, bound water (as referred for the ensemble of non-freezing and freezing interfacial water) can either do not present any phase transition temperature, as in the case of non-freezing water, or it can present different transitions temperature and enthalpy in comparison with bulk water. The bound water is in a thermodynamically non-equilibrium state.<sup>218</sup>



**Figure 5.1:** Schematic representation of the different states of water in hydrogel system. The bound water (red circles) is strongly associated with the gel network by non-covalent interactions (hydrogen bonds or polar interactions); the intermediate water (pale yellow circles) is a transitory state between the bound and the bulk water and it is characterised by hydrophobic interaction with the macromolecule; the bulk water (light blue circles) is free water whose properties are not affected by the presence of the gel matrix.

The most common methodologies employed for water characterisation in hydrogels are differential scanning calorimetry (DSC) and NMR spectroscopy (water  $T_1$  and  $T_2$  relaxation times and water self-diffusion coefficient).<sup>211, 218, 219</sup> In performing DSC and NMR experiments, there are some important considerations to be taken into account to avoid misinterpretation of data. For instance, the experimental timescale should be considerably longer than the timescale of a thermal event to be probed (i.e. freezing of water, which is a quite slow process). Hence the rate of temperature change in DSC experiments should be appropriately tailored. Timescale of phenomena such as chemical exchange, spin diffusion and cross-relaxation are important for NMR measurements.<sup>219</sup>

Furthermore, attention must be paid in comparing data from different gel networks. Indeed, their individual character (i.e. polarity and hydrophilicity of the gelator molecules and morphology of the network), their self-assembly habit and thermal history are important factors to take into account.<sup>219</sup> Finally, results cannot be routinely extrapolated from one temperature regime to another, as changes in temperature would highly affect the equilibrium conditions of the interactions established to hold the gel system.<sup>215</sup>

In general, the presence of bound water to insoluble hydrophilic polymers has been reported to break down the hydrogen bonds established between the numerous hydroxyl groups present in the polymeric chain.<sup>218</sup> In this chapter we will answer some of the questions related to the state(s) of water in highly hydrophilic hydrogels, focusing on the interactions between water and cellulose (either unfunctionalised or oxidised cellulose nanofibrils, OCNF).

### 5.1.2.1 Cellulose and water interactions

Cellulose is the most abundant hydrophilic (but water-insoluble) polymer on Earth. Its high hydrophilicity determines cellulose ability to adsorb and entrap water. Indeed, water is able to bind and interact with the hydroxyl group exposed on the surface of the cellulose fibrils. *In turn*, the changes in water mobility due to these interactions determine a change in water physical properties, such as melting and crystallisation temperature.<sup>218</sup>

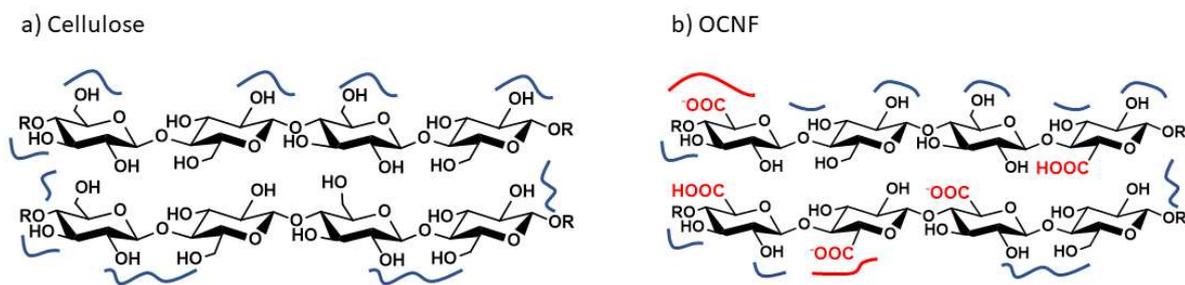
Molecular mechanics simulations showed short-range ordering of water on cellulose surfaces by a combination of both hydrogen-bonds and hydrophobic interactions with nonpolar surfaces.<sup>220</sup> O'Neill *et al.*<sup>221</sup> recently reported the existence of two different populations of water associated with cellulose, one more mobile and associated with the cellulose surface, the second more rigid and confined in the narrow spaces between fibrils. It is important to note that unfunctionalised cellulose fibrils present intrinsic structural anisotropy determined by (i) the high polarity of the lateral side in the glucopyranose ring due to the presence of equatorial hydroxyl groups, and (ii) the hydrophobicity of the axial orientation of the ring due to the presence of C-H bonds (see Chapter 1, Figure 1.12).<sup>222, 223</sup> The presence of hydroxyl groups on the fibrils surface and their consistent equatorial configuration determines structuring of the water molecules at the surface (Figure 5.2, a). In addition, MD simulations reported that only the dynamics of the surface hydrogen bonds between water and cellulose is affected by cellulose hydration, while the

hydrogen bonds in the bulk of cellulose remain unaffected. This indicates that the hydration-driven fast hydrogen bond relaxation is not transmitted from the surface to the core.<sup>224</sup> Hence, no drastic rearrangement of the fibril network is expected upon hydration.

A different situation is encountered when considering functionalised cellulose. In the presented study, we focused on oxidised cellulose nanofibrils (OCNF), where about 25% of the hydroxymethyl groups are substituted with carboxylic acid functional group. Importantly, the  $pK_a$  of OCNF was reported to be 3.9.<sup>225, 226</sup> Thus, in our experimental conditions (the pH of the analysed sample was measured as 7), about 10% of the carboxylic groups are protonated while the rest are negatively charged (Figure 5.2, b).<sup>225</sup> This results in repulsive forces among the cellulosic fibrils and favour dispersion in water.

SAXS studies at increasing concentration of OCNF fibrils showed that: (i) no interactions are established for OCNF fibrils concentrations below 1 wt%, (ii) between 1 wt% and 1.5 wt% repulsive interactions start emerging, and (iii) above 1.5 wt% repulsive interactions between the OCNF fibrils are evident.<sup>227</sup> In addition, rheological data demonstrated that measurable  $G'$  values were obtained only for concentrations above 1 wt%, while  $G' \gg G''$  only for samples of 2 wt% and above. Hence, the formation of gel was explained by the repulsion of the fibrils due to the negative charges, but no insight into the role played by water has been provided.

Previous studies on negatively charged hydrogelators described the ability of water to structure around the negative charges. For instance, a work published by Peppas *et al.* on poly(methacrylic acid) grafted with poly(ethylene glycol) at low ionic strength reported a higher swelling ratio for copolymers containing a higher amount of methacrylic acid, attributed to a larger electrostatic repulsion.<sup>228</sup> In a later work on the same copolymer, the authors showed that an increase in the solution ionic strength would have a shielding effect on the negative charges, hence reducing the repulsive forces contribution on swelling.<sup>229</sup>



**Figure 5.2:** Schematic representation of water interaction with different cellulose systems; a) binding of water around cellulose fibrils with ubiquitous distribution of hydroxyl groups on the surface; b) interactions of water with different functional groups present in OCNF nanofibrils. The blue and red lines help us to visualise differences in the interactions established by water and the different functional groups onto cellulose surface.

### 5.1.3 NMR for solvent investigation in gels

Traditionally, NMR investigation of solvent (i.e. water for hydrogels) in swollen systems relied on  $T_1$  and  $T_2$  relaxation times and solvent self-diffusion coefficient, providing insight into the dynamics and local environment of the solvent molecules. A large number of studies have been reported with applications to several different systems, and a full discussion is beyond the scope of this thesis. Nonetheless, it is important to understand the mechanism behind those methodologies, their possible uses and their inherent limitations. As hydrogels are the main interest of this thesis, from now on we will refer to water as the solvent in question.

#### 5.1.3.1 $T_1$ and $T_2$ relaxation times

As specified in *Chapter 2*,  $T_1$  and  $T_2$  relaxation constants are affected by the rate of molecular tumbling. Bulk water, fast and isotropic tumbling molecule, generally presents the same magnitude for proton  $T_1$  (spin-lattice) and  $T_2$  (spin-spin) relaxation constants, with  $T_1$  usually varying between 2 or 3 seconds depending on the amount of dissolved oxygen and/or other impurities.<sup>212</sup> On the contrary, in gels water not only presents  $T_1$  and  $T_2$  values different from each other, but they are also smaller in magnitude in comparison to the bulk water. This change in the relaxation properties depends on the water-network surface contact volume, on nature of the interactions (exchange, cross-relaxation) and the slower dynamics of bound water molecules.<sup>215, 230</sup>

## Chapter 5 – Spin Diffusion Transfer Difference (SDTD) NMR

In the simplest description of the swollen water-network system, water can be considered to exist in free and bound states, which are in exchange between each other. Two hierarchical processes govern this exchange: (i) the diffusion from the bulk to the gel surface, and (ii) the exchange with the molecule in the bound state. Importantly, exchange of water molecules or a chemical exchange of protons are both possible.<sup>212</sup> A stochastic theory of relaxation times for multiple phase systems has been developed which accounts specifically for the essential features of very slow and very fast exchange regimes.<sup>231</sup> For very slow exchange two distinct average nuclear correlation times exist, which give rise to two different longitudinal ( $T_1$ ) and transverse ( $T_2$ ) relaxation constants associated to each phase. On the contrary, when unique  $T_1$  and  $T_2$  values are found, it can be concluded to be in the fast exchange regime in relation to the NMR time scale between the water bound and free states.<sup>215</sup>

Two processes are involved in the levelling of the relaxation values: 1) chemical exchange (between distinct molecules of water or between the water and the exchangeable protons – OH and NH<sub>2</sub> – in the gel network) and 2) diffusion processes, such as physical diffusion of water and spin diffusion within proton spin system.<sup>219</sup> When chemical exchange is ineffective, different transverse relaxation times are observable as the spin diffusion process is slow on a  $T_2$  time scale.<sup>232</sup> On the contrary, spin diffusion equalises the longitudinal relaxation. In other words, the observation of a distribution of  $T_2$  values for water indicates the existence of water microenvironments large enough to cause an inefficient spin diffusion process on a spin-spin relaxation time scale.<sup>232</sup> Importantly, the distinction between mechanisms of spin diffusion in solids and liquids will be discussed in the spin diffusion subsection 5.1.5.1.

The presence of a slower  $T_2$  component has been attributed, in some cases, to the presence of structured water. This has been reported by Blinc *et al.*<sup>233</sup> in the study of collagen-water system in which a third water component (in addition to the bound and free water) was found and identified as "structural" water stabilizing the helical structure of collagen.

Importantly, the interpretation of  $T_1$  and  $T_2$  relaxation data can be hindered by the relaxation constants intrinsic limitations. If we look back at Figure 2.3, b in Chapter 2, we can clearly see that  $T_1$  presents symmetric behaviour at short and long correlation times. Hence, high  $T_1$  values might indicate either slow or fast dynamics regimes.<sup>69</sup> In addition,  $T_1$  represent the average contribution

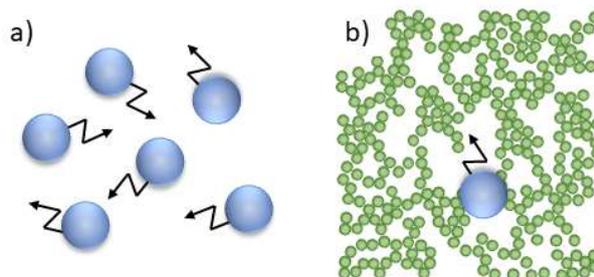
of both bulk and confined water molecules. On the contrary,  $T_2$  relaxation presents a monotonic dependence with molecular motion (slow tumbling molecules have shorter  $T_2$  time than fast tumbling molecules). Nonetheless, the measured value is strongly affected by chemical exchange.<sup>207</sup>

### 5.1.3.2 Self-Diffusion

Self-diffusion in solution (i.e. translational motion, as defined in Chapter 4) plays a central role in science reflecting not only the intrinsic properties of a biomacromolecule, such as its hydrodynamics, but also the features of the surrounding environment such as intermolecular dynamics or motional restriction.

When talking about translational molecular motion, the term diffusion is used to denote self-diffusion, mutual diffusion and 'distinct' diffusion, phenomena that share the same units ( $\text{m}^2\text{s}^{-1}$ ) and are related within each other, although being physically different: the first describes the random translational motion of molecules driven by their internal kinetic energy, the second the mass fluxes that arise when a concentration inhomogeneity exists between molecules in a multicomponent system, while the third has been proposed as direct measures of particle interactions in fluids.<sup>182</sup>

NMR methodologies are used for the measurement of self-diffusion, defined as the motion of molecules in a pure liquid at thermal equilibrium and depicted as a Brownian motion without any applied force (Figure 5.3).<sup>182</sup> Being a random process, the mean displacement,  $\langle r_1 - r_0 \rangle$  or  $\langle R \rangle$ , is equal to zero, although vicinal molecules are separated and hence dispersed. This phenomenon is defined by a self-diffusion coefficient  $D$  ( $\text{m}^2\text{s}^{-1}$ ) which range from *ca.*  $10^{-20} \text{m}^2\text{s}^{-1}$  for solids up to *ca.*  $1 \text{m}^2\text{s}^{-1}$  for dilute gases.<sup>182</sup>



**Figure 5.3:** a) Diffusion in the random thermal motion of the molecules, characterised by a diffusion coefficient  $D$  ( $\text{m}^2\text{s}^{-1}$ ); b) Brownian motion of a particle in a liquid formed by so small molecules that they effectively form a continuum with respect to the diffusing solute particle.

Source: Adapted from Price *et al.*, 2009<sup>182</sup>

Importantly, in a multicomponent system each of the components will present its own characteristic diffusion coefficient. In addition, the structure of a system in which another species is diffusing will restrict its motion and forbid an isotropic diffusion, hence affecting the rate of propagation and the pattern of its time-dependence.

Self-diffusion can be seen from a molecular level perspective, hence as a many-body effect in which each molecule is considered individually, or from a macroscopic level, in which the particles are taken as a continuum, reducing the phenomena to a one-body stochastic problem. Most of the experimental techniques work on a timescale sensitive to the macroscopic behaviour, with the self-diffusion coefficient  $D$  defined as (Equation 5.1)

$$D = \lim_{t \rightarrow \infty} \frac{1}{2t} \langle [r_i(t) - r_i(0)]^2 \rangle \quad \text{Equation 5.1}$$

where  $r_i(t)$  is the location of particle  $i$  at time  $t$  and  $\langle [r_i(t) - r_i(0)]^2 \rangle$  is the mean square displacement (MSD).<sup>182</sup>

### Self-Diffusion coefficient measurements - PGSE NMR:

NMR provides a unique method for the non-invasive measurement of translational motion. Self-diffusion measurements rely on spin-echo, at first in the presence of a static magnetic gradients (i.e., Steady Gradient Spin-Echo or SGSE NMR), following improved by the application of the magnetic gradients as pulses (i.e., Pulsed Gradient Spin-Echo NMR or PGSE NMR).<sup>182</sup> In addition, the most common sequences in PGSE NMR are the Hahn spin-echo sequence (SE) and the stimulated echo sequence (STE). In PGSE NMR the nuclei are spatially labelled using a magnetic field gradient pulse which imparts a position-dependent phase shift. A delay is elapsed before a second gradient pulse is applied to reverse the effect of the first. Diffusion of the molecules within the delay period leads to an incomplete refocusing and a reduced signal intensity. Typically, the

signal intensity is recorded as a function of the amplitude of the gradient pulse. The self-diffusion coefficient can then be extracted by fitting the data to the Stejskal-Tanner Equation:<sup>234</sup>

$$\ln\left(\frac{I}{I_0}\right) = -\gamma^2 \delta^2 D \left(\frac{\Delta - \delta}{3}\right) g^2 \quad \text{Equation 5.2}$$

where  $D$  is the diffusion coefficient,  $\gamma$  is the nuclei gyromagnetic ratio,  $\delta$  is the gradient length,  $g$  the gradient strength and  $\Delta$  the diffusion time.

The application of magnetic field gradients allows to spatially differentiate the NMR signal, as nuclei in different section of the sample will experience a different magnetic field and hence precess at different frequency.<sup>235</sup>

### 5.1.4 STD NMR applied to hydrogel systems

*Chapter 2* presents STD NMR methodology in detail. Briefly, for macromolecules undergoing Brownian motion in solution, STD NMR consists of the selective saturation of the macromolecule followed by saturation spread throughout the whole macromolecule *via* intramolecular and intermolecular NOE. Subsequently, the magnetisation is transferred to fast-exchanging binders by intermolecular NOE.

In soft matter, STD NMR has been used to study both gelator and solvent molecules. The first application of STD NMR to study water in gels system was reported by Yadama-Nosaka *et al.*, in the study of a hydrophilic poly(hydroxyethyl methacrylate) and a hydrophobic of poly(methyl methacrylate) hydrogels.<sup>236</sup> Specifically, they classified the water contained in these membranes as bound, intermediate and free water. The bound water was claimed to be too rigid to be detected by NMR (signal too broad), hence the detected NMR water signal was attributed to intermediate and free water. By selective saturation 20 ppm up-field of the observed water signal, reduction of water signal intensities was recorded for both hydrogels and attributed to saturation transfer via fast-exchange between observable and unobservable water.

About 15 years later Mahajan *et al.*, reported the use of STD NMR to monitor the site-specific interactions between water and gelator molecules within gels formed by pyrene-grafted glutathione ( $\gamma$ -glutamyl-cysteinyl-glycine; GSH).<sup>237</sup> The water peak was selectively saturated and

## Chapter 5 – Spin Diffusion Transfer Difference (SDTD) NMR

the transfer of magnetisation from water to the GHS-pyrene protons monitored in a sample in the *process of gelling* (hence, in viscous fluid 'gels' state to reduce line broadening). Thus, qualitative indication of which protons are affected upon gelation was achieved. Lozano *et al.*, applied STD NMR to study asparagine/tryptophan organogel.<sup>53</sup> Importantly, they reported that setting the on-resonance frequency at -1 ppm, a reduced peak intensity was recorded for both the gelator and the solvent resonances. This demonstrated the existence of a rigid gel solid phase hidden below the baseline due to fast transverse relaxation rate.

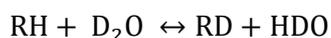
In addition to its role to unveil solvent-network interactions, STD NMR has been used to monitor the interaction of other substrates within the gel network. Segarra-Maset *et al.*,<sup>238</sup> investigated the interaction of dopamine and related substances with amino acid-based hydrogels and obtained relative affinity values by competition experiments either with free gelator molecules or with an added substrate.

A literature research with Google Scholar and Web of Science revealed that, in the last 5 years, only 6 papers have been published with STD NMR applied for the study of gels, and all of them from our research group at UEA, with the exception of the work carried by Wallace *et al.*<sup>239</sup>, University of Liverpool (currently also at UEA). Ramalhelte *et al.* proved STD NMR a valuable tool to assess exchange phenomena at the gel/solution interface in supramolecular amino acid based hydrogels.<sup>240</sup> Wallace *et al.* focused instead in the indirect characterisation of fibre surface properties (i.e. charge, hydrophobicity and ion-binding dynamics) in *N*-functionalised dipeptides hydrogels and used STD NMR to characterise interactions of probe molecules to the gel surface.<sup>239</sup>

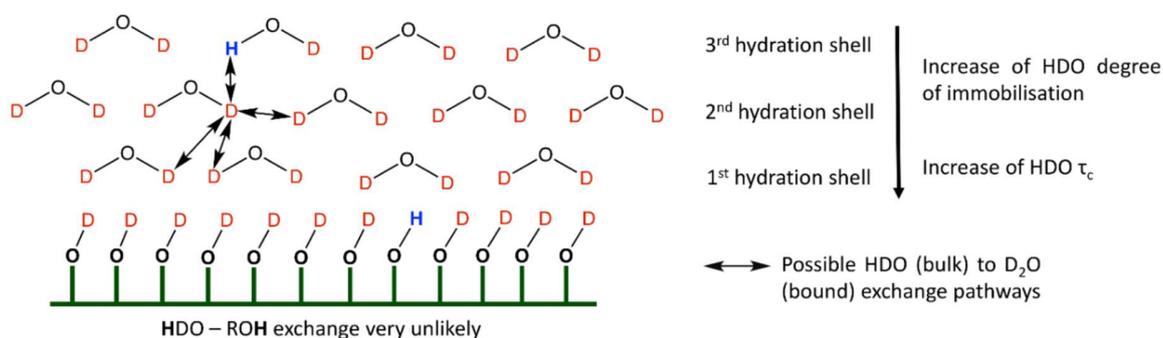
Finally, the work published by Calabrese *et al.* is the first example of STD NMR applied to the investigation of the state of water in carbohydrate-based hydrogels and it is an important starting point for our study.<sup>207</sup> Briefly, the work focused on the study of heat-induced gelation in TEMPO-oxidised cellulose hydrogels (OCNF - 1wt% and 4 wt% concentrations). These were compared with cationic cellulose hydrogels (CCNF - 1wt% and 4 wt% concentrations), which did not present any changes in their physical properties upon temperature raise.

STD NMR was measured by selective saturation of the polysaccharide fibrils network at 1 ppm and the saturation transfer to the solvent monitored at 4 seconds saturation time and variable temperature. Importantly, the gels were prepared in D<sub>2</sub>O. This detail is of extreme importance for

two main reasons: (i) the recorded solvent signal in the spectra belongs to HDO, derived from the chemical exchange process which takes place from D<sub>2</sub>O and exchangeable protons from OH and NH<sub>2</sub> functional groups:



The signal is proportional to the number of exchangeable protons present in the system; (ii) the low presence of <sup>1</sup>H nuclei strongly limits the contribution of chemical exchange to the saturation transfer from the gel fibrils to the solvent (Figure 5.4).



**Figure 5.4:** For gels in D<sub>2</sub>O the contribution from direct transfer through chemical exchange between HDO and the hydroxyl group of cellulose can be discarded as most of the cellulose hydroxyl group will be deuterated. Minimising the contribution of chemical exchange to the apparent STD factor of water, a more accurate estimation of the population of network-bound water is achieved. *Source: Calabrese et al., 2019<sup>207</sup>*

### 5.1.5 A new methodological tool: Spin Diffusion Transfer Difference (SSTD) NMR

We rationalised a strong dependence of STD NMR absolute signal on gelator and solvent concentrations. For hydrogels formulated in D<sub>2</sub>O, the final concentration of HDO in the sample depends on the gelator chemical nature, on its concentration and on the formulation conditions (e.g. environmental humidity, exposure to air, purity of D<sub>2</sub>O - just to mention few). All those factors determine the presence of an unpredictable concentration of HDO in the sample, which in turn

would affect the STD signal recorded, hence precluding comparison between different sets of hydrogels.

In addition, after selective saturation of the gelator network, the magnetisation will spread among the network components at different rate: the stronger dipole-dipole interactions established by thicker, more solid-like gels determine a faster magnetisation spreading in comparison to softer gels. This precludes comparison between different concentrations of the same materials, as well as comparison between gelators which presents different thickening effects.

Three mechanisms can be ascribed to the gel network - water magnetisation transfer observed: (i) proton exchange, (ii) NOE and (iii) spin diffusion. Following formulation of gels in D<sub>2</sub>O the proton exchange mechanism has a lower statistical probability, while NOE is expected to be less effective in anisotropic conditions, hence spin diffusion is the major active component for magnetisation transfer from the fibrils to the bound HDO.<sup>241-243</sup>

### 5.1.5.1 Spin Diffusion

Spin diffusion is a cross-relaxation phenomenon, which causes a mutual exchange of spin magnetisation at a rate larger than the spin-lattice relaxation of the protons, and can be depicted as the recovery of spatially uniform magnetisation in a system removed from a state of thermodynamics equilibrium.<sup>241, 243</sup> In solution state, multiple spin systems relaxation theory usually ignores the correlation between dipolar coupling and motion of a proton pair, simplification acceptable for systems showing an overall isotropic rotational diffusion.<sup>243</sup>

Nonetheless, in solids and gel-like samples (e.g. particulate gels) the dipolar coupling is not averaged out by the isotropic tumbling of the molecules (we are in the presence of anisotropic interactions). Thus, the spin diffusion phenomenon is not related to molecular motion but to a coherent effect due to the <sup>1</sup>H-<sup>1</sup>H dipolar interactions.<sup>244,245</sup> For this reason, NOE-based solid-state NMR pulse sequences are often referred to as spin diffusion, as the incomplete MAS averaging of the strong homonuclear dipolar coupling determines spin diffusion to be the dominant mechanism of magnetisation transfer in solids. To avoid confusion, both definitions of spin diffusion are usually classified as coherent spin diffusion (solids) and incoherent spin diffusion (liquids).

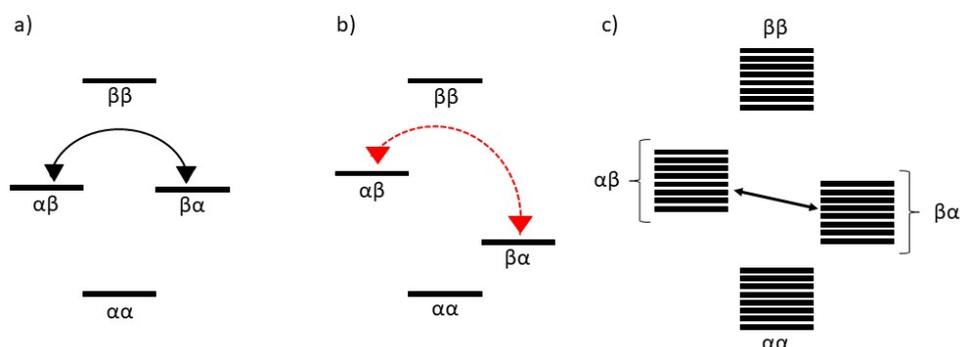
## Chapter 5 – Spin Diffusion Transfer Difference (SDTD) NMR

In semisolid samples (e.g. gels), both coherent and incoherent spin diffusion can play a role, although the contribution of coherent spin diffusion is expected to be much larger than the incoherent mechanism. In this work, when we use the term spin diffusion, we will be referring to coherent spin diffusion.

Two factors contribute to the spin diffusion coefficient,  $D_{sp}$  (Equation 5.3): (i) the intermolecular flip-flop process ( $D_f$ ) and (ii) the translational displacement of the molecules ( $D_{tr}$ ).<sup>241, 246, 247</sup>

$$D_{sp} = D_f + D_{tr} \quad \text{Equation 5.3}$$

The intermolecular flip-flop process is a spontaneous phenomenon when the energy levels of two spins connected by dipolar coupling (e.g. the  $\alpha\beta$  and  $\beta\alpha$  levels in a two spins  $I=1/2$  system) have the same energy (Figure 5.5, a), although this is rarely the case for isotropically tumbling molecules in solution (Figure 5.5, b). The situation is different for solids and solid-like systems, where the existence of a dense network of dipolar coupled nuclear spins causes overlapping between some of the  $\alpha\beta$  and  $\beta\alpha$  levels (Figure 5.5, c). In this case the flip-flop transitions between these levels is energy conserving and spin diffusion occurs.<sup>248</sup> Importantly, protonated solids and solid-like systems present significant  $^1\text{H}$ - $^1\text{H}$  dipolar couplings compared to the chemical shift differences,<sup>247, 248</sup> hence those dipolar couplings are the driving force for spin diffusion to occur.<sup>247, 248</sup> In addition, characteristic times of these intermolecular flip-flop transitions ( $\tau_f \gg 10^{-2} - 1$  s, where  $\tau_f$  is the mean flip-flop time of a dipolar coupled spin pair) are greater for colloidal dispersions because of the higher mobility of the molecules in liquid.<sup>241</sup>



**Figure 5.5:** Energy levels for: (a) two equivalent dipolar coupled spins with degenerate  $\alpha\beta$  and  $\beta\alpha$  energy levels. The energy-conserving “flip-flop” transitions is allowed, and cross-correlation occurs; (b) two inequivalent dipolar coupled spins, hence the transition is not energy conserving

## Chapter 5 – Spin Diffusion Transfer Difference (SDTD) NMR

and its probability is low, and (c) two inequivalent spins coupled to (many) other spins. Broadening of the two-spin system energy levels causes overlap between some of the  $\alpha\beta$  and  $\beta\alpha$  levels, hence cross-correlation has high probability and spin diffusion occurs. *Source:* adapted from Emsley *et al.*, 2009.<sup>248</sup>

In the case of macromolecular solutions and high molecular weight particles dispersions,  $\tau_f$  shows a molecular mass dependence, i.e. it initially decreases with the increase of the MW and becomes independent after a critical molecular mass  $N^*$  (hence, it reaches a constant value), as expressed by Equation 5.4:

$$\tau_f = \tau_s N^{*3.4} \quad \text{Equation 5.4}$$

where  $\tau_s$  is the segmental relaxation time ( $10^{-11}$  -  $10^{-9}$  s) and  $N$  represent the number of Kuhn segments per macromolecule. Importantly, the relationship between the spin diffusion coefficient  $D_{sp}$  and the self-diffusion coefficient  $D_{sd}$  is dependent on the relationship between  $N^*$  and the molecular mass of the dispersed particles, with  $D_{sp} \approx D_{sd}$  for  $N \ll N^*$ , and  $D_{sp} \gg D_{sd}$  for  $N \gg N^*$  (as  $\tau_f$  of the macromolecules resembles anomalous diffusion). Indeed, for  $N \ll N^*$  the rate of flip-flop transition is slow in comparison with the macromolecule relaxation time and the levelling of the spatially distributed magnetisation lines up with the macromolecule self-diffusion. On the contrary, when  $N \gg N^*$  the flip-flop transition is faster than relaxation, hence spatially distributed magnetisation will equilibrate at times  $\tau_f$ , while the macromolecule segments will self-diffuse with a longer time, hence  $D_{sp} \gg D_{sd}$ .<sup>241</sup>

Fisher, Kimmich, Fatkullin<sup>247</sup> and Yatsenko<sup>246</sup> measured intermolecular spin diffusion process *via* field gradient NMR diffusion in the steady gradient of a superconducting fringe field. For example, Fisher *et al.*,<sup>247</sup> measured spin diffusion in a polyethylene oxide sample by division of the stimulated-echo amplitude attenuation derived from experiments performed with and without field gradient. In the first experiment (with field gradient, where the sample was subjected to the magnet fringe field) the obtained echo-attenuation depended on the overall contribution of transverse relaxation, spin–lattice relaxation, dipolar correlation, and diffusion. In the second experiment the sample was placed in the centre on the corresponding magnet and therefore subjected to a homogeneous field. Consequently, the echo-attenuation was governed by relaxation and dipolar correlation effects only, with no contribution of translational diffusion.

## Chapter 5 – Spin Diffusion Transfer Difference (SDTD) NMR

In addition, solid-state NMR experiments on membrane proteins and plant cell wall demonstrated the effectiveness of intermolecular magnetization transfer at the solid-liquid interface either from a mobile<sup>249</sup> or rigid<sup>250</sup> source *via* spin diffusion.

It was reported that the magnetization time-space evolution via spin diffusion is described by a Gaussian displacement probability<sup>251-253</sup> and follows a macroscopic 1D diffusion equation.<sup>247, 251-253</sup> The selective saturation applied to the gelator network creates a nonuniform magnetization profile throughout the sample, which is then transferred to the HDO molecules in proximity with the gelator 3D network with a length scale evolution on the order of the root mean square displacement during the saturation time ( $\Delta z_{rms} = \sqrt{Dt_{sat}}$ ).<sup>251</sup> The overall time required to the magnetisation to achieve spatial equilibration depends on the morphology of the molecular entity, hence on its proton-proton distances and protons density.

Spin diffusion kinetics is described by Fick's 2<sup>nd</sup> law of diffusion

$$M(r, t) = \nabla[D(r)\nabla M(r, t)] \quad \text{Equation 5.5}$$

Where  $\nabla$  is the Laplace operator,  $D$  is the diffusion gradient,  $r$  is the space vector,  $t$  is the diffusion time and  $M(r, t)$  is defined as the ratio between the z-magnetisation  $m(r, t)$  and the mass fraction of protons  $m_H(r)$ , which *in turn* depends on the proton density  $Q_H$  and on the molecular entity total volume  $V_{tot}$

$$M(r, t) = m(r, t)/m_H(r) = m(r, t)/(Q_H, V_{tot})(r) \quad \text{Equation 5.6}$$

The solution of the diffusion equation for a point source is the Gaussian function:

$$M(r, t) = (M_0/4\pi D_t)e^{-r^2/4Dt} \quad \text{Equation 5.7}$$

While for an infinite solid the error function of the Gaussian function can be a solution for the diffusion equation

$$M(r, t) = 1/2 M_0 \operatorname{erfc}(r - r_0/\sqrt{4Dt}) \quad \text{Equation 5.8}$$

Let us consider a two-phase system A and B in which the nonequilibrium spatial distribution of the magnetisation is achieved by selective saturation of protons of phase A, and the magnetisation diffusion to phase B is detected. The diffusion of magnetisation for each phase is described by the error function

$$M_A(r, t) = E_A + F_A \operatorname{erfc}(r - r_0 / \sqrt{4D_A t}) \quad \text{Equation 5.9}$$

$$M_B(r, t) = E_B + F_B \operatorname{erfc}(r - r_0 / \sqrt{4D_B t}) \quad \text{Equation 5.10}$$

where  $E_A$  and  $E_B$  represent the spin populations in phase A and phase B, while  $F_A$  and  $F_B$  represent the field experienced by in phase A and phase B, respectively. For  $t = 0$

$$E_A + F_A = M_{A,0} \quad \text{Equation 5.11}$$

$$E_B + F_B = M_{B,0} = 0 \quad \text{Equation 5.12}$$

Using the interface condition that  $E_A = E_B$  and the flux equilibrium at the interface  $j_A(r_0, t) = j_B(r_0, t)$  it can be shown that

$$M_B(r, t) = (M_{A,0} \sqrt{D_A Q_{HA}} / \sqrt{D_A Q_{HA}} + \sqrt{D_B Q_{HB}}) \operatorname{erfc}(r - r_0 / \sqrt{4D_B t}) \quad \text{Equation 5.13}$$

And in the conditions  $D_A \gg D_B$  Equation X can be simplified to

$$M_B(r, t) = M_{A,0} \operatorname{erfc}(r - r_0 / \sqrt{4D_B t}) \quad \text{Equation 5.14}$$

### 5.1.5.2 Transforming the STD data

By acquiring experiments at increasing saturation times, the build-up curve of the rigid-to-mobile magnetisation transfer is obtained. For the experiment to be successful, the following conditions must be fulfilled:

1. The small molecule (HDO) in phase B is in fast exchange conditions between the free and the bound state with phase A, so that the overall exchange constant (Equation 2.30, Chapter 2) is high due to a  $k_{on}$  that can be considered at the diffusion limit and a high  $k_{off}$  constant, so that  $k_{ex} \gg D_{interface}$ ;
2. Hence, the half-life time of the small molecule/macromolecule instantaneous interaction is expected to be low compared to the interfacial diffusion time. Therefore, several cycles of association-dissociation of the small molecule will take place before the magnetization can be efficiently transported from phase A to phase B, where it will be detected.
3. Free HDO molecules in phase B present almost no differences between the spin diffusion and the self-diffusion coefficient ( $D_{sp} \approx D_{sd}$ , a condition that we have already encountered in section 5.1.4.1). The small molecules, once received the magnetisation interacting with the macromolecule (phase A), will maintain it for long time. Indeed, the

## Chapter 5 – Spin Diffusion Transfer Difference (SDTD) NMR

process of relaxation through spin diffusion for small molecules is highly inefficient and therefore they must relax through different mechanisms (i.e. longitudinal relaxation). This will create a new nonequilibrium spatial distribution of magnetization along the protons of phase B, described by Equation 5.14. Spin diffusion at the interface is the slowest and rate-limiting process and it dominates the kinetics of magnetisation transfer in phase B. Hence, it is possible to substitute  $D_B$  within  $D_{interface}$  in Equation 5.14.

4. Finally  $D_{interface}$  can be easily obtained experimentally by changing the time of saturation in  $^1\text{H}$  NMR STD pulse sequence, selecting specific protons of phase A and detecting the diffusion of magnetization in the phase B through the proportionality  $M_B \propto I/I_0$ .

Point 4 defines that it is necessary to relate them to the equilibrium value for  $t_{sat} \rightarrow \infty$  by scaling the data against the highest obtained STD signal. Hence, we perform a standard normalisation using the highest STD signal as a scaling factor to finally obtain the Spin Diffusion Transfer Difference (SDTD) values. The SDTD values are then plotted against the square root of the saturation time ( $\sqrt{t_{sat}}$ ) and the obtained spin diffusion build-up curve can be simulated numerically *via* mathematical fit to the 1D diffusion equation<sup>252, 253</sup> (Equation 5.15):

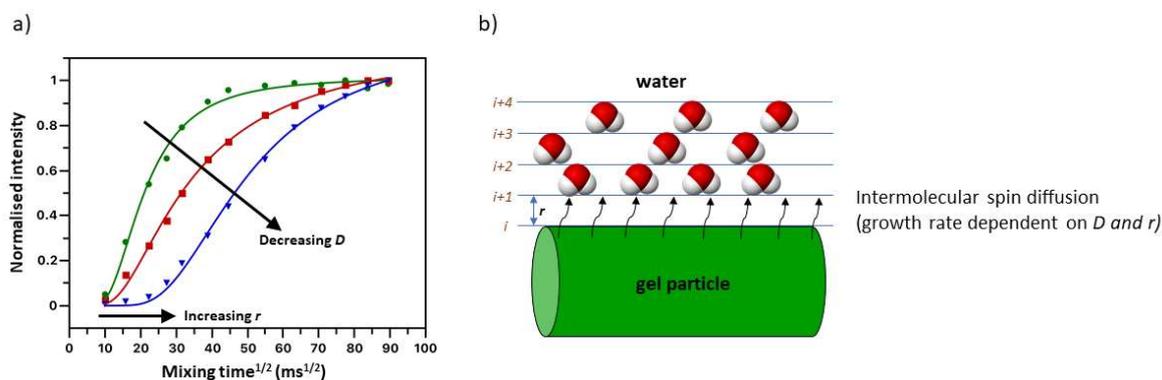
$$SDTD = C \cdot \text{erfc} \left[ \frac{r}{2 \cdot \sqrt{D \cdot t_{sat}}} - b \right] \quad \text{Equation 5.15}$$

where SDTD is the normalised STD intensity values, C is the proportionally constant of the fit, erfc is the complementary error function, r is the minimum distance between the gel network and the HDO molecules (expressed in nm), D is the spin diffusion coefficient at the solid-liquid interface (expressed in  $\text{nm}^2/\text{ms}$ ),  $t_{sat}$  is the saturation time (expressed in ms) and b is a correction factor to centre the gaussian distribution. In this way, for each saturation time  $t_{sat}$  one can determine the amount of magnetisation which has diffused from the gelator network to the surrounding water. In addition, this approach allows comparison of the speed of growth for different STD curves.

Through comparison of different curves, information can be inferred. Two factors determine the shape of the SDTD curves (Figure 5.6, a): (i) the minimal distance in the solvent-network r (Figure 5,6, b) determines the initial part of the curve, also called lag phase. The longer is r, the greater is the lag phase. To obtain a correct fit of the curve and sample the lag phase, it is necessary to perform the experiments starting from shorter saturation times than in conventional

## Chapter 5 – Spin Diffusion Transfer Difference (SDTD) NMR

STD (0.05 seconds and above); and (ii) the spin diffusion coefficient  $D$  determines the slope of the curve. The larger the value, the higher the slope. Importantly, both  $r$  and  $D$  are parameters related to the degree of solvent structuration within the gel network, as faster spin diffusion rates  $D$  and shorter distances  $r$  reflect increased solvent structuration.



**Figure 5.6:** a) Schematic representation of the parameters affecting the SDTD curves shape; b) grid approximation of the spin diffusion model.

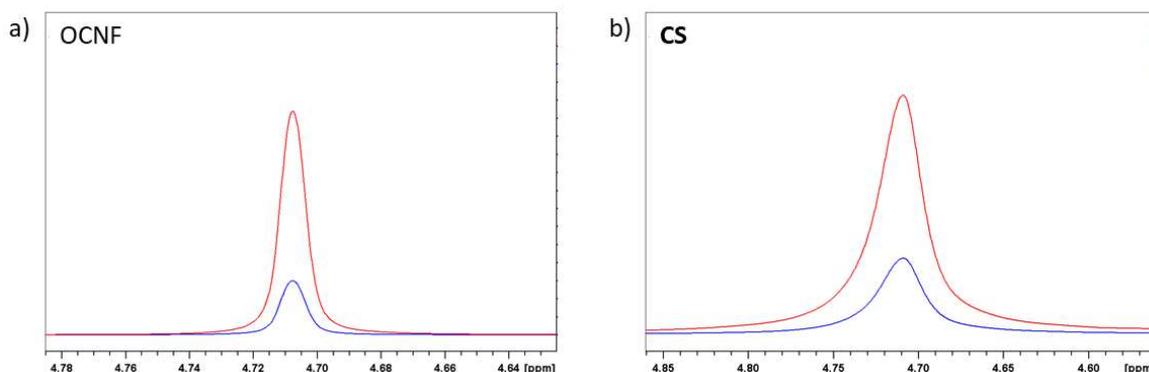
Mathematical fit of the SDTD curve is performed by variation of either the distance  $r$  or the spin diffusion coefficient  $D$  parameters, while maintaining the other parameter constant. Throughout this work, the value  $r$  was fixed to 0.2 nm in accordance to previous studies reporting on water-particle interfaces.<sup>249</sup> Since spin-diffusion is primarily determined by the rate-limiting steps of magnetization transfer across the gelator-water interface, the spin diffusion build-up curves are largely independent from the thickness of the hydration shell. Hence, only the closer bound water is represented by the SDTD curves.

## 5.2 Experimental section

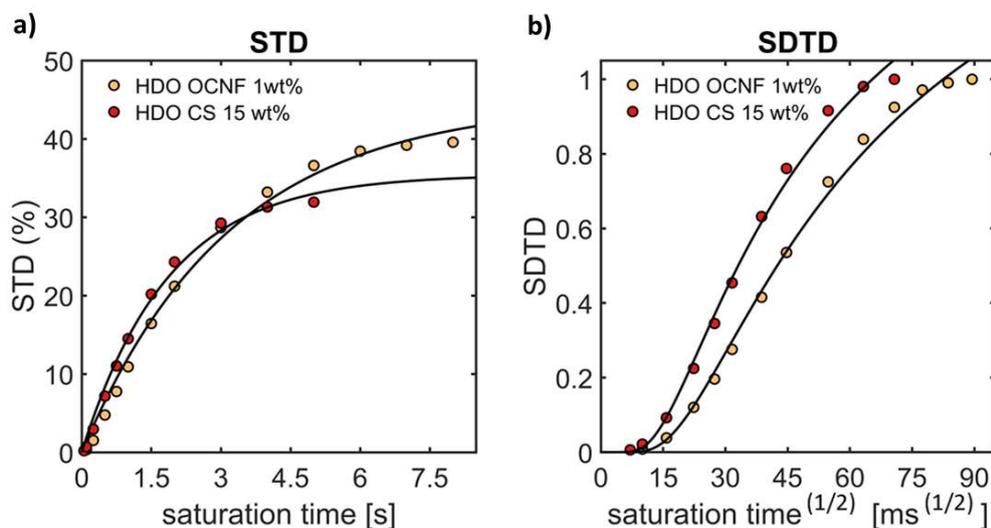
This chapter presents the results of a systematic study on the experimental conditions affecting the recorded absolute STD values and in the development of a novel STD NMR methodology, which we named Spin Diffusion Transfer Difference (SDTD) NMR. The obtained data (SDTD curves) can be modelled by the 1D diffusion equation (Equation 5.15), which describes the “rigid-to-mobile” magnetisation transfer *via* spin-diffusion that occurs between the gel network and the interacting water molecules. The SDTD data are important to explore changes in solvent structuration in different systems and to establish correlations between solvent structuration and material properties.

### 5.2.1 Model validation

The modelling of the spin diffusion kinetics for large particles dispersed in solution (i.e. of very slow rotational and translational diffusion) by Equation 5.15 has been validated experimentally by monitoring the evolution of the HDO peak of two carbohydrate-based dispersions in D<sub>2</sub>O: (i) a liquid-like OCNF 1 wt% dispersion, and (ii) a viscous corn starch (CS) 15 wt% gel (Figure 5.7 a and b). STD NMR experiments at different  $t_{sat}$  ranging from 50 ms to 8 s and 5 s for OCNF and CS, respectively (Figure 5.8, a) were recorded. Normalisation of the obtained STD intensities against the highest STD value was performed, the resulting data were plotted against the  $t_{sat}$  square-root and fit to Equation 5.15 (Figure 5.8, b). Several values of the  $b$  parameter were tested to reach the best fit. In this regard, it should be noted that, as the parameter  $D$  is dependent on the  $b$  value used during the fit, only  $D$  values obtained from curve fits carried out using the same  $b$  can be compared. Thus, when comparing SDTD curves for which different  $b$  values give the best fit, a compromise value must be chosen to determine  $D$ .



**Figure 5.7:** Off-Resonance (red) and STD (blue) HDO peak for a) OCNF 1wt% and b) CS 15wt% at 5 s saturation time and 298 K. *on-resonance* frequency of -1 and 0 ppm for OCNF and CS, respectively, and *off-resonance* frequency of 50 ppm.



**Figure 5.8:** a) STD and b) SDTD build-up curves of the HDO peak for the OCNF 1 wt% dispersion (pale yellow) and corn starch (CS) 15 wt% gel (red) prepared in D<sub>2</sub>O. Solid lines are fits to Equation 5.15.

The comparison of the SDTD build-up curves of OCNF 1 wt% and CS 15 wt% shows the faster spin diffusion growth at the CS-water interface (Figure 5.8, b and Table 5.1), indicating that, as expected, the degree of structuration of water is significantly higher in the viscous CS gel compared to the liquid-like OCNF dispersion. Table 5.1 reports the parameters used to fit the SDTD curves to Equation 5.15 and the resulting  $D$  values.

**Table 5.1:** Calculated values for the  $C$  and  $D$  parameters obtained from the fit to Equation 5.15 of the SDTD build-up curves of the HDO peak for the OCNF 1 wt% and CS 15 wt% dispersions.  $b$  and  $r$  parameters were fixed at a value of 1 and 0.2 (nm), respectively. The errors associated to each  $C$  and  $D$  value are shown and correspond to the 99% confidence level. The  $R^2$  values of each fit to Equation 5.15 are shown.

	OCNF 1 wt%	CS 15 wt%
$C$	1.14	1.12
<i>error</i>	$\pm 0.13$	$\pm 0.14$
$D$ (nm <sup>2</sup> /ms)	$9.80 \times 10^{-5}$	$1.28 \times 10^{-4}$
<i>error</i>	$\pm 1.13 \times 10^{-5}$	$\pm 1.70 \times 10^{-5}$
$R^2$	0.9951	0.9952

### 5.2.2 The effect of solvent concentration on STD and SDTD build-up curves

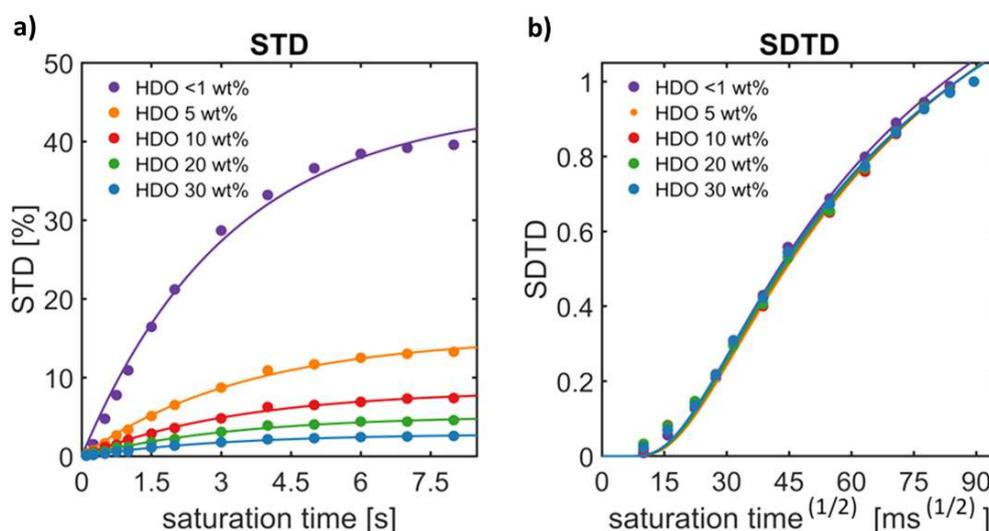
As it is mentioned above, the chemical exchange mechanism for saturation transfer is limited by the preparation of the dispersions/gels in pure D<sub>2</sub>O, hence limiting the HDO concentration to the proportion of exchangeable protons originating from the gelator molecules. Nonetheless, it is extremely difficult to maintain the HDO concentration under precise control due to its dependence on (i) the <sup>2</sup>H purity of the batch of deuterated solvents employed, (ii) the relative humidity of the environment and air exposition, and (iii) the solid content of the gel and (iv) the chemical structure of the gelator.

In this regard, it is important to note that the STD NMR signal is proportional to the fraction of ligand bound, henceforth referred as fraction of bound water ( $f_{WB}$ ). As expressed in *Chapter 2*, Equation 2.36, the fraction of ligand bound is inversely proportional to the total concentration of ligand (HDO in the presented case). Therefore, an increase of the total HDO concentration ( $[HDO]_{TOT}$ ) determines a reduction in  $f_{WB}$  (as expressed by Equation 5.16, where the total HDO concentration -  $[HDO]_{TOT}$  - appears in the denominator) and *in turn* a decrease in the recorded STD factor.

$$f_{WB} = \frac{[Network - HDO]}{[HDO]_{TOT}} \quad \text{Equation 5.16}$$

To investigate the effect of  $[HDO]_{TOT}$  on the SDTD build-up curve, we carried out H<sub>2</sub>O titrations (used concentrations: below 1 wt%, 5 wt%, 10 wt%, 20 wt% and 30 wt%) to OCNF 1 wt% dispersions prepared in D<sub>2</sub>O. Comparison of the STD (Figure 5.9, a) and SDTD (Figure 5.9, b) build-up curves for HDO binding to OCNF particles shows that, when we applied the SDTD methodology to these data, the effect of HDO concentration on the observed SDTD values was cancelled out, resulting in the overlapping curves for all the HDO concentrations sampled. Importantly, in the considered experimental conditions the contribution of the chemical exchange phenomenon to the magnetisation transfer can still be neglected.

The mathematical fit the Equation 5.15 yielded C and D parameters, reported in Table 5.2.



**Figure 5.9:** a) STD and b) SDTD build-up curves for HDO binding to OCNF 1 wt% dispersion acquired at different H<sub>2</sub>O/D<sub>2</sub>O ratios. The H<sub>2</sub>O concentrations used range from < 0.1 wt% (purple), 5 wt% (yellow), 10 wt% (red), 20 wt% (green) and 30 wt% (light blue). A *b* value of 1 was used to obtain the best fit for the EpC SDTD curves.

**Table 5.2:** Calculated values for the  $C$  and  $D$  parameters obtained from the fit to Equation 5.15 of the SDTD build-up curves of the HDO peak for OCNF 1 wt% dispersions prepared with different concentrations of  $H_2O$ . An  $r$  value of 0.2 nm and a  $b$  value of 1 were kept constant during the fit. The errors associated to each  $C$  and  $D$  value are shown and correspond to the 99% confidence level. The  $R^2$  values of each fit to Equation 5.15 are shown.

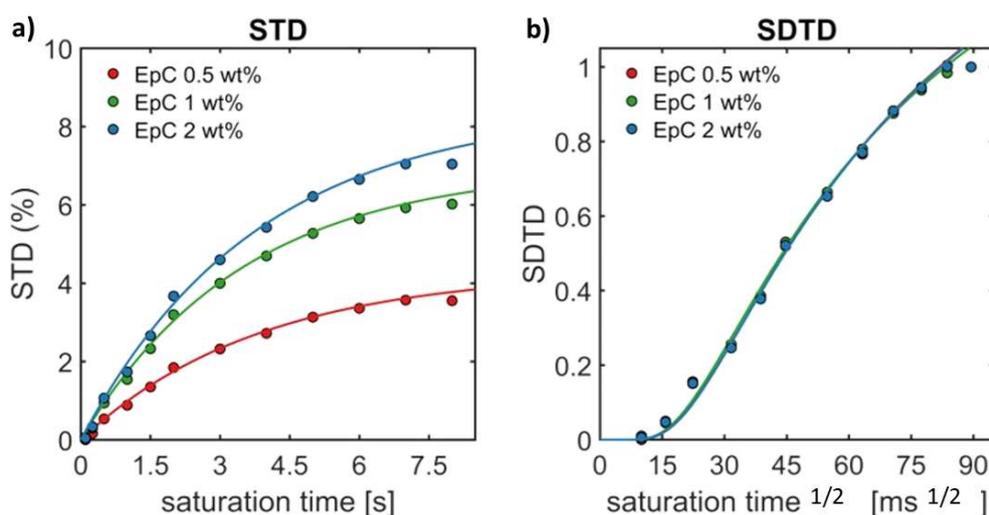
<b>HDO</b>	<b>&lt;1%</b>	<b>5%</b>	<b>10%</b>	<b>20%</b>	<b>30%</b>
<b><math>C</math></b>	1.14	1.15	1.13	1.13	1.11
<b>error</b>	$\pm 0.13$	$\pm 0.11$	$\pm 0.15$	$\pm 0.18$	$\pm 0.12$
<b><math>D</math> (<math>nm^2/ms</math>)</b>	$9.80 \times 10^{-5}$	$9.36 \times 10^{-5}$	$9.64 \times 10^{-5}$	$9.66 \times 10^{-5}$	$9.93 \times 10^{-5}$
<b>error</b>	$\pm 1.13 \times 10^{-5}$	$\pm 9.35 \times 10^{-6}$	$\pm 1.31 \times 10^{-5}$	$\pm 1.59 \times 10^{-5}$	$\pm 1.12 \times 10^{-5}$
<b><math>R^2</math></b>	0.9951	0.9963	0.9930	0.9897	0.9952

### 5.2.3 The effect of gelator concentration on STD and SDTD build-up curves: *SDTD reports on changes in the degree of solvent structuration*

The solution-to-gel transition which characterise the gelation process is generally dependent on the gelator concentration. In the STD NMR context, it should be noted that an increase in the gelator concentration determines a change in the number of binding sites available for the solvent, an increase in the  $f_{WB}$  and, *in turn*, higher STD factors. In addition, the stronger dipolar interactions in more viscous or stiffer network boost spin diffusion and hence a higher amount of magnetisation will be transferred from rigid to mobile components. The combination of these factors hinders the understanding of solvent structuration *via* conventional STD NMR.

To prove the independence of the SDTD build-up curves from gelator concentration, we applied the SDTD method to a dispersion of neutral enzymatically produced cellulose (EpC).<sup>128</sup> As expected, STD NMR absolute signal increased with the increasing of EpC concentration (Figure 5.10, a), while the SDTD build-up curves showed a perfect overlap for the three EpC concentrations tested (Figure 5.10, b). This and the three reported almost identical spin diffusion coefficients  $D$  derived

from Equation 5.15 (Table 5.3) reveal no significant changes in the magnetisation transfer at the solid-liquid interface. The data prove that (i) solvent structure is not influenced by EpC concentration, and (ii) the differences observed in the STD build-up curve are strictly due to changes in the fraction of bound solvent.



**Figure 5.10:** a) STD and b) SDTD build-up curves for HDO binding to EpC 0.5 wt% (light blue), 1 wt% (green) and 2 wt% (red) gels; A  $b$  value of 1 was used to obtain the best fit for the EpC SDTD curves.

**Table 5.3:** Calculated values for the  $C$  and  $D$  parameters obtained from the fit to Equation 5.15 of the SDTD build-up curves of the HDO peak for EpC dispersions at different concentrations. An  $r$  value of 0.2 nm was kept constant during the fit. A  $b$  value of 1 was used to fit the EpC SDTD curves. The errors associated to each  $C$  and  $D$  value are shown and correspond to the 99% confidence level. The  $R^2$  values of each fit to Equation 5.15 are shown.

<b>Gelator conc.</b>	<b>0.5 wt%</b>	<b>1 wt%</b>	<b>2 wt%</b>
<b><math>C</math></b>	1.24	1.19	1.24
<b>error</b>	$\pm 0.06$	$\pm 0.04$	$\pm 0.04$
<b><math>D</math> (nm<sup>2</sup>/ms)</b>	$8.85 \times 10^{-5}$	$9.25 \times 10^{-5}$	$8.88 \times 10^{-5}$
<b>error</b>	$\pm 3.06 \times 10^{-6}$	$\pm 2.97 \times 10^{-6}$	$\pm 5.40 \times 10^{-6}$
<b><math>R^2</math></b>	0.9942	0.9953	0.9949

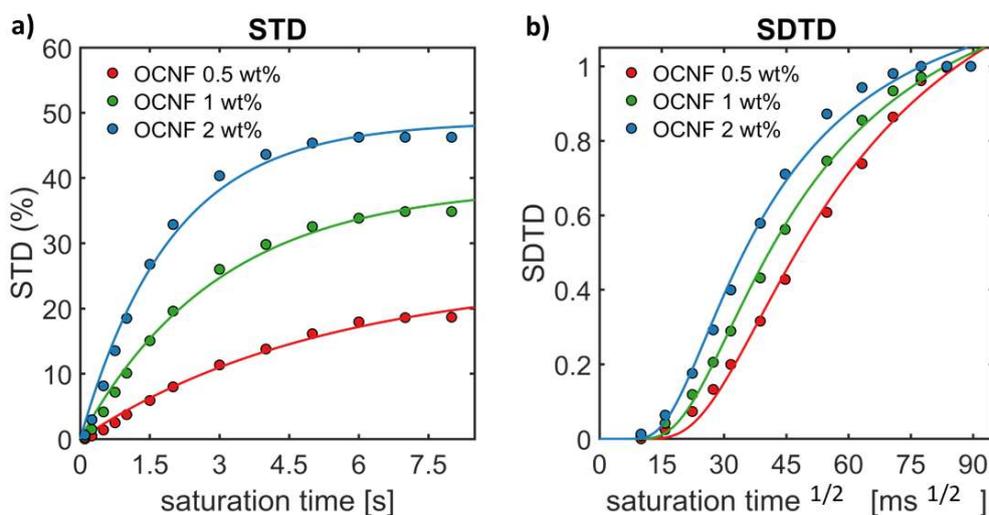
## Chapter 5 – Spin Diffusion Transfer Difference (SDTD) NMR

A different scenario is expected for systems where gelation occurs due to a change in solvent structuration, in which case changes of the network-solvent interactions would correspond to a variation in the magnetization transfer at the network-solvent interface. We predicted this to be reported by changes in our SDTD curves, as well as in the spin diffusion coefficient  $D$ .

We used OCNF dispersions at increasing concentrations as a case study. As explained in section 5.2.1.1, the gelation mechanism OCNF has been attributed to the establishment of fibril-fibril interactions with (i) no interactions for OCNF concentrations below 1 wt%, (ii) between 1 wt% and 1.5 wt% repulsive interactions start emerging, and (iii) above 1.5 wt% repulsive interactions between the OCNF fibrils are evident.<sup>227</sup>

Acquisition of the STD build-up curves and application of the SDTD methodology to three different OCNF concentrations (0.5 wt%, 1 wt% and 2 wt%) show an increase of both STD and SDTD factors (Figure 5.11, a and b), and the faster SDTD build-up implies that water becomes more structured upon increasing the concentration of gelator (Figure 5.11, b). The mathematical fit the Equation 5.15 yielded C and D parameters are reported in Table 5.4.

The increased degree of structuration of water at higher OCNF concentration might be due to (i) the formation of denser networks of structured water that shields the increased concentration of repulsive interactions, (ii) the increased presence of  $\text{Na}^+$  ions onto the surface of OCNF fibrils leading to reduced fibril-fibril repulsion and, therefore, increased fibril-fibril overlap and water confinement, and (iii) the enhanced structuration of water around the  $\text{Na}^+$  ions bound to the fibrils.<sup>227</sup> However, as  $\text{Na}^+$  is only present at stoichiometric ratios in our samples (very small compared to the HDO concentration), we expect a negligible contribution of the latter and a major role of the denser networks of structured water within the more confined space between fibrils.



**Figure 5.11:** a) STD and b) SDTD build-up curves for HDO binding to OCNF 0.5 wt% (light blue), 1 wt% (green) and 2 wt% (red) gels. A  $b$  value of 2 was used to obtain the best fit for the OCNF SDTD curves.

**Table 5.4:** Calculated values for the  $C$  and  $D$  parameters obtained from the fit to Equation 5.15 of the SDTD build-up curves of the HDO peak for OCNF dispersions at different concentrations. An  $r$  value of 0.2 nm was kept constant during the fit. A  $b$  value 2 was used to fit the OCNF SDTD curves. The errors associated to each  $C$  and  $D$  value are shown and correspond to the 99% confidence level. The  $R^2$  values of each fit to Equation 5.15 are shown.

<i>Gelator conc.</i>	0.5 wt%	1 wt%	2 wt%
<b><math>C</math></b>	0.76	0.69	0.65
<b>error</b>	$\pm 0.06$	$\pm 0.04$	$\pm 0.04$
<b><math>D</math> (<math>\text{nm}^2/\text{ms}</math>)</b>	$3.93 \times 10^{-5}$	$4.86 \times 10^{-5}$	$5.91 \times 10^{-5}$
<b>error</b>	$\pm 3.06 \times 10^{-6}$	$\pm 2.97 \times 10^{-6}$	$\pm 5.40 \times 10^{-6}$
<b><math>R^2</math></b>	0.9964	0.9975	0.9940

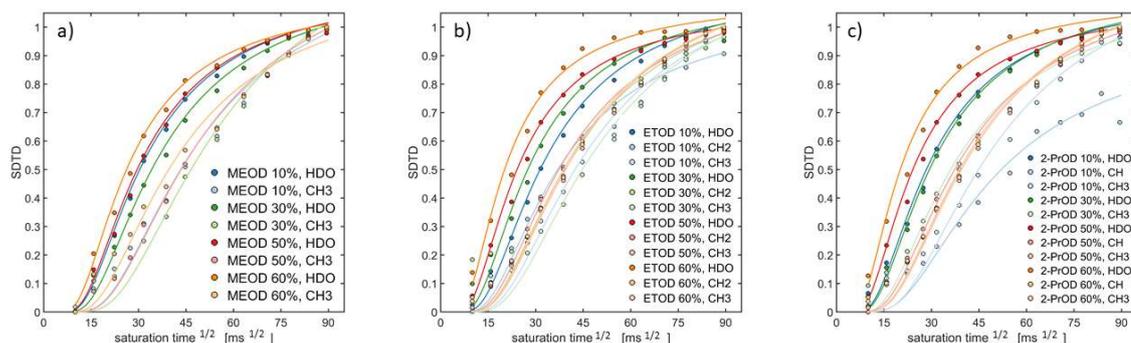
Hence, the independence of the SDTD factors from solvent and gelator concentrations allow to monitor changes in the degree of water structuration within gel networks of different gelator concentrations.

### 5.2.4 SDTD NMR characterisation of the role of co-solvents on the alcohol-induced gelation of OCNF hydrogels

The alcohol-induced gelation of OCNF was recently investigated by rheology and SAXS.<sup>254</sup> Three alcohols (i.e. methanol, ethanol and 2-propanol - in order of decreasing hydrophilicity) were tested in their ability to induce OCNF gelation in mixtures with water. Alcohol's hydrophilicity/hydrophobicity on its own did not explain the observed macroscopic properties. For example, methanol was able to induce gelation at the lowest concentration, followed by 2-propanol and ethanol. However, methanol and ethanol gels gave the weakest and stronger gels at the point of gelation, respectively. 2-propanol showed reduced stiffness compared to ethanol. With the exception of methanol gels, the analysis of SAXS data showed an increase in the cross-section of the OCNF nanofibrils above the gelation and phase separation concentrations.<sup>254</sup> The authors proposed that, as  $\text{Na}^+$  is less soluble in ethanol and 2-propanol than in water (similar solubility in methanol and water), the gel formation could be driven to some extent by the aggregation of OCNF fibrils due to the increased condensation of  $\text{Na}^+$  ions onto the surface of OCNF fibrils at higher ethanol and 2-propanol concentrations.<sup>254</sup> However, they also pointed out that the difference in the solubility of NaCl in ethanol and methanol is not large enough to explain the substantial differences of stiffness and fibril-fibril overlap between these alcohols (methanol gels are much weaker than ethanol gels), and suggested that other mechanisms must be involved. Thus, we hypothesised that water structuration must play an important role in the gelation mechanism of OCNF-alcohol mixtures.

To test our hypothesis, we studied a series of OCNF gels prepared in cosolvent mixtures of water ( $\text{D}_2\text{O}$ ) and low molecular weight alcohols. The  $\text{D}_2\text{O}$ -exchanged alcohols methanol (MeOD), ethanol (EtOD) and 2-propanol (2-PrOD) were studied at several concentrations (10 wt%, 30 wt%, 50 wt% and 60 wt%). An OCNF concentration of 1 wt% was used for all gels, and a dispersion of OCNF 1 wt% prepared in  $\text{D}_2\text{O}$  was employed as control sample.

Notably, the SDTD curves clearly demonstrate the preferential binding of HDO to OCNF at all alcohol concentrations (much faster growth of the SDTD build-up curves of HDO compared to the alcohols; Figure 5.12). These results indicate that water constitutes the first solvation shell(s) of OCNF nanofibrils, while the alcohol component would only establish indirect interactions mediated by water.

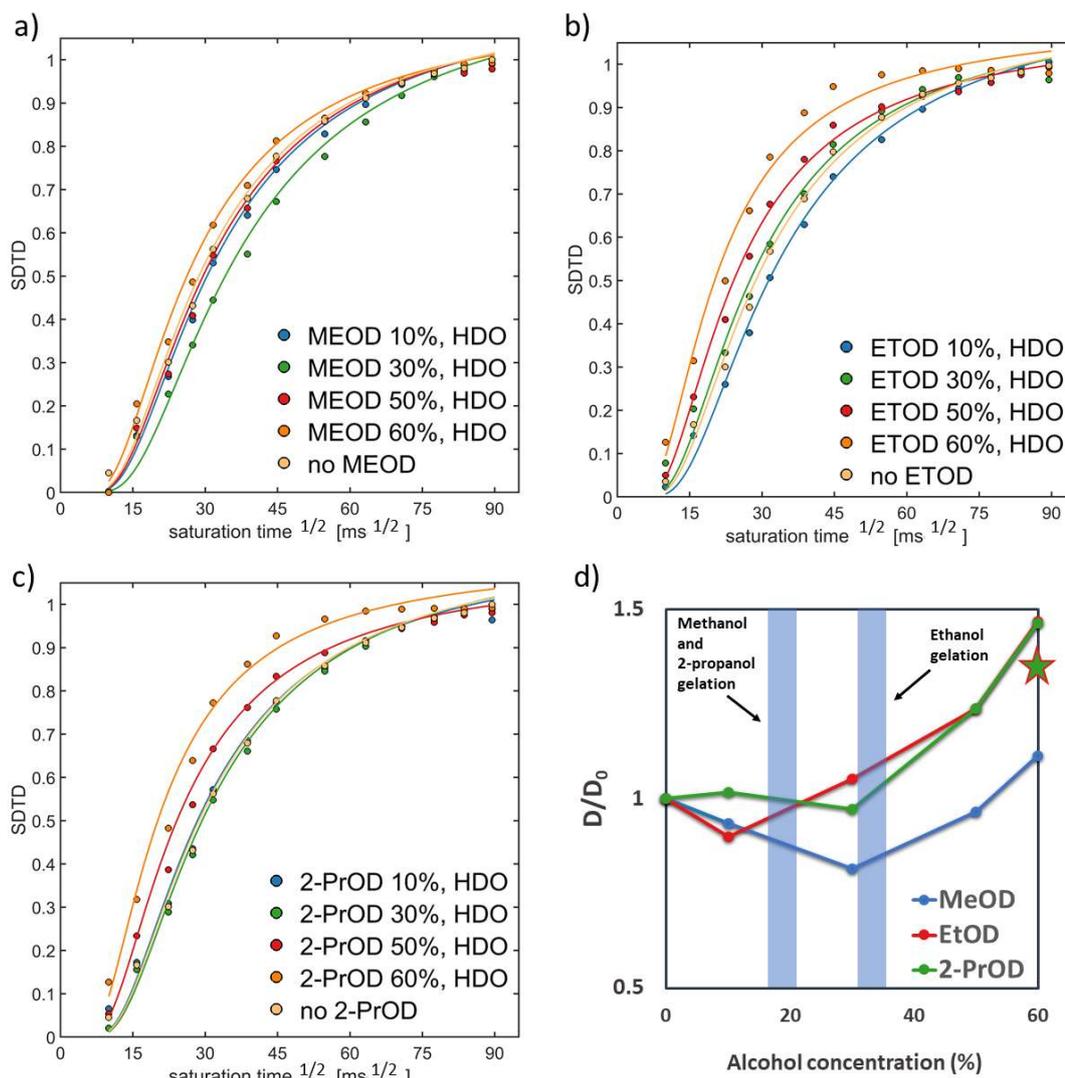


**Figure 5.12:** SDTD NMR build-up curves of the HDO alcohol peaks in OCNF 1 wt% gels prepared in a) D<sub>2</sub>O/MeOD, b) D<sub>2</sub>O/EtOD and c) D<sub>2</sub>O/2-PrOD cosolvent mixtures of 10 wt% (blue symbols), 30 wt% (green symbols), 50 wt% (red symbols) and 60 wt% (orange symbols) alcohol content. Note the faster growth of the SDTD build-up curves for HDO compared to the alcohols in all the gels.

Aiming to monitor the degree of structuration of HDO before, during and after gelation (i.e. upon addition of increasing concentrations of alcohol), we calculated the spin diffusion rate  $D$  of HDO for each water/alcohol gel (Figure 5.13, a, b, c and Table 5.5, 5.6 and 5.7), and normalised it against the spin diffusion rate  $D_0$  of HDO ( $D_0 = 7.71 \times 10^{-5} \text{ nm}^2/\text{ms}$ ) for the control sample without alcohol (OCNF 1 wt% in D<sub>2</sub>O). Thus, the  $D/D_0$  ratio of HDO was plotted as a function of alcohol content (Figure 5.13, d). Firstly, we note lower  $D/D_0$  values for the MeOD compared to the EtOD and 2-PrOD gels above the point of gelation. Also, MeOD gels showed a significant  $D/D_0$  decrease up to 30 wt% followed by an increase up to the  $D/D_0 \sim 1$  (i.e. very similar to the control sample) at 60 wt% of MeOD (no syneresis observed at this concentration). This suggests a lower capacity of MeOD to induce water structuration in OCNF gels compared to EtOD and 2-PrOD, which correlates with previous SAXS studies showing that the cross-section and overlap of OCNF fibrils do not vary significantly with methanol concentration.<sup>254</sup>

On the other hand, ethanol gels showed a continuous increase of  $D/D_0$  upon gelation (30 wt% of ethanol) and up to 60 wt% of alcohol content, while the  $D$  of ethanol was not significantly affected (Figure 5.13, d - point of syneresis indicated with a star). This suggests that the degree of structuration of water within the gel network increases with ethanol concentration, which correlates to the ethanol-induced increase of the average OCNF fibril cross-section (i.e. increase of fibril-fibril overlap), whereas the structuration of ethanol is barely affected. This further confirms that ethanol does not interact directly with the OCNF network, but possibly forms microdomains similar to what was described before for the mechanism of alcohol-induced gelation of clays.<sup>255</sup> Regarding 2-propanol gels, a behaviour very similar to ethanol gels was observed above 30 wt% of 2-propanol, although no significant differences of  $D/D_0$  were observed for concentrations below 30 wt%.

In conclusion, our SDTD NMR approach highlights the essential role of water structuration on the gelation properties of OCNF gels prepared in water and low molecular weight alcohol mixtures. In particular, the higher stiffness of water/ethanol and water/2-propanol gels correlates with their best ability to form networks of highly structured water compared to water/methanol gels, most likely due to the increased water confinement within the denser OCNF particle network (increased OCNF particle cross-section).<sup>254</sup> On the other hand, ethanol and 2-propanol would organise into microdomains due to the more favourable water-water and alcohol-alcohol compared to water-ethanol interactions. Notably, a similar mechanism was proposed for the alcohol-induced gelation of clays, where the clay particles were also in the sodium-salt form.<sup>255</sup> Overall, the SDTD NMR method has clarified the partially incomplete picture on the molecular features governing the mechanism of gelation and macroscopic properties of OCNF water/alcohol gels, highlighting that, besides fibril-fibril overlap and NaCl solubility in the alcohols, the degree of water structuration also plays a critical role on directing gel properties.



**Figure 5.13:** SDTD NMR build-up curves of the HDO peak in OCNF 1 wt% gels prepared in a)  $D_2O$ /MeOD, b)  $D_2O$ /EtOD and c)  $D_2O$ /2-PrOD cosolvent mixtures. The SDTD curve for the control sample (OCNF 1 wt% in 100%  $D_2O$ , 0% alcohol), is shown in pale yellow. The SDTD curves for the  $D_2O$ /alcohol-OD gels are shown in blue (10 wt% of alcohol-OD), green (30 wt% of alcohol-OD), red (50 wt% of alcohol-OD), and orange (60 wt% of alcohol-OD). A  $b$  value of 2 was used for all curves. d) Plot showing the evolution of the normalised spin diffusion rate ( $D/D_0$ ) of HDO binding to OCNF 1 wt% containing different concentrations of MeOD, EtOD and 2-PrOD.  $D_0$  represents the value of  $D$  of HDO calculated in the absence of alcohol (OCNF 1 wt% in  $D_2O$ , control sample). The  $D$  value of HDO in each water-alcohol gel sample was obtained from the fit of the SDTD build-up

## Chapter 5 – Spin Diffusion Transfer Difference (SDTD) NMR

curves shown in a), b) and c). The ranges of alcohol concentrations leading to gelation are shown as blue areas. The concentration (*ca.* 60 wt%) at which phase separation occurs for ethanol and 2-propanol gels is indicated with a star.

**Table 5.5:** Calculated values for the *C* and *D* parameters obtained from the fit to Equation 5.15 of the SDTD build-up curves of the HDO peak for OCNF 1 wt% dispersions with different concentrations of MeOD alcohol. An *r* value of 0.2 nm was kept constant during the fit. A *b* value 2 was used to fit the OCNF SDTD curves. The errors associated to each *C* and *D* value are shown and correspond to the 99% confidence level. The *R*<sup>2</sup> values of each fit to Equation 5.15 are shown.

MeOD	10%		30%		50%		60%	
	CH <sub>3</sub>	HDO						
<b>C</b>	0.66	0.59	0.74	0.61	0.68	0.59	0.60	0.57
<b>error</b>	±0.04	±0.01	±0.06	±0.02	±0.04	±0.01	±0.05	±0.02
<b>D</b>	4.70	7.20	4.08	6.29	4.61	7.45	5.46	8.59
<b>(nm<sup>2</sup>/ms)</b>	x10 <sup>-5</sup>							
<b>error</b>	±2.78	±1.30	±3.21	±2.47	±2.62	±2.88	±1.10	±5.48
	x10 <sup>-6</sup>							
<b>R<sup>2</sup></b>	0.9786	0.9978	0.9644	0.9894	0.9813	0.9898	0.9352	0.947

## Chapter 5 – Spin Diffusion Transfer Difference (SDTD) NMR

**Table 5.6:** Calculated values for the *C* and *D* parameters obtained from the fit to Equation 5.15 of the SDTD build-up curves of the HDO peak for OCNF1 wt% dispersions with different concentrations of EtOD alcohol. An *r* value of 0.2 nm was kept constant during the fit. A *b* value 2 was used to fit the OCNF SDTD curves. The errors associated to each *C* and *D* value are shown and correspond to the 99% confidence level. The *R*<sup>2</sup> values of each fit to Equation 5.15 are shown.

EtOD	10%			30%			50%			60%		
	CH <sub>2</sub>	CH <sub>3</sub>	HDO									
<b>C</b>	0.55	0.64	0.60	0.62	0.67	0.58	0.61	0.65	0.55	0.64	0.65	0.56
<b>error</b>	± 0.05	± 0.04	± 0.01	± 0.05	± 0.04	± 0.01	± 0.02	± 0.02	± 0.01	± 0.02	± 0.02	± 0.01
<b>D (nm<sup>2</sup>/ ms)</b>	6.24 x10 <sup>-5</sup>	4.87 x10 <sup>-5</sup>	6.93 x10 <sup>-5</sup>	5.45 x10 <sup>-5</sup>	4.51 x10 <sup>-5</sup>	8.12 x10 <sup>-5</sup>	5.74 x10 <sup>-5</sup>	5.23 x10 <sup>-5</sup>	9.55 x10 <sup>-5</sup>	5.35 x10 <sup>-5</sup>	5.30 x10 <sup>-5</sup>	1.13 x10 <sup>-4</sup>
<b>error</b>	±7.14 x10 <sup>-6</sup>	±3.17 x10 <sup>-6</sup>	±1.17 x10 <sup>-6</sup>	±5.90 x10 <sup>-6</sup>	±2.93 x10 <sup>-6</sup>	±3.30 x10 <sup>-6</sup>	±2.41 x10 <sup>-6</sup>	±2.03 x10 <sup>-6</sup>	±2.46 x10 <sup>-6</sup>	±1.56 x10 <sup>-6</sup>	±1.82 x10 <sup>-6</sup>	±5.78 x10 <sup>-6</sup>
<b>R<sup>2</sup></b>	0.914	0.973	0.997	0.912	0.974	0.983	0.989	0.990	0.995	0.994	0.992	0.978

**Table 5.7:** Calculated values for the  $C$  and  $D$  parameters obtained from the fit to Equation 5.15 of the SDTD build-up curves of the HDO peak for OCNF 1 wt% dispersions with different concentrations of 2-PrOD alcohol. An  $r$  value of 0.2 nm was kept constant during the fit. A  $b$  value 2 was used to fit the OCNF SDTD curves. The errors associated to each  $C$  and  $D$  value are shown and correspond to the 99% confidence level. The  $R^2$  values of each fit to Equation 5.15 are shown.

2-PrOD	10%			30%		50%			60%		
	CH	(CH <sub>3</sub> ) <sub>2</sub>	HDO	(CH <sub>3</sub> ) <sub>2</sub>	HDO	CH	(CH <sub>3</sub> ) <sub>2</sub>	HDO	CH	(CH <sub>3</sub> ) <sub>2</sub>	HDO
<b>C</b>	0.51	0.66	0.58	0.58	0.58	0.61	0.65	0.55	0.64	0.65	0.56
<b>error</b>	± 0.08	± 0.04	± 0.01	± 0.02	± 0.01	± 0.01	± 0.02	± 0.01	± 0.02	± 0.02	± 0.01
<b>D</b> <b>(nm<sup>2</sup>/ms)</b>	4.70 x10 <sup>-5</sup>	4.51 x10 <sup>-5</sup>	7.84 x10 <sup>-5</sup>	6.14 x10 <sup>-5</sup>	7.50 x10 <sup>-5</sup>	5.74 x10 <sup>-5</sup>	5.23 x10 <sup>-5</sup>	9.55 x10 <sup>-5</sup>	5.35 x10 <sup>-5</sup>	5.30 x10 <sup>-5</sup>	1.13 x10 <sup>-5</sup>
<b>error</b>	±8.13 x10 <sup>-6</sup>	±2.95 x10 <sup>-6</sup>	±2.2 x10 <sup>-6</sup>	±2.48 x10 <sup>-6</sup>	±1.26 x10 <sup>-6</sup>	±2.26 x10 <sup>-6</sup>	±2.03 x10 <sup>-6</sup>	±2.45 x10 <sup>-6</sup>	±1.56 x10 <sup>-6</sup>	±1.82 x10 <sup>-6</sup>	±85.3 x10 <sup>-6</sup>
<b>R<sup>2</sup></b>	0.819	0.973	0.994	0.989	0.998	0.987	0.990	0.990	0.995	0.992	0.982

### 5.3 Discussion and future perspective

The chapter presents for the first time the application of STD NMR for the investigation of solvent and cosolvents structuration in hydrogels systems. We demonstrate that, under conditions of negligible translational diffusion of the receptor within the NMR time scale, i.e., for large particles, the normalised STD build-up curve is 1) independent on the gelator and solvent concentrations, and 2) dependent on the receptor-to-solvent spin diffusion rates ( $D$ ) and the minimum receptor-to-solvent distance ( $r$ ). Our new method, called Spin Diffusion Transfer Difference (SDTD), relies on the normalisation of the standard STD NMR build-up curve against the highest STD factor determined at long saturation times. The SDTD build-up curve can be simulated using the general 1D diffusion equation. The results report on the degree of solvent structuration (e.g. the extension of structured solvent networks or number of solvation shells).

## Chapter 5 – Spin Diffusion Transfer Difference (SDTD) NMR

The SDTD method presents an advantage over traditional water polarisation transfer (WPT) solid-state NMR experiments, of similar aim, as it relies on monitoring the well-resolved solvent peaks instead of the broad or often invisible particle peaks, in solution or HR-MAS NMR. Hence, SDTD method does not require the solid-state NMR setup. Also, the method is based on the quick acquisition of one-dimensional NMR spectra, does not require isotopic labelling, nor assignment of the network peaks (being a solvent-observed method). This represents a significant advantage over gelator-observed solid-state NMR methods such as WPT, which rely on the cross-polarization efficiency and the observation of low abundant nuclei. In addition, SDTD allows for the study of diluted dispersions, very limited by WPT solid-state NMR experiments, as well as highly viscous gels by HR-MAS SDTD NMR.

Importantly, SDTD NMR allows for the simultaneous and rapid characterisation of the degree of structuration of different solvents in cosolvent gels. In this regard, the application of the SDTD methodology to OCNF-water/alcohol gels enabled the understanding of the role these cosolvents (methanol, ethanol and 2-propanol) on the macroscopic properties of these materials. In particular, the SDTD build-up curves demonstrated that (i) water binds preferentially to OCNF over any of the three alcohols tested, and (ii) the degree of water structuration increases with alcohol concentration for the water/ethanol and water/2-propanol gels. This effect correlates with the much higher gel strength of water/ethanol and water/2-propanol gels compared to methanol gels.

To conclude, we have demonstrated that the application of STD NMR can be extended beyond its traditional boundaries for very high molecular weight receptors such as carbohydrate particles. We show that by simulating the SDTD build-up curves with the 1D diffusion model it is possible to derive structural parameters at the particle-solvent interface reporting on the degree of solvent structuration. Our novel method will provide the community of soft matter with a straightforward, fast and robust ligand-observed NMR technique for better understanding of the role of the solvent(s) in the gelation mechanism and the rheological and mechanical properties of a wide range of particulate soft matter materials.

### 5.4 Material and methods

#### 5.4.1 Sample preparation

##### 5.4.1.1 *Gels prepared in D<sub>2</sub>O*

Dispersions of TEMPO-oxidised cellulose nanofibrils (OCNF), corn starch (CS) and enzymatically produced cellulose (EpC) at different concentrations were prepared in D<sub>2</sub>O. OCNF of a degree of oxidation of *ca.* 25%, produced from purified softwood fibre and processed *via* high pressure homogenisation, was kindly provided by Croda. These were further purified by dialysis against ultra-pure water (DI water, 18.2 MO cm) and stirred at room temperature for 30 min. Then the dispersion was acidified to pH 3 using HCl solution and dialysed against ultra-pure water (cellulose dialysis tubing MWCO 12400) for 3 days with the DI water replaced twice daily. The dialysed OCNF suspension was processed via mechanical shear (ULTRA TURRAX, IKA T25 digital, 30 minutes at 6500 rpm) and the pH was adjusted to 7 using NaOH solution. This dispersion was further dialysed to remove any remaining salts and dispersed using a sonication probe (Ultrasonic Processor, FB-505, Fisher), via a series of 1 s *on* and 1 s *off* pulses for a total time of 60 min at 30% amplitude in an ice bath, and subsequently freeze-dried.

To prepare the OCNF dispersions for NMR investigation, OCNF powder and water were weighed to provide the desired concentrations of OCNF, and then probe sonicated for 30 min at 20% amplitude using pulses of 1 s *on* and 2 s *off*, using an ultrasonic processor vibracell VCX 130 sonicator. CS samples were first gelatinised in a boiling water bath for 30 minutes. The CS samples were sonicated for 2 min at 40% amplitude using 1 s *on* and 2 s *off* pulses.

For the H<sub>2</sub>O titration experiments, OCNF 1 wt% dispersions were prepared using MilliQ® water and D<sub>2</sub>O of 99.9 atom % D to achieve the desired H<sub>2</sub>O/D<sub>2</sub>O ratio (5:95, 10:90, 20:80 and 30:70). For the variable gelator concentration experiments (OCNF and EpC at 0.5, 1 and 2 wt%), the samples were prepared by dilution from the 2 wt% dispersions to avoid error propagation.

##### 5.4.1.2 *OCNF 1 wt% gels prepared in mixtures of D<sub>2</sub>O and alcohol-OD*

First, stock dispersions of OCNF 2 wt% were prepared by redispersing OCNF powder in D<sub>2</sub>O by probe sonication for 1 min at 30% amplitude using 1 s *on* 1 s *off* pulses, using an ULTRA TURRAX,

IKA T25 digital sonicator. Subsequently, all the gels were prepared by dilution of the OCNF 2wt% dispersions using the corresponding alcohol-OD and D<sub>2</sub>O weight concentrations.

D<sub>2</sub>O (151882) and 2-propanol-OD (615080) were purchased from Sigma-Aldrich. Ethanol-OD and methanol-OD were purchased from Cambridge Isotopes Lab, Inc.

### 5.4.2 Nuclear magnetic resonance (NMR) spectroscopy

Solution state NMR experiments were performed using a Bruker Avance I spectrometer equipped with a 5 mm triple resonance probe operating at frequency of 499.69 MHz (<sup>1</sup>H). Saturation transfer difference (STD) NMR experiments of CS and OCNF dispersions were acquired at 298 K using a train of 50 ms Gaussian shaped pulses for selective saturation of the gelator particles, using an *on-resonance* frequency of 0 and -1 ppm for CS and OCNF dispersions, respectively, and an *off-resonance* frequency of 50 ppm. For the CS 15 wt% dispersion, saturation times ranging from 50 ms to 5 s were used. For the experiments carried out on OCNF dispersions in water (*i.e.* H<sub>2</sub>O titrations and variable OCNF concentration), STD NMR experiments were performed using saturation times ranging from 100 ms to 8 s. A constant time length per scan (saturation time + recycle delay) of 8 s was used. Depending on saturation time, STD NMR experiments were performed with 128 scans or less (with a minimum of 16 scans), in inverse relation to the saturation time, and 8 dummy scans.

STD NMR experiments for different concentrations of EpC were carried out using an Avance II Bruker 800.23 MHz spectrometer equipped with a 5 mm inverse triple-resonance probe. The experiments were acquired at 298 K at saturation times ranging from 100 ms to 8 s, using a constant time length per scan (saturation time + recycle delay) of 8 s. The *on-* and *off-resonance* frequencies were set to -1 and 50 ppm, respectively. Depending on saturation time, STD NMR experiments were performed with 512 scans or less, in inverse relation to the saturation time, and 8 dummy scans.

The D<sub>2</sub>O/alcohol-OD OCNF gels were characterised by high-resolution magic angle spinning (HR-MAS) using a solid-state Bruker Avance III spectrometer operating at a <sup>1</sup>H frequency of 400.22 MHz with a triple resonance HR-MAS probe (<sup>1</sup>H,<sup>31</sup>P,<sup>13</sup>C). All samples were spun at 6 kHz. HR-MAS NMR was required for these samples due to large <sup>1</sup>H peak broadening precluding enough

resolution in the absence of magic angle spinning. The large spectral broadening of the D<sub>2</sub>O/alcohol-OD OCNF gels is due to their high viscosity leading to very strong dipolar couplings, particularly for D<sub>2</sub>O/ethanol-OD and D<sub>2</sub>O/2-propanol-OD at high alcohol concentrations.

Saturation transfer difference (STD) NMR experiments were carried out by <sup>1</sup>H selective irradiation (*on*-resonance) of OCNF peaks (2.3-2.5 ppm). A train of 50 ms Gaussian-shaped pulses were employed for saturation, with a field strength of 50 Hz.<sup>2</sup> STD NMR experiments using 0.5, 0.75, 1, 1.5, 2, 3, 4, 5, 6, 7 and 8 s saturation times were carried out, using a total relaxation time of 8.1 s. The *off*-resonance frequency was set to 56 ppm.

The STD spectra ( $I_{STD}$ ) were obtained by subtracting the *on*- ( $I_{sat}$ ) to the *off*-resonance ( $I_0$ ) spectra. To determine the STD response or STD factor ( $\eta_{STD}$ ), the peak intensities in the difference spectrum ( $I_{STD}$ ) were integrated relative to the peak intensities in the *off*-resonance spectrum ( $I_0$ ). The SDTD build-up curves were obtained by normalising all the STD factors against the highest value (usually corresponding to the longest saturation time).

### 5.4.3 Simulation of the SDTD build-up curves

To obtain a good fit of the SDTD build-up curve, it is essential to achieve a good sampling of both the lag phase and the plateau of the curve. To do this, saturation times ranging from tens of milliseconds to 6-8 seconds need to be used. The SDTD build-up curves were represented as a function of the square root of the saturation time and simulated in Matlab (Appendix, Script A.1) using Equation 5.15. Here, the dependent variable is the normalized intensity of the NMR observable and the independent variable is the square root of the saturation time (in ms),  $r$  is the minimum distance of the grid (in nm),  $D$  is the spin diffusion rate (in nm<sup>2</sup>/ms) at the particle-solvent interface,  $erfc$  is the complementary error function,  $C$  is the proportionally constant of the fit, and  $b$  is a parameter to centre the function around  $x$ . Notably, the growth rate of the SDTD curve presents a proportional and inversely proportional relationship to the spin diffusion rate  $D$  and the minimum distance  $r$ , respectively, both related to the degree of solvent structuration within the gel network. Hence, faster spin diffusion rates  $D$  and shorter distances  $r$  reflect increased solvent structuration.

## Chapter 5 – Spin Diffusion Transfer Difference (SDTD) NMR

By visual investigation, a tendency of the fitting curves to misfit the final points could be noticed. Nonetheless, the fitting was considered valid in line with the acceptable value of the standard deviations obtained. In addition, comparison of curves fitting done with origin resulted in the same values of the spin diffusion coefficient  $D$ , indicating the veracity of the reported data.

## Chapter 6

### Final remarks

## 6. Final Remarks

Carbohydrates are abundant biomolecules with a strong tendency to form supramolecular networks via the formation of a multitude of hydrogen bonds. Due to their inherent biocompatibility, it is becoming increasingly important to develop strategies to functionalise these materials with specific and tuneable properties. Hence, controlling the composition of carbohydrate-based polymers at the molecular level could provide materials with novel self-assembly arrangements and tailored properties.

Abundant plant-based polysaccharides comprise starch, cellulose, hemicelluloses and chitin, and from these building blocks, value-added products can be obtained by chemical, thermal or enzymatic treatments. These treatments expand the design space for creating completely novel polysaccharide-based compounds. Carbohydrates active enzyme (CAZy) can be used *in vitro* to produce polysaccharide-based products. In addition, the growing portfolio of techniques for genetic engineering has the potential to be expand the enzyme specificity towards non-cognate substrates for the production of functionalised polysaccharides.

To achieve desired properties by design, it is important to understand both the mechanistic details behind the molecular recognition of building blocks in the enzyme catalytic cleft, and the mechanism of the polymer self-assembly. This will allow us to control arrangement, morphology and properties of the polysaccharide chains. The aim of this chapter is to emphasize our hypothesis and main findings, as well as to highlight the novelty of this doctoral work.

The core results of this thesis are:

- 1) the investigation of the binding mode of series of cognate and non-cognate substrates to Cellodextrin phosphorylase (CDP), deciphering important contacts for substrates specificity;
- 2) the application of CDP for the synthesis of novel cellulosic materials, together with the atomic level characterisation of the particles self-assembly;
- 3) the implementation of a novel STD NMR approach to study solvent structuring in hydrogel systems.

Due to the potential wide-ranging applications of the presented work, we think the content of this thesis will have considerable appeal, not just to the cellulose community, but more broadly in materials science, pharmaceuticals and biosciences.

Throughout the thesis, we covered the fields of glycobiology, intermolecular interactions, materials development and characterisation. The field of Nuclear Resonance was also explored, with an eye on its extensive application in structural studies. The great variability of the presented studies could be achieved only thanks to the coexistence of these fields of expertise in our research group.

At this point, it is important to emphasise the vast potential of NMR spectroscopy to study both relevant weak and dynamics interactions (either protein-ligand or gel network-solvent molecules), as well as to obtain a fingerprint of the molecular arrangement of novel materials. Carbohydrates are an interesting subject for NMR observation given their high degree of stereochemistry, flexibility and complexity. Still, carbohydrates generally present a low chemical shift dispersion (proton chemical shifts concentrate between 3.2 ppm and 4.1 ppm for the ring protons H2 to H6, and between 4.4 ppm and 5.2 ppm for the anomeric protons). Carbon chemical shifts can be found around 60 ppm for C6, between 67 ppm and 80 ppm for the ring carbon C2 to C5 and between 90 ppm and 105 ppm for the anomeric carbons).<sup>256</sup> Consequently, it is often necessary to use high field spectrometers and, in the case of solid-state NMR, to achieve ultra-fast spinning of the sample, to obtain atomic details. In addition, carbohydrates usually present weak-to-medium binding affinities to proteins, which makes NMR the ideal tool to study their interactions.

The accurate knowledge of the structure of polymeric material is an essential step for the complete understanding of the solids and for monitoring their properties. Diffraction techniques such as single-crystal and powder X-ray diffraction, small-angle X-ray scattering and small angle neutron scattering analysis are commonly used to characterise molecular and supramolecular interactions in the solid state. For single-crystal XRD a large, stable and good quality crystal is required, and when not available, the structural analysis is conducted on polycrystalline powders, named powder XRD. This determines the contemporary presence of small crystals, each of them with their own orientation, and consequently poorer data are acquired. To overcome these requirements of strict

periodicity and crystal size, solid-state NMR spectroscopy is usually applied. The technique allows not only the characterisation of ordered materials, but also the study of materials that are not necessarily periodic and may present local disorder or mobility, such as amorphous powders and gels.<sup>257, 258</sup> NMR spectroscopy is sensitive to the local environment of atoms, therefore the same atom in different molecules of an asymmetric unit cell can show different chemical shift values (as example, C1 in cellulose type II gives a doublet).<sup>191</sup> Therefore, correlations can be made between chemical shift and tensor components with the number and multiplicity of crystallographically non-equivalent positions in the asymmetric unit.<sup>257</sup> In addition, NMR provides information on through-bond connectivities, through-space proximities and intermolecular distances.<sup>87, 257</sup>

The well-equipped NMR facility available at the School of Pharmacy at UEA combines a 800 MHz Bruker spectrometer with both solution state and HR-MAS probeheads, a solution state 500 MHz Bruker spectrometer, as well as two solid-state NMR spectrometers: a 400 MHz Bruker spectrometer featured with four different probeheads, three of them able to achieve different spinning rates (4 mm, 2.5 mm and a 1.3 mm rotor sizes, for sample spinning of 15 KHz, 35 KHz and 67 KHz), the fourth being a HR-MAS probehead, and a 300 MHz Bruker spectrometer. It was also possible to resort to the UK solid-state NMR national facility in the University of Warwick for the use of the 850 MHz solid-state Bruker spectrometer for material structural elucidation. In addition, the performance of molecular docking calculations to model the 3D structure of the interactions in solution, on the bases of the structural experimental data, could be achieved thanks to availability of the Maestro Schrödinger suite (reported to be the most efficient tool to model flexible carbohydrates in relatively shallow binding pockets).<sup>259</sup> It is therefore evident that the combination of expertise from our research group and the available facility created the perfect environment to make considerable research advances.

### 6.1 Cellodextrin phosphorylase from *Clostridium thermocellum*: structural studies on substrates recognition

In Chapter 3, we investigated the specificity and the mechanism of molecular recognition of donor and acceptor substrates by CDP, with an emphasis on the ability to discriminate between cognate

and non-cognate donor-like molecules and between substrate lengths and stereochemistry of acceptors molecules.

Thanks to the combination of NMR approaches and protein-ligand docking calculations, we obtained structural elucidation of the binding of cognate and non-cognate substrates molecules to the -1, +1 and +2 subsites of CDP. The natural leads were Glc-1-P, D-cellobiose and D-celotriose, while the non-cognate substrates were the glucose isomers Gal-1-P and Man-1-P, the functionalised glucose analogues GIN-1P, 6F-Glc-1-P and the cellobiose regioisomer D-laminaribiose. Our aim was to investigate the binding event in terms of affinity, specificity and mechanisms of recognition. Through this investigation, we wanted to unveil:

- the main elements of the ligands for the molecular recognition
- the key elements of the interactions
- the minimum sugar entity recognised by CDP
- the selectivity of CDP towards glycosidic linkages

### 6.1.1 Molecular recognition of donor and donor-like substrates:

Our main results indicate the high specificity for the hexopyranose ring of donor and donor-like substrates to the -1 subsite of CDP. The main pieces of structural information that we obtained investigating Glc-1-P and the non-cognate donor-like substrates were:

#### **For Glc-1-P**

- The proof that the glucopyranose ring fits tightly in the -1 subsite
- The hint that the specificity of binding Glc-1-P is driven by the closer contacts showed for H1, H3, H5 and H6s (above 90% of relative STD) with the enzyme surface in comparison with H2 and H4
- the location of Glc-1-P hydroxymethyl group between Trp622 and Asp624 residues in CDP -1 subsite, together with the close contact with Trp622 side chain and the establishment of an H-bond between the hydroxyl group in C6 and the backbone NH of Asp624

**For Gal-1-P and Man-1-P**

- Gal-1-P rearrangement of the hexopyranose ring in binding CDP in comparison to Glc-1-P
- The evidence that Man-1-P tightly binds CDP and presents the same binding mode of Glc-1-P, proving that the C2 configurational change does not impair binding
- The presence of a steric clash between the axial hydroxyl in C2 of Man-1-P and the catalytically competent Asp624

**For GIN-1-P and 6F-Glc-1-P**

- A change in the binding mode of both GIN-1-P and 6F-Glc-1-P, in comparison with Glc-1-P, characterised by a reduction in H1 and H2 contacts
- The possibility of developing steric hindrance and electrostatic repulsion between the amine group in C2 of GIN-1-P and the positively charged side chain of Arg496
- The tolerance of CDP for a group at position 6 isosteric to OH acting as an H-bond acceptor

### 6.1.2 Molecular recognition of acceptor and acceptor-like substrates:

Our main results indicate that specificity of CDP -1 subsite towards the glucopyranose ring of acceptor molecules is not sufficient to anchor the ligand in the binding pocket, and additional cooperative interactions need to be established to ensure the binding of acceptor substrates. This was proven by the lack of binding for D-glucose indicated by no detection of STD signal.

The main pieces of structural information that we obtained investigating the longer substrates D-cellobiose, D-celotriose and the non-cognate acceptor-like D-laminaribiose were:

**General information**

- The proof that the ligand exclusively enters the binding pocket with the nonreducing ring
- The importance of the contact established between H2 of the non-reducing ring and the enzyme surface for acceptor recognition
- The thermodynamics contribution played by inorganic phosphate in acceptor substrates recognition

**For D-cellobiose**

- The anomeric selectivity towards the  $\beta$ -configuration of the reducing ring, driven by the establishment of an H-bond between the reducing ring hydroxyl group in C1 and Asp297 residue in the +1 subsite of CDP binding pocket

**For D-cellobiose**

- The anomeric selectivity towards the  $\beta$ -configuration of the reducing ring, explained by the C-H/ $\pi$ -stacking established between the D-cellobiose reducing ring and Tyr300 residue in the +2 subsite of CDP binding pocket

**For D-laminaribiose**

- The preferential molecular recognition of the  $\alpha$ -configuration of the reducing ring (inverted anomeric selectivity in comparison with D-cellobiose and D-cellobiose)

### 6.1.3 Additional notes

With molecular docking we developed the 3D molecular models of Glc-1-P and D-cellobiose bound to CDP, which was validated against experimental STD data. Our 3D model of the CDP/Glc-1-P is the first Michaelis complex of the CDP enzyme acting on a donor substrate available to date. Unfortunately, at the current stage we were not able to experimentally validate the 3D model of the ternary complex CDP/Glc-1-P/ D-cellobiose. The investigation was hindered by the stronger affinity demonstrated by D-cellobiose, in comparison with the donor-like substrate (Man-1-P) used to conduct our studies. It would be interesting, for future work, to investigate different substrates and experimental conditions (i.e. temperature and pH) in order to provide experimental data to validate (or not) our model.

This investigation was conducted with the aim to uncover the structural details of the molecular recognition of donor and acceptor substrates by CDP, paramount in directing novel pathways for the synthesis of functionalised cellulosic materials. It is in our interest to extend the applicability of CDP, by enhancing the binding affinity of accepted substrates, or by allowing recognition of non-accepted substrates.

## 6.2 Deoxyfluorinated cellodextrin derivatives: structural characterisation

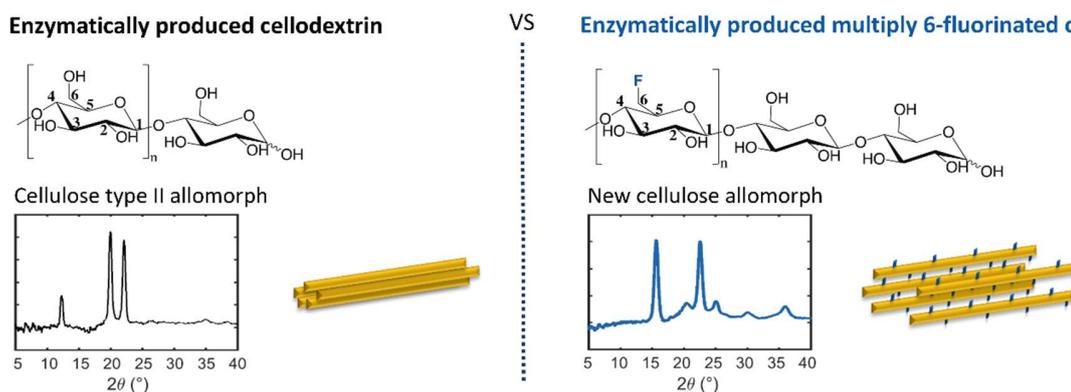
In Chapter 4, we introduced a bottom-up (from single building blocks to larger molecules), chemoenzymatic synthesis approach to access site-selectively functionalized cellulose oligomers, along with their detailed structural analysis. Our approach expands the scope of chemoenzymatic syntheses in the production of tailor-made materials and evidences the potential to generate unique structural arrangements.

The synthesis of mono- and multi-fluorinated cellulose-like materials by CDP was achieved using either non-cognate acceptors or donor-like molecules. The detailed structural characterisation of a novel cellodextrin derivatives at different length scales was performed using complementary analytical techniques (TEM, IR-AFT, AFM, PXRD and NMR). In our project, more emphasis was given to NMR spectroscopy, due to its proven ability to detect a wide range of domains (surface/core, organised/amorphous, rigid/mobile). The heterogeneous character of these materials requires the combination of solution and solid-state NMR techniques.

The core interests of this thesis are:

- 1) the combination of PXRD and solid-state NMR to characterise the molecular scale arrangement of the mono- and multi-fluorinated cellodextrin chains, and
- 2) the combination of a wide range of solution- and solid-state NMR experiments to decipher water exposed and interior chemical environments for different carbon sites.

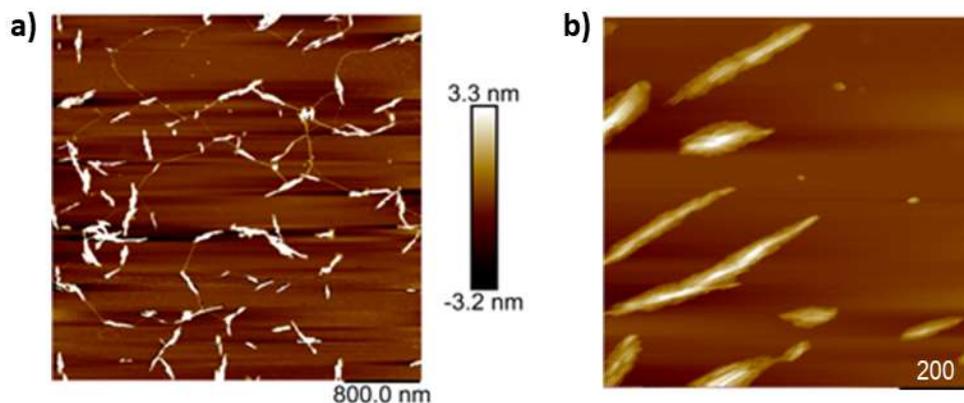
We believe that these studies represent a significant advancement for the field, proving that glycoside phosphorylases can tolerate functionalized donor substrates, generating unique structural arrangement. Indeed, the presence of multiple units of 6F-glucose yielded novel fluorinated cellulose oligomers with unique structural features. In one instance, the crystalline arrangement discovered has not been reported before and constitutes a new long-range ordering motif unprecedented for cellulose-based biomaterials (Figure 6.1).



**Figure 6.1:** The different PXRD pattern for enzymatically produced cellulodextrin (left) and the enzymatically produced multiply 6-fluorinated cellulodextrin (right), demonstrating a distinct organisation of the cellulodextrin fibrils.

The site selective introduction of  $^{19}\text{F}$  nuclei was adopted as these nuclei represent a highly sensitive NMR probe for chemical and biological interactions in complex systems. As previously mentioned in Chapter 4,  $^{19}\text{F}$  is of increasing interest due to its 100% natural abundance, its almost total absence in biological systems and its pronounced chemical shift dispersion and sensitivity to local environment in NMR spectroscopy. Therefore, the introduction of  $^{19}\text{F}$  may give us useful information about conformational changes and interaction processes in complex systems such as interpenetrating or composite hydrogels.

In the context of our project aiming to develop cellulose based materials with tailored properties, the fluorine inserted onto enzymatically produced cellulodextrin oligomers can be used to probe the associations of the oligomers when packing into crystalline and/or amorphous regions, their interaction with other polysaccharides, and their process of mesoscale assembly. For instance, we prepared blends between the monofluorinated 3F-EpC and the multi 6F-EpC with crystalline nanocellulose (CNC). Our preliminary studies using AFM indicate that, in the case of 3F-EpC, CNC forms a network of particles which coats and bridges the 3F-EpC fibrils (Figure 6.2, a). On the contrary, multi 6F-EpC assembles around CNCs (Figure 6.2, b). However, complete NMR data needed for full characterisation of the structure of this novel material are yet to be acquired. Importantly, due to the very similar  $^{13}\text{C}$  chemical shifts of CNC and fluorinated cellulose, high resolution and ultra-fast MAS spinning rate are of fundamental importance to attempt to differentiate both components in the  $^1\text{H}$ - $^{13}\text{C}$  CP/MAS NMR spectra.



**Figure 6.2:** AFM height images of a) 3F-EpC and b) multi 6F-EpC cellulose fibrils blended with CNC. 3F-EpC fibrils (light brown narrow areas, left image) coat and bridge the CNC particles (white shapes, left image), while multi 6F-EpC (light brown shapes, right image) assembles around CNCs (white shapes, right image).

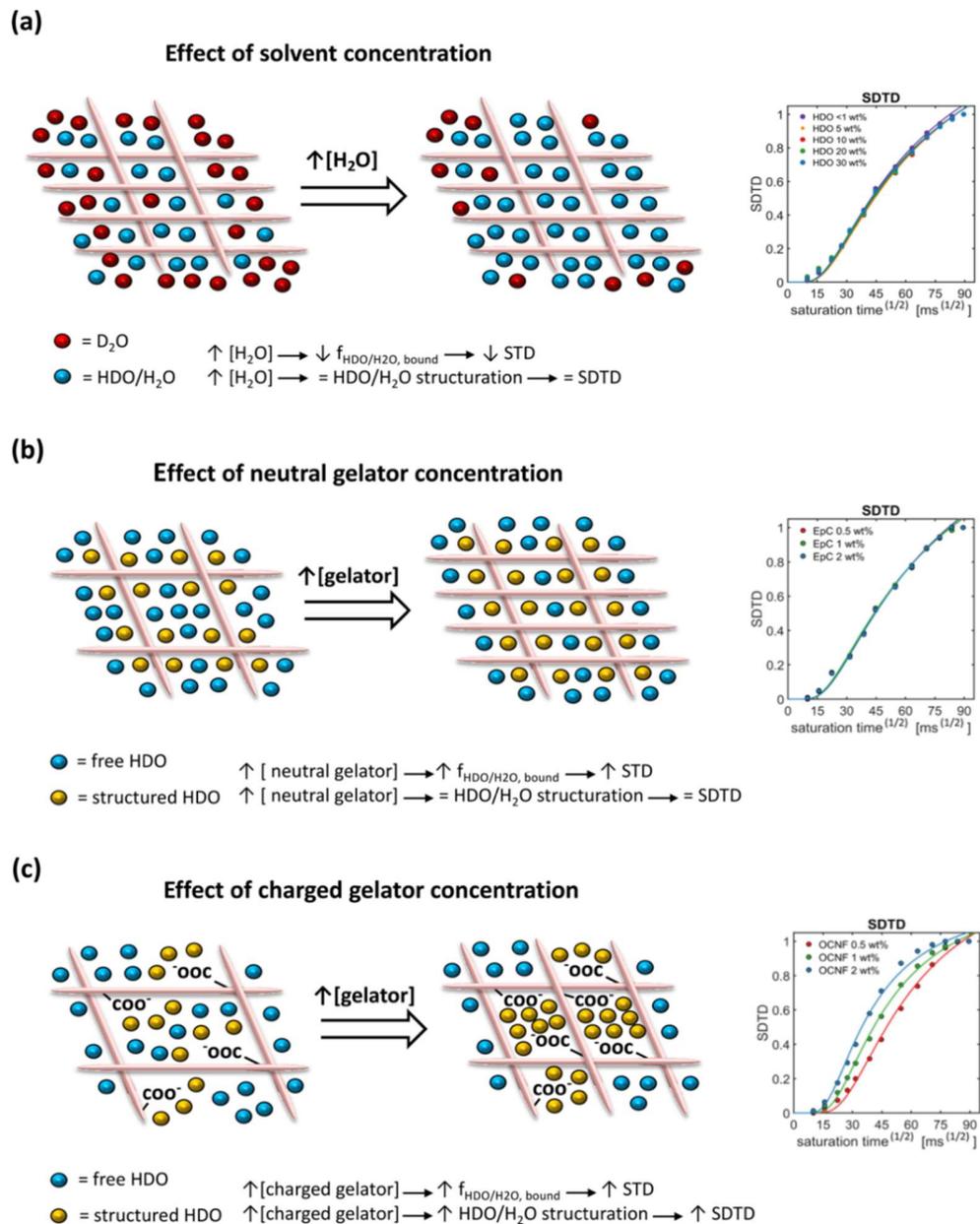
### 6.3 SDTD: a novel NMR tool to study solvent (and co-solvents) in 3D gels network

In Chapter 5, we introduced a novel solution state NMR methodology based on STD NMR to investigate solvent structuration in gels systems. Gels are neither conventional solids nor liquids, and their heterogeneous nature (the coexistence of a very rigid component together with a dynamics and highly mobile solution phase) makes them very challenging systems to characterise. Several complementary analytical techniques are required to gain a complete understanding of morphology, dynamics and supramolecular organisation of these systems.

In this work, we focused on gaining understanding of the role played by solvent in the gelation. Solvent properties play a central role in mediating aggregation and self-assembly of gelators and have huge influence on the solvation effect (depending on the system pH, which affects the competing solvation of donor and acceptor sites of hydrogen bonds) and salt effects (depending on the types of salt ion and their ionic strength, impacting the strength of electrostatic interactions).<sup>260</sup>

Our novel methodology, named Spin Diffusion Transfer Difference (SDTD) NMR, cancels the effect of solvent and gelator concentrations on the signal recorded by STD NMR experiments (Figure 6.3,

a and b), and allows the comparison between different gels, thus unveiling changes in solvent structuration (Figure 6.3, c). In addition, the application of our methodology gave access to important parameters for the characterisation of solvent structuring around the solid network, such as solvent-network minimum distance ( $r$ ) and diffusion rate at the solid-liquid interface ( $D$ ).



**Figure 6.3:** Summary of the main findings reported in Chapter 4. The effect of solvent (a) and gelator concentration, for neutral (b, EpC) and charged (c, OCNF) gelators, on the SDTD build-up curves and the degree of structuration of water within the gel network is shown.

This study was conducted on carbohydrate-based hydrogels (namely, corn starch, enzymatically produced cellodextrin and oxidised cellulose nanofibrils), as well as by using hydrogels in which the gelation process is induced by the presence of alcohols. Therefore, our methodology was proven to be a robust approach to monitor both main solvent and co-solvents present in the systems, unveiling preferential interactions between the gelator network and the surrounding solvents.

We expect the use of SDTD NMR to be expanded to other systems, including non-carbohydrate-based hydrogels, organogels, and multi-components systems (i.e. copolymeric or double network gels, as an example). Solvation plays a very important role in the stimuli-responsive properties of these systems.<sup>261, 262</sup> In addition, the methodology can be expanded to monitor the interactions of small ligand within system, such as in the case of drug-loaded hydrogels,<sup>238</sup> or even to monitor hydrogel-protein interactions.

The chance to apply SDTD NMR to a diverse range of systems (and with promising results so far) is very important to strengthen the potential of the technique. Further perspectives include the additional methodological investigation of the effect of different parameters on the SDTD outcome (including temperature, irradiation power, etc.).

We emphasise that the monitoring of the broad and high intensity peak of HDO, as well as the broad nature of the solid resonance hidden in the baseline, does not require the application of high magnetic field, nor the acquisition of an prohibitive number of scans. - The sample is thus easy to prepare, and the experiments are fast, reliable and easy to set up. Nonetheless, in the case if the solvent peak(s) are difficult to be resolved (i.e. in stiff gel, the solvent peak might be too broad) it will be necessary to use HR-MAS.

## References

1. McNaught, A. D., International Union of Pure and Applied Chemistry and International Union of Biochemistry and Molecular Biology - Joint Commission on Biochemical Nomenclature - Nomenclature of carbohydrates - Recommendations 1996. *Pure Appl Chem* **1996**, *68* (10), 1919-2008.
2. Dowd, M. K.; French, A. D.; Reilly, P. J., Modeling of aldopyranosyl ring puckering with MM3 (92). **1994**, *264* (1), 1-19.
3. Jeffrey, G. A.; Ja Park, Y., Application of molecular mechanics to the structure of 1,6-anhydroxyranoses. *Carbohydrate Research* **1979**, *74* (1), 1-5.
4. Pople, D. C. a. J. A., A General Definition of Ring Puckering Coordinates. *Journal of the American Chemical Society* **1975**, *97* (6).
5. Lutteke, T., Carbohydrate Structure Suite (CSS): analysis of carbohydrate 3D structures derived from the PDB. *Nucleic Acids Research* **2004**, *33* (Database issue), D242-D246.
6. Kirschner, K. N.; Woods, R. J., Solvent interactions determine carbohydrate conformation. *Proceedings of the National Academy of Sciences* **2001**, *98* (19), 10541-10545.
7. Hans Heinrich Schlubach, E. L., Einwirkung von Chlorwasserstoff auf Glucose. Synthese eines Polyglucosans. *EurJOC* **1941**.
8. Nishimura, T.; Akiyoshi, K., Amylose engineering: phosphorylase-catalyzed polymerization of functional saccharide primers for glycobiomaterials. **2016**.
9. Boltje, T. J.; Buskas, T.; Boons, G.-J., Opportunities and challenges in synthetic oligosaccharide and glycoconjugate research. *Nature Chemistry* **2009**, *1* (8), 611-622.
10. Panza, M.; Pistorio, S. G.; Stine, K. J.; Demchenko, A. V., Automated Chemical Oligosaccharide Synthesis: Novel Approach to Traditional Challenges. *Chemical Reviews* **2018**, *118* (17), 8105-8150.
11. Nakatsubo, F.; Kamitakahara, H.; Hori, M., Cationic Ring-Opening Polymerization of 3,6-Di-O-benzyl- $\alpha$ -D-glucose 1,2,4-Orthopivalate and the First Chemical Synthesis of Cellulose. **1996**, *118* (7), 1677-1681.
12. Husemann, E.; Muller, G. J. M., Über Die Synthese Unverzweigter Polysaccharide. *Makromolekul Chem* **1966**, *91* (Feb), 212-&.
13. Mydock, L. K.; Demchenko, A. V., Mechanism of chemical O-glycosylation: from early studies to recent discoveries. *Org. Biomol. Chem.* **2010**, *8* (3), 497-510.
14. Kröck, L.; Esposito, D.; Castagner, B.; Wang, C.-C.; Bindschädler, P.; Seeberger, P. H., Streamlined access to conjugation-ready glycans by automated synthesis. *Chemical Science* **2012**, *3* (5), 1617.
15. Shoda, S.-I.; Uyama, H.; Kadokawa, J.-I.; Kimura, S.; Kobayashi, S., Enzymes as Green Catalysts for Precision Macromolecular Synthesis. *Chemical Reviews* **2016**, *116* (4), 2307-2413.
16. Fischer, E., Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft* **1894**, *27* (3), 2985-2993.
17. Kollman, P. A.; Kuhn, B.; Donini, O.; Perakyla, M.; Stanton, R.; Bakowies, D., Elucidating the Nature of Enzyme Catalysis Utilizing a New Twist on an Old Methodology: Quantum Mechanical-Free Energy Calculations on Chemical Reactions in Enzymes and in Aqueous Solution. **2001**, *34* (1), 72-79.
18. Bairoch, A., The ENZYME database in 2000. *Nucleic Acids Research* **2000**, *28* (1), 304-305.
19. Jun-ichi Kadokawa, Y. K., *Engineering of polysaccharides materials by phosphorylase-catalysed enzymatic chain-elongation*. Pan Stanford Publishing: 2012.
20. Kadokawa, J.-I., Precision Polysaccharide Synthesis Catalyzed by Enzymes. *Chemical Reviews* **2011**, *111* (7), 4308-4345.
21. Kadokawa, J.-I., Precision Synthesis of Functional Polysaccharide Materials by Phosphorylase-Catalyzed Enzymatic Reactions. *Polymers* **2016**, *8* (4), 138.

22. O'Neill, E. C.; Field, R. A., Enzymatic synthesis using glycoside phosphorylases. *Carbohydrate Research* **2015**, *403*, 23-37.
23. Zhong, C.; Luley-Goedl, C.; Nidetzky, B., Product solubility control in cellooligosaccharide production by coupled cellobiose and celloextrin phosphorylase. *Biotechnology and Bioengineering* **2019**, *116* (9), 2146-2155.
24. Fersht, A., *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. World Scientific publishing Co. Pvt. Ltd.: Singapore, 2017; Vol. 9.
25. Dunitz, J. D., Win Some, Lose Some - Enthalpy-Entropy Compensation in Weak Intermolecular Interactions. *Chem Biol* **1995**, *2* (11), 709-712.
26. Mobley, D. L.; Chodera, J. D.; Dill, K. A., On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. *J Chem Phys* **2006**, *125* (8), 084902.
27. Grunwald, E.; Steel, C., Solvent Reorganization and Thermodynamic Enthalpy-Entropy Compensation. *Journal of the American Chemical Society* **1995**, *117* (21), 5687-5692.
28. Breiten, B.; Lockett, M. R.; Sherman, W.; Fujita, S.; Al-Sayah, M.; Lange, H.; Bowers, C. M.; Heroux, A.; Krilov, G.; Whitesides, G. M., Water networks contribute to enthalpy/entropy compensation in protein-ligand binding. *J Am Chem Soc* **2013**, *135* (41), 15579-84.
29. Du, X.; Li, Y.; Xia, Y. L.; Ai, S. M.; Liang, J.; Sang, P.; Ji, X. L.; Liu, S. Q., Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. *Int J Mol Sci* **2016**, *17* (2).
30. Ryde, U., A fundamental view of enthalpy-entropy compensation. *Med. Chem. Commun.* **2014**, *5* (9), 1324-1336.
31. Alisdair; David; Harry; Gideon, Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochemical Journal* **2004**, *382* (3), 769-781.
32. Gabius, H.-J.; André, S.; Jiménez-Barbero, J.; Romero, A.; Solís, D., From lectin structure to functional glycomics: principles of the sugar code. *Trends in Biochemical Sciences* **2011**, *36* (6), 298-313.
33. Lemieux, R. U., How water provides the impetus for molecular recognition in aqueous solution. *Accounts Chem Res* **1996**, *29* (8), 373-380.
34. Dashnau, J. L.; Sharp, K. A.; Vanderkooi, J. M., Carbohydrate intramolecular hydrogen bonding cooperativity and its effect on water structure. *J Phys Chem B* **2005**, *109* (50), 24152-9.
35. Hudson, K. L.; Bartlett, G. J.; Diehl, R. C.; Agirre, J.; Gallagher, T.; Kiessling, L. L.; Woolfson, D. N., Carbohydrate-Aromatic Interactions in Proteins. *J Am Chem Soc* **2015**, *137* (48), 15152-60.
36. Perez, S.; Samain, D., Structure and Engineering of Celluloses. *Adv Carbohydr Chem Bi* **2010**, *64*, 25-116.
37. Munoz-Garcia, J. C.; Corbin, K. R.; Hussain, H.; Gabrielli, V.; Koev, T.; Iuga, D.; Round, A. N.; Mikkelsen, D.; Gunning, P. A.; Warren, F. J.; Khimyak, Y. Z., High Molecular Weight Mixed-Linkage Glucan as a Mechanical and Hydration Modulator of Bacterial Cellulose: Characterization by Advanced NMR Spectroscopy. *Biomacromolecules* **2019**, *20* (11), 4180-4190.
38. Nishiyama, Y.; Langan, P.; Chanzy, H., Crystal Structure and Hydrogen-Bonding System in Cellulose I $\beta$  from Synchrotron X-ray and Neutron Fiber Diffraction. *Journal of the American Chemical Society* **2002**, *124* (31), 9074-9082.
39. M. Wada, L. H., A. Isogai, and J. Sugiyama, Polymorphism of Cellulose I family: Reinvestigation of Cellulose IVi. *Biomacromolecules* **2004**, *5*, 1385-1391.
40. Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P., Crystal Structure and Hydrogen Bonding System in Cellulose Ia from Synchrotron X-ray and Neutron Fiber Diffraction. *Journal of the American Chemical Society* **2003**, *125* (47), 14300-14306.
41. Langan, P.; Nishiyama, Y.; Chanzy, H., X-ray Structure of Mercerized Cellulose II at 1 Å Resolution. **2001**, *2* (2), 410-416.
42. Wada, M.; Chanzy, H.; Nishiyama, Y.; Langan, P., Cellulose III I Crystal Structure and Hydrogen Bonding by Synchrotron X-ray and Neutron Fiber Diffraction. **2004**, *37* (23), 8548-8555.

43. Wada, M.; Heux, L.; Nishiyama, Y.; Langan, P., X-ray Crystallographic, Scanning Microprobe X-ray Diffraction, and Cross-Polarized/Magic Angle Spinning <sup>13</sup>C NMR Studies of the Structure of Cellulose III II. **2009**, *10* (2), 302-309.
44. Wada, M.; Heux, L.; Isogai, A.; Nishiyama, Y.; Chanzy, H.; Sugiyama, J., Improved structural data of cellulose III prepared in supercritical ammonia. *Macromolecules* **2001**, *34* (5), 1237-1243.
45. Zugenmaier, P., Conformation and packing of various crystalline cellulose fibers. *Progress in Polymer Science* **2001**, *26* (9), 1341-1417.
46. Graham, T., Liquid Diffusion applied to Analysis. *Phil. Trans. R. Soc.* **1861**, *151*, 183–224.
47. Lloyd, D., *The problem of gel structure*. The Chemical Catalog Co.: New York, 1926; Vol. 1.
48. Flory, P. J., Introductory lecture: levels of order in amorphous polymers. *Faraday Discussions of the Chemical Society* **1979**, *68*, 14.
49. Everett, D. H., : Basic Principles of Colloid Science. *Royal society of chemistry* **2007**.
50. Wyss, H. M., Rheology of Soft Materials. In *Fluids, Colloids, and Soft Materials: An Introduction to Soft Matter Physics*, Puertas, A. F. N. A. M., Ed. John Wiley & Sons: 2016.
51. Cazer, C., Elastic Solids. In *Notes on the Thermodynamics of Solids* Morris, J. W., Ed. 2007; pp 366-411.
52. Chen, D. T. N.; Wen, Q.; Janmey, P. A.; Crocker, J. C.; Yodh, A. G., Rheology of Soft Materials. *Annual Review of Condensed Matter Physics* **2010**, *1* (1), 301-322.
53. Lozano, V.; Hernández, R.; Ardá, A.; Jiménez-Barbero, J.; Mijangos, C.; Pérez-Pérez, M.-J., An asparagine/tryptophan organogel showing a selective response towards fluoride anions. **2011**, *21* (24), 8862.
54. Guilherme K. Gonzatti; Netz, P. A.; Fiel, L. A.; Pohlmann, A. R., Colloidal Dispersion Stability: Kinetic Modeling of Agglomeration and Aggregation. *J. Braz. Chem. Soc.* **2015**, *26* (2).
55. Borsos, A.; Gilányi, T., Transition from thermodynamically stable solution to colloid dispersion state. *Colloid and Polymer Science* **2012**, *290* (5), 473-479.
56. Ahmed, E. M., Hydrogel: Preparation, characterization, and applications: A review. *Journal of Advanced Research* **2015**, *6* (2), 105-121.
57. Estroff, L. A.; Hamilton, A. D., Water Gelation by Small Organic Molecules. *Chemical Reviews* **2004**, *104* (3), 1201-1218.
58. Jagur-Grodzinski, J., Polymeric gels and hydrogels for biomedical and pharmaceutical applications. **2009**, n/a-n/a.
59. Yu, G.; Yan, X.; Han, C.; Huang, F., Characterization of supramolecular gels. **2013**, *42* (16), 6697.
60. Hoffman, A. S., Hydrogels for biomedical applications. *Advanced Drug Delivery Reviews* **2012**, *64*, 18-23.
61. Myszka, D. G., Kinetic analysis of macromolecular interactions using surface plasmon resonance biosensors. **1997**, *8* (1), 50-57.
62. Kabiri, M.; Unsworth, L. D., Application of Isothermal Titration Calorimetry for Characterizing Thermodynamic Parameters of Biomolecular Interactions: Peptide Self-Assembly and Protein Adsorption Case Studies. *Biomacromolecules* **2014**, *15* (10), 3463-3473.
63. Woods, R. J.; Tessier, M. B., Computational glycoscience: characterizing the spatial and temporal properties of glycans and glycan–protein complexes. *Current Opinion in Structural Biology* **2010**, *20* (5), 575-583.
64. Yip, K. M.; Fischer, N.; Paknia, E.; Chari, A.; Stark, H., Breaking the next Cryo-EM resolution barrier – Atomic resolution determination of proteins! Cold Spring Harbor Laboratory: 2020.
65. Felli, I. C.; Brutscher, B., Recent Advances in Solution NMR: Fast Methods and Heteronuclear Direct Detection. **2009**, *10* (9-10), 1356-1368.

66. Lin, Y. L.; Cheng, Y. S.; Ho, C. I.; Guo, Z. H.; Huang, S. J.; Org, M. L.; Oss, A.; Samoson, A.; Chan, J. C. C., Preparation of fibril nuclei of beta-amyloid peptides in reverse micelles. *Chem Commun (Camb)* **2018**, 54 (74), 10459-10462.
67. Chien, P.-H.; Griffith, K. J.; Liu, H.; Gan, Z.; Hu, Y.-Y., Recent Advances in Solid-State Nuclear Magnetic Resonance Techniques for Materials Research. *Annual Review of Materials Research* **2020**, 50 (1), 493-520.
68. Chen, P.; Albert, B. J.; Gao, C.; Alaniva, N.; Price, L. E.; Scott, F. J.; Saliba, E. P.; Sesti, E. L.; Judge, P. T.; Fisher, E. W.; Barnes, A. B., Magic angle spinning spheres. *Science Advances* **2018**, 4 (9), eaau1540.
69. Claridge, T. D. W., *High-Resolution NMR Techniques in Organic Chemistry*. 3rd ed.; Elsevier: 2016.
70. Keeler, J., *Understanding NMR Spectroscopy* - 2nd ed.; Wiley: 2010.
71. Levitt, M. H., *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. 2nd ed.; Wiley: 2008.
72. Bryant, R. G., The NMR time scale. **1983**, 60 (11), 933.
73. Ni, F., Recent developments in transferred NOE methods. **1994**, 26, 517-606.
74. Post, C., Exchange-transferred NOE spectroscopy and bound ligand structure determination. **2003**, 13 (5), 581-588.
75. Muñoz-García, J. C. Insights on the Structure and Dynamics of Glycosaminoglycans and their Interactions with Langerin: NMR and Computational Studies. UNIVERSIDAD DE SEVILLA, 2013.
76. Aksel A. Bothner-By, R. L. S., and Ju-mee Lee, Structure Determination of a Tetrasaccharide: Transient Nuclear Overhauser Effects in the Rotating Frame. *J. Am. Chem. Soc* **1984**, (106), 811-813.
77. Mayer, M.; Meyer, B., Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew Chem Int Edit* **1999**, 38 (12), 1784-1788.
78. Samuel Walpole, S. M., Ridvan Nepravishta and Jesus Angulo, STD NMR as a Technique for Ligand Screening and Structural Studies. In *Methods in Enzymology*, 2019; Vol. 615, pp 423-451.
79. Angulo, J.; Enríquez-Navas, P. M.; Nieto, P. M., Ligand-Receptor Binding Affinities from Saturation Transfer Difference (STD) NMR Spectroscopy: The Binding Isotherm of STD Initial Growth Rates. *Chemistry - A European Journal* **2010**, 16 (26), 7803-7812.
80. Brand, T.; Cabrita, E. J.; Berger, S., Intermolecular interaction as investigated by NOE and diffusion studies. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2005**, 46 (4), 159-196.
81. Mayer, M.; Meyer, B., Group Epitope Mapping by Saturation Transfer Difference NMR To Identify Segments of a Ligand in Direct Contact with a Protein Receptor. *Journal of the American Chemical Society* **2001**, 123 (25), 6108-6117.
82. Mayer, M.; James, T. L., NMR-Based Characterization of Phenothiazines as a RNA Binding Scaffold†. *Journal of the American Chemical Society* **2004**, 126 (13), 4453-4460.
83. Monaco, S.; Tailford, L. E.; Juge, N.; Angulo, J., Differential Epitope Mapping by STD NMR Spectroscopy To Reveal the Nature of Protein-Ligand Contacts. *Angewandte Chemie International Edition* **2017**, 56 (48), 15289-15293.
84. Han, B.; Liu, Y.; Ginzinger, S. W.; Wishart, D. S., SHIFTX2: significantly improved protein chemical shift prediction. *J Biomol NMR* **2011**, 50 (1), 43-57.
85. Nepravishta, R.; Walpole, S.; Tailford, L.; Juge, N.; Angulo, J., Deriving Ligand Orientation in Weak Protein-Ligand Complexes by DEEP-STD NMR Spectroscopy in the Absence of Protein Chemical-Shift Assignment. *ChemBioChem* **2019**, 20 (3), 340-344.
86. Pickhardt, M.; Larbig, G.; Khlistunova, I.; Coksezen, A.; Meyer, B.; Mandelkow, E.-M.; Schmidt, B.; Mandelkow, E., Phenylthiazolyl-Hydrazide and Its Derivatives Are Potent Inhibitors of  $\tau$  Aggregation and Toxicity in Vitro and in Cells†. *Biochemistry* **2007**, 46 (35), 10016-10023.
87. Duer, M. J., *Introduction to Solid-State NMR Spectroscopy*.

88. Andrew, E. R.; Bradbury, A.; Eades, R. G., Nuclear Magnetic Resonance Spectra from a Crystal rotated at High Speed. *Nature* **1958**, *182* (4650), 1659-1659.
89. Andrew, E., Rotational narrowing of nuclear magnetic resonance spectra. *Arch. Sci* **1959**, *12* (103).
90. Laws, D. D.; Bitter, H. M. L.; Jerschow, A., Solid-state NMR spectroscopic methods in chemistry. *Angew Chem Int Edit* **2002**, *41* (17), 3096-3129.
91. Demers, J.-P.; Chevelkov, V.; Lange, A., Progress in correlation spectroscopy at ultra-fast magic-angle spinning: Basic building blocks and complex experiments for the study of protein structure and dynamics. **2011**, *40* (3), 101-113.
92. Baldwin, A. J.; Egan, D. L.; Warren, F. J.; Barker, P. D.; Dobson, C. M.; Butterworth, P. J.; Ellis, P. R., Investigating the mechanisms of amylolysis of starch granules by solution-state NMR. *Biomacromolecules* **2015**, *16* (5), 1614-21.
93. Mitchell, D. J.; Evans, J. N. S., Improved heteronuclear decoupling in REDOR with the use of TPPM. **1998**, *292* (4-6), 656-660.
94. Bennett, A. E.; Rienstra, C. M.; Auger, M.; Lakshmi, K. V.; Griffin, R. G., Heteronuclear decoupling in rotating solids. *The Journal of Chemical Physics* **1995**, *103* (16), 6951-6958.
95. Detken, A.; Hardy, E. H.; Ernst, M.; Meier, B. H., Simple and efficient decoupling in magic-angle spinning solid-state NMR: the XiX scheme. **2002**, *356* (3-4), 298-304.
96. Fung, B. M.; Khitrin, A. K.; Ermolaev, K., An Improved Broadband Decoupling Sequence for Liquid Crystals and Solids. **2000**, *142* (1), 97-101.
97. Lesage, A.; Duma, L.; Sakellariou, D.; Emsley, L., Improved Resolution in Proton NMR Spectroscopy of Powdered Solids. **2001**, *123* (24), 5747-5752.
98. Shu, J.; Li, P.; Chen, Q.; Zhang, S., Quantitative Measurement of Polymer Compositions by NMR Spectroscopy: Targeting Polymers with Marked Difference in Phase Mobility. **2010**, *43* (21), 8993-8996.
99. Kolodziejcki, W.; Klinowski, J., Kinetics of Cross-Polarization in Solid-State NMR: A Guide for Chemists. *Chemical Reviews* **2002**, *102* (3), 613-628.
100. White, P. B.; Wang, T.; Park, Y. B.; Cosgrove, D. J.; Hong, M., Water-Polysaccharide Interactions in the Primary Cell Wall of Arabidopsis thaliana from Polarization Transfer Solid-State NMR. **2014**, *136* (29), 10399-10409.
101. Lesley E. Smart, E. A. M., *Solid State Chemistry: An Introduction*. Third Edition ed.; CRC Press: 2005.
102. R E Dinnebier, S. J. L. B., Armel Le Bail, Ian Madsen, Lachlan M D Cranswick, *Powder Diffraction: Theory and Practice*. Royal Society of Chemistry 2008.
103. Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A., Mercury CSD 2.0- new features for the visualization and investigation of crystal structures. *Journal of Applied Crystallography* **2008**, *41* (2), 466-470.
104. Hans-Dieter Höltje, W. S., Didier Rognan, Gerd Folkers, *Molecular Modeling: Basic Principles and Applications*. WILEY-VCH: 2008.
105. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry* **2004**, *47* (7), 1739-1749.
106. Garrett M. Morris, L.-W. M., *Molecular Docking*. Humana Press: 2008; Vol. 443.
107. Nurisso, A.; Kozmon, S.; Imberty, A., Comparison of docking methods for carbohydrate binding in calcium-dependent lectins and prediction of the carbohydrate binding mode to sea cucumber lectin CEL-III. **2008**, *34* (4), 469-479.

108. Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A., OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *Journal of Chemical Theory and Computation* **2016**, *12* (1), 281-296.
109. Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R., Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *Journal of Medicinal Chemistry* **2006**, *49* (2), 534-553.
110. Baleja, J. D.; Moulton, J.; Sykes, B. D., Distance measurement and structure refinement with NOE data. **1990**, *87* (2), 375-384.
111. Artimo, P.; Jonnalagedda, M.; Arnold, K.; Baratin, D.; Csardi, G.; De Castro, E.; Duvaud, S.; Flegel, V.; Fortier, A.; Gasteiger, E.; Grosdidier, A.; Hernandez, C.; Ioannidis, V.; Kuznetsov, D.; Liechti, R.; Moretti, S.; Mostaguir, K.; Redaschi, N.; Rossier, G.; Xenarios, I.; Stockinger, H., ExpASY: SIB bioinformatics resource portal. *Nucleic Acids Research* **2012**, *40* (W1), W597-W603.
112. Nakai, H.; Kitaoka, M.; Svensson, B.; Ohtsubo, K. I., Recent development of phosphorylases possessing large potential for oligosaccharide synthesis. *Current Opinion in Chemical Biology* **2013**, *17* (2), 301-309.
113. Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B., The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research* **2009**, *37* (Database), D233-D238.
114. Nakai, H.; Hachem, M. A.; Petersen, B. O.; Westphal, Y.; Mannerstedt, K.; Baumann, M. J.; Dilokpimol, A.; Schols, H. A.; Duus, J. Ø.; Svensson, B., Efficient chemoenzymatic oligosaccharide synthesis by reverse phosphorolysis using cellobiose phosphorylase and cellodextrin phosphorylase from *Clostridium thermocellum*. *Biochimie* **2010**, *92* (12), 1818-1826.
115. Sheth, K.; Alexander, J., Cellodextrin phosphorylase from *Clostridium thermocellum*. **1967**, *148* (3), 808-810.
116. Reichenbecher, M.; Lottspeich, F.; Bronnenmeier, K., Purification and Properties of a Cellobiose Phosphorylase (CepA) and a Cellodextrin Phosphorylase (CepB) from the Cellulolytic Thermophile *Clostridium Stercorarium*. **1997**, *247* (1), 262-267.
117. Jianrong Lou, K. A. D., Herbert J. Strobel, Cellobiose and Cellodextrin Metabolism by the Ruminant Bacterium *Ruminococcus albus*. *CURRENT MICROBIOLOGY* **1997**, *35*, 221-227.
118. Wu, Y.; Mao, G.; Fan, H.; Song, A.; Zhang, Y.-H. P.; Chen, H., Biochemical properties of GH94 cellodextrin phosphorylase THA\_1941 from a thermophilic eubacterium *Thermosiphon africanus* TCF52B with cellobiose phosphorylase activity. *Scientific Reports* **2017**, *7* (1).
119. Liu, N.; Fosses, A.; Kampik, C.; Parsiegla, G.; Denis, Y.; Vita, N.; Fierobe, H.-P.; Perret, S., In vitro and in vivo exploration of the cellobiose and cellodextrin phosphorylases panel in *Ruminococcus cellulosilyticus*: implication for cellulose catabolism. *Biotechnology for Biofuels* **2019**, *12* (1).
120. Krishnareddy, M.; Kim, Y.-K.; Kitaoka, M.; Mori, Y.; Hayashi, K., Cellodextrin Phosphorylase from *Clostridium thermocellum* YM4 Strain Expressed in *Escherichia coli*. **2002**, *49* (1), 1-8.
121. Sawano, T.; Saburi, W.; Hamura, K.; Matsui, H.; Mori, H., Characterization of *Ruminococcus albus* cellodextrin phosphorylase and identification of a key phenylalanine residue for acceptor specificity and affinity to the phosphate group. **2013**, *280* (18), 4463-4473.
122. Petrović, D. M.; Kok, I.; Woortman, A. J. J.; Ćirić, J.; Loos, K., Characterization of Oligocellulose Synthesized by Reverse Phosphorolysis Using Different Cellodextrin Phosphorylases. *Analytical Chemistry* **2015**, *87* (19), 9639-9646.
123. Serizawa, T.; Fukaya, Y.; Sawada, T., Self-Assembly of Cellulose Oligomers into Nanoribbon Network Structures Based on Kinetic Control of Enzymatic Oligomerization. *Langmuir* **2017**, *33* (46), 13415-13422.

124. Samain, E.; Lancelon-Pin, C.; Férido, F.; Moreau, V. S., 1995 #25); Chanzy, H.; Heyraud, A.; Driguez, H., Phosphorolytic synthesis of cellodextrins. **1995**, *271* (2), 217-226.
125. Kawaguchi, T.; Ikeuchi, Y.; Tsutsumi, N.; Kan, A.; Sumitani, J.-I.; Arai, M., Cloning, nucleotide sequence, and expression of the *Clostridium thermocellum* cellodextrin phosphorylase gene and its application to synthesis of cellulase inhibitors. **1998**, *85* (2), 144-149.
126. O'Neill, E. C.; Pergolizzi, G.; Stevenson, C. E. M.; Lawson, D. M.; Nepogodiev, S. A.; Field, R. A., Cellodextrin phosphorylase from *Ruminiclostridium thermocellum*: X-ray crystal structure and substrate specificity analysis. *Carbohydrate Research* **2017**, *451*, 118-132.
127. Arai, M.; Tanaka, K.; Kawaguchi, T., Purification and properties of cellodextrin phosphorylase from *Clostridium thermocellum*. **1994**, *77* (3), 239-242.
128. Hiraishi, M.; Igarashi, K.; Kimura, S.; Wada, M.; Kitaoka, M.; Samejima, M., Synthesis of highly ordered cellulose II in vitro using cellodextrin phosphorylase. *Carbohydrate Research* **2009**, *344* (18), 2468-2473.
129. Moreau, V.; Viladot, J.-L.; Samain, E.; Planas, A.; Driguez, H., Design and chemoenzymatic synthesis of thiooligosaccharide inhibitors of 1,3:1,4- $\beta$ -D-glucanases. **1996**, *4* (11), 1849-1855.
130. Shintate, K.; Kitaoka, M.; Kim, Y.-K.; Hayashi, K., Enzymatic synthesis of a library of  $\beta$ -(1 $\rightarrow$ 4) hetero- D-glucose and D-xylose-based oligosaccharides employing cellodextrin phosphorylase. *Carbohydrate Research* **2003**, *338* (19), 1981-1990.
131. Singh, R. P.; Pergolizzi, G.; Nepogodiev, S. A.; Andrade, P.; Kuhaudomlarp, S.; Field, R. A., Preparative and Kinetic Analysis of  $\beta$ -1,4- and  $\beta$ -1,3-Glucan Phosphorylases Informs Access to Human Milk Oligosaccharide Fragments and Analogues Thereof. *ChemBioChem* **2020**, *21* (7), 1043-1049.
132. Yataka, Y.; Sawada, T.; Serizawa, T., Enzymatic synthesis and post-functionalization of two-dimensional crystalline cellulose oligomers with surface-reactive groups. **2015**, *51* (63), 12525-12528.
133. Serizawa, T.; Kato, M.; Okura, H.; Sawada, T.; Wada, M., Hydrolytic activities of artificial nanocellulose synthesized via phosphorylase-catalyzed enzymatic reactions. *Polymer Journal* **2016**, *48* (4), 539-544.
134. Hata, Y.; Kojima, T.; Koizumi, T.; Okura, H.; Sakai, T.; Sawada, T.; Serizawa, T., Enzymatic Synthesis of Cellulose Oligomer Hydrogels Composed of Crystalline Nanoribbon Networks under Macromolecular Crowding Conditions. **2017**, *6* (2), 165-170.
135. Hata, Y.; Sawada, T.; Serizawa, T., Effect of solution viscosity on the production of nanoribbon network hydrogels composed of enzymatically synthesized cellulose oligomers under macromolecular crowding conditions. *Polymer Journal* **2017**, *49* (7), 575-581.
136. Hata, Y.; Fukaya, Y.; Sawada, T.; Nishiura, M.; Serizawa, T., Biocatalytic oligomerization-induced self-assembly of crystalline cellulose oligomers into nanoribbon networks assisted by organic solvents. *Beilstein Journal of Nanotechnology* **2019**, *10*, 1778-1788.
137. Hata, Y.; Sawada, T.; Marubayashi, H.; Nojima, S.; Serizawa, T., Temperature-Directed Assembly of Crystalline Cellulose Oligomers into Kinetically Trapped Structures during Biocatalytic Synthesis. *Langmuir* **2019**, *35* (21), 7026-7034.
138. Zhang, Y. H. P.; Lynd, L. R., Biosynthesis of radiolabeled cellodextrins by the *Clostridium thermocellum* cellobiose and cellodextrin phosphorylases for measurement of intracellular sugars. **2006**, *70* (1), 123-129.
139. Nohara, T.; Sawada, T.; Tanaka, H.; Serizawa, T., Enzymatic Synthesis of Oligo(ethylene glycol)-Bearing Cellulose Oligomers for in Situ Formation of Hydrogels with Crystalline Nanoribbon Network Structures. **2016**.
140. Yataka, Y.; Sawada, T.; Serizawa, T., Multidimensional Self-Assembled Structures of Alkylated Cellulose Oligomers Synthesized via in Vitro Enzymatic Reactions. **2016**, *32* (39), 10120-10125.

141. Nohara, T.; Sawada, T.; Tanaka, H.; Serizawa, T., Enzymatic synthesis and protein adsorption properties of crystalline nanoribbons composed of cellulose oligomer derivatives with primary amino groups. *J Biomater Sci Polym Ed* **2017**, *28* (10-12), 925-938.
142. Wang, J.; Niu, J.; Sawada, T.; Shao, Z.; Serizawa, T., A Bottom-Up Synthesis of Vinyl-Cellulose Nanosheets and Their Nanocomposite Hydrogels with Enhanced Strength. *Biomacromolecules* **2017**, *18* (12), 4196-4205.
143. Adharis, A.; Petrović, D. M.; Özdamar, I.; Woortman, A. J. J.; Loos, K., Environmentally friendly pathways towards the synthesis of vinyl-based oligocelluloses. *Carbohydrate Polymers* **2018**, *193*, 196-204.
144. Serizawa, T.; Maeda, T.; Sawada, T., Neutralization-Induced Self-Assembly of Cellulose Oligomers into Antibiofouling Crystalline Nanoribbon Networks in Complex Mixtures. *ACS Macro Letters* **2020**, 301-305.
145. Ambar; Kitaoka, M.; Hayashi, K., Synthesis of a Cellobiosylated Dimer and Trimer and of Cellobiose-Coated Polyamidoamine (PAMAM) Dendrimers to Study Accessibility of an Enzyme, Cellobiose Phosphorylase. **2003**, *2003* (13), 2462-2470.
146. ALEXANDER, K. S. A. J. K., Purification and Properties of  $\beta$ -1,4-Oligoglucan:Orthophosphate Glucosyltransferase from *Clostridium thermocellum*. *Journal Biological Chemistry* **1969**.
147. Pergolizzi, G.; Kuhadomlarp, S.; Kalita, E.; Field, R. A., Glycan Phosphorylases in Multi-Enzyme Synthetic Processes. *Protein & Peptide Letters* **2017**, *24* (8).
148. Hrmova, M.; Fincher, G. B.; Viladot, J.-L.; Planas, A.; Driguez, H., Chemoenzymic synthesis of (1 $\rightarrow$ 3,1 $\rightarrow$ 4)- $\beta$ -D-glucooligosaccharides for subsite mapping of (1 $\rightarrow$ 3,1 $\rightarrow$ 4)- $\beta$ -D-glucan endohydrolases. **1998**, (21), 3571-3576.
149. Maestro *Schrödinger*, LLC, New York, 2020.
150. Nerinckx, W.; Desmet, T.; Claeysens, M., A hydrophobic platform as a mechanistically relevant transition state stabilising factor appears to be present in the active centre of all glycoside hydrolases. *FEBS Letters* **2003**, *538* (1-3), 1-7.
151. O'Neill, E. C.; Pergolizzi, G.; Stevenson, C. E. M.; Lawson, D. M.; Nepogodiev, S. A.; Field, R. A., Cellodextrin phosphorylase from *Ruminiclostridium thermocellum* : X-ray crystal structure and substrate specificity analysis. *Carbohydrate Research* **2017**, *451*, 118-132.
152. Field, R.; Khimiyak, Y.; Angulo, J.; Eichhorn, S. J.; Harniman, R.; Johns, M.; A; Nigmatullin, R.; Fábíán, L.; Iuga, D.; Nepogodiev, S.; Gabrielli, V.; Pergolizzi, G.; Munoz, J.; de Andrade, P., Bottom-up chemoenzymatic synthesis towards novel fluorinated cellulose-like materials. *cheRxiv* **2020**.
153. Bubb, W. A., NMR spectroscopy in the study of carbohydrates: Characterizing the structural complexity. *Concepts in Magnetic Resonance* **2003**, *19A* (1), 1-19.
154. Böhm, M.; Bohne-Lang, A.; Frank, M.; Loss, A.; Rojas-Macias, M. A.; Lütteke, T., Glycosciences.DB: an annotated data collection linking glycomics and proteomics data (2018 update). *Nucleic Acids Research* **2019**, *47* (D1), D1195-D1201.
155. Monaco, S.; Tailford, L. E.; Juge, N.; Angulo, J., Differential Epitope Mapping by STD NMR Spectroscopy To Reveal the Nature of Protein-Ligand Contacts. *Angew Chem Int Ed Engl* **2017**, *56* (48), 15289-15293.
156. Tanaka, T.; Nagai, H.; Noguchi, M.; Kobayashi, A.; Shoda, S.-I., One-step conversion of unprotected sugars to  $\beta$ -glycosyl azides using 2-chloroimidazolium salt in aqueous solution. *Chemical Communications* **2009**, (23), 3378.
157. Hidaka, M.; Kitaoka, M.; Hayashi, K.; Wakagi, T.; Shoun, H.; Fushinobu, S., Structural dissection of the reaction mechanism of cellobiose phosphorylase. **2006**, *398* (1), 37-43.
158. Hai Tran, G.; Desmet, T.; De Groeve, M. R. M.; Soetaert, W., Probing the active site of cellobiose phosphorylase from *Clostridium stercorarium*: Kinetic characterization, ligand docking, and site-directed mutagenesis. **2011**, *27* (2), 326-332.

159. Becattini, B.; Pellicchia, M., SAR by ILOEs: An NMR-Based Approach to Reverse Chemical Genetics. *Chemistry - A European Journal* **2006**, *12* (10), 2658-2662.
160. Dixon, D. A.; Smart, B. E.; Krusic, P. J.; Matsuzawa, N., Bond energies in organofluorine systems: applications to Teflon® and fullerenes. **1995**, *72* (2), 209-214.
161. Biffinger, J. C.; Kim, H. W.; Dimagno, S. G., The Polar Hydrophobicity of Fluorinated Compounds. *ChemBioChem* **2004**, *5* (5), 622-627.
162. Hoang, K. C.; Mecozzi, S., Aqueous Solubilization of Highly Fluorinated Molecules by Semifluorinated Surfactants. **2004**, *20* (18), 7347-7350.
163. John A. Gladysz, D. P. C., Istvan T. Horvath, *Handbook of Fluorous Chemistry*. Wiley-VCH: 2004.
164. Fried, J. R.; Hu, N., The molecular basis of CO<sub>2</sub> interaction with polymers containing fluorinated groups: computational chemistry of model compounds and molecular simulation of poly[bis(2,2,2-trifluoroethoxy)phosphazene]. **2003**, *44* (15), 4363-4372.
165. Pagliaro, M.; Ciriminna, R., New fluorinated functional materials. **2005**, *15* (47), 4981.
166. Chan, K. K. J.; O'Hagan, D., The Rare Fluorinated Natural Products and Biotechnological Prospects for Fluorine Enzymology. Elsevier: 2012; pp 219-235.
167. O'Hagan, D.; Deng, H., Enzymatic Fluorination and Biotechnological Developments of the Fluorinase. *Chemical Reviews* **2015**, *115* (2), 634-649.
168. Cunha, A. G.; Gandini, A., Turning polysaccharides into hydrophobic materials: a critical review. Part 1. Cellulose. *Cellulose* **2010**, *17* (5), 875-889.
169. Gonçalves, G.; Marques, P. A. A. P.; Trindade, T.; Neto, C. P.; Gandini, A., Superhydrophobic cellulose nanocomposites. **2008**, *324* (1-2), 42-46.
170. Zhang, W.; Chen, S.; Hu, W.; Zhou, B.; Yang, Z.; Yin, N.; Wang, H., Facile fabrication of flexible magnetic nanohybrid membrane with amphiphobic surface based on bacterial cellulose. *Carbohydrate Polymers* **2011**, *86* (4), 1760-1767.
171. Kasuya, N.; Iiyama, K.; Meshitsuka, G.; Okano, T., A structural study of fluorinated cellulose: crystallization of fluorinated cellulose by conversion treatments used for cellulose. **1997**, *34* (4), 229-234.
172. O'Hagan, D.; Schaffrath, C.; Cobb, S. L.; Hamilton, J. T. G.; Murphy, C. D., Biosynthesis of an organofluorine molecule. *Nature* **2002**, *416* (6878), 279-279.
173. Deng, H.; Ma, L.; Bandaranayaka, N.; Qin, Z.; Mann, G.; Kyeremeh, K.; Yu, Y.; Shepherd, T.; Naismith, J. H.; O'Hagan, D., Identification of Fluorinases from *Streptomyces* sp. MA37, *Nocardia brasiliensis*, and *Actinoplanes* sp. N902-109 by Genome Mining. *ChemBioChem* **2014**, *15* (3), 364-368.
174. Huang, S.; Ma, L.; Tong, M. H.; Yu, Y.; O'Hagan, D.; Deng, H., Fluoroacetate biosynthesis from the marine-derived bacterium *Streptomyces xinghaiensis* NRRL B-24674. *Org. Biomol. Chem.* **2014**, *12* (27), 4828-4831.
175. Ma, L.; Li, Y.; Meng, L.; Deng, H.; Li, Y.; Zhang, Q.; Diao, A., Biological fluorination from the sea: discovery of a SAM-dependent nucleophilic fluorinating enzyme from the marine-derived bacterium *Streptomyces xinghaiensis* NRRL B24674. **2016**, *6* (32), 27047-27051.
176. Tessa Keenan, F. P., Julien Malassis, Clement Fontenelle, Jean-Baptiste Vendeville, Wendy Offen, Peter Both, kun huang, Andrea Marchesi, Alex Heyam, Carl Young, Simon J. Charnock, Gideon J. Davies, Bruno Linclau, Sabine Flitsch, Martin Fascione, Profiling Substrate Promiscuity of Wild-Type Sugar Kinases for Multi-Fluorinated Monosaccharides. *ChemRxiv* **2020**, <https://doi.org/10.26434/chemrxiv.11950647.v1>.
177. Yu, H.; Thon, V.; Lau, K.; Cai, L.; Chen, Y.; Mu, S.; Li, Y.; Wang, P. G.; Chen, X., Highly efficient chemoenzymatic synthesis of  $\beta$ 1-3-linked galactosides. **2010**, *46* (40), 7507.
178. Yan, J.; Chen, X.; Wang, F.; Cao, H., Chemoenzymatic synthesis of mono- and di-fluorinated Thomsen-Friedenreich (T) antigens and their sialylated derivatives. *Org. Biomol. Chem.* **2013**, *11* (5), 842-848.

179. Li, L.; Liu, Y.; Li, T.; Wang, W.; Yu, Z.; Ma, C.; Qu, J.; Zhao, W.; Chen, X.; Wang, P. G., Efficient chemoenzymatic synthesis of novel galacto-N-biose derivatives and their sialylated forms. **2015**, *51* (51), 10310-10313.
180. Dalvit, C.; Vulpetti, A., Ligand-Based Fluorine NMR Screening: Principles and Applications in Drug Discovery Projects. *Journal of Medicinal Chemistry* **2019**, *62* (5), 2218-2244.
181. Pethrick, R. A., Molecular motion in polymer systems. **2002**, *6* (3), 221-225.
182. Price, W. S., *NMR Studies of Translational Motion: Principles and Applications*. Cambridge University Press: 2009.
183. Pethrick, R. A., Molecular motion of polymers in solution. *Science Progress* **1975**, 599-631.
184. Singer, P. M.; Asthagiri, D.; Chen, Z.; Valiya Parambathu, A.; Hirasaki, G. J.; Chapman, W. G., Role of internal motions and molecular geometry on the NMR relaxation of hydrocarbons. *The Journal of Chemical Physics* **2018**, *148* (16), 164507.
185. Sillen, A.; Leroy, A.; Wieruszkeski, J. M.; Loyens, A.; Beauvillain, J. C.; Buee, L.; Landrieu, I.; Lippens, G., Regions of tau implicated in the paired helical fragment core as defined by NMR. *Chembiochem* **2005**, *6* (10), 1849-56.
186. Vanderhart, D. L.; Earl, W. L.; Garroway, A. N., Resolution in <sup>13</sup>C NMR of organic solids using high-power proton decoupling and magic-angle sample spinning. **1981**, *44* (2), 361-401.
187. William L. Earl, D. L. V., Observation by High-Resolution Carbon-13 Nuclear Magnetic Resonance of Cellulose I Related to Morphology and Crystal Structure. *Macromolecules* **1981**, *14*, 570 - 574.
188. Atalla, D. L. V. a. R. H., Studies of Microstructure in Native Celluloses Using Solid-State <sup>13</sup>C NMR. *Macromolecules* **1984**, *17*, 1465-1472.
189. Earl, W. L.; VanderHart, D. L., High resolution, magic angle sampling spinning carbon-13 NMR of solid cellulose I. *Journal of the American Chemical Society* **1980**, *102* (9), 3251-3252.
190. R. H. Atalla, J. C. G., <sup>13</sup>C NMR Spectra of Cellulose Polymorphs. *Journal American Chemical Society* **1980**, *102*:9, 3249 - 3251.
191. Isogai, A.; Usuda, M.; Kato, T.; Uryu, T.; Atalla, R. H., Solid-state CP/MAS carbon-13 NMR study of cellulose polymorphs. *Macromolecules* **1989**, *22* (7), 3168-3172.
192. Atalla, R. H.; Vanderhart, D. L., Native Cellulose: A Composite of Two Distinct Crystalline Forms. *Science* **1984**, *223* (4633), 283-285.
193. Teer, R.; Lippmaa, E., Solid state <sup>13</sup>C-NMR of cellulose. A relaxation study. *Polymer Bulletin* **1984**, *12* (4), 315-318.
194. Fumitaka Horii, A. H. a. R. K., CP/MAS <sup>13</sup>C NMR Spectra of the Crystalline Components of Native Celluloses. *Macromolecules* **1987**, *20*, 2117-2120.
195. Gary E. Maciel, W. L. K., Maria S. Bertrán, and Bruce E. Dale, <sup>13</sup>C NMR and Order in Cellulose. *Macromolecules* **1982**, *15*, 686-687.
196. Horii, F.; Hirai, A.; Kitamaru, R., Solid-state <sup>13</sup>C-NMR study of conformations of oligosaccharides and cellulose. *Polymer Bulletin* **1983**, *10* (7-8), 357-361.
197. John J. Cael, D. L. W. K., and Shyam S. Bhattacharjee, Cellulose Crystallites: A Perspective from Solid-State <sup>13</sup>C NMR. *Macromolecules* **1985**, *18*, 819-821.
198. R. L. Dudley, C. A. F., P. J. Stephenson, Y. Deslandes, G. K. Hamer, and R. H. Marchessault, High-resolution carbon-13 CP/MAS NMR spectra of solid cellulose oligomers and the structure of cellulose II. *Journal American Chemistry Society* **1983**, *105*, 2469 - 2472.
199. Kulshreshtha, A. K., 2—a Review of the Literature on the Formation of Cellulose IV, Its Structure, and Its Significance in the Technology of Rayon Manufacture. *The Journal of The Textile Institute* **2008**, *70* (1), 13-18.
200. Newman, R. H., Simulation of X-ray diffractograms relevant to the purported polymorphs cellulose IVI and IVII. **2008**, *15* (6), 769-778.

201. Kono, H.; Numata, Y., Two-dimensional spin-exchange solid-state NMR study of the crystal structure of cellulose II. **2004**, *45* (13), 4541-4547.
202. French, A. D.; Johnson, G. P.; Kelterer, A.-M.; Csonka, G. I., Fluorinated cellobiose and maltose as stand-ins for energy surface calculations. **2005**, *16* (2), 577-586.
203. Larsson, P. T.; Wickholm, K.; Iversen, T., A CP/MAS<sup>13</sup>C NMR investigation of molecular ordering in celluloses. *Carbohydrate Research* **1997**, *302* (1-2), 19-25.
204. Mittal, A.; Katahira, R.; Himmel, M. E.; Johnson, D. K., Effects of alkaline or liquid-ammonia treatment on crystalline cellulose: changes in crystalline structure and effects on enzymatic digestibility. **2011**, *4* (1), 41.
205. Muñoz-García, J. C.; Corbin, K. R.; Hussain, H.; Gabrielli, V.; Koev, T.; Iuga, D.; Round, A. N.; Mikkelsen, D.; Gunning, P. A.; Warren, F. J.; Khimyak, Y. Z., High Molecular Weight Mixed-Linkage Glucan as a Mechanical and Hydration Modulator of Bacterial Cellulose: Characterization by Advanced NMR Spectroscopy. *Biomacromolecules* **2019**.
206. Wang, T.; Hong, M., Solid-state NMR investigations of cellulose structure and interactions with matrix polysaccharides in plant primary cell walls. *J Exp Bot* **2016**, *67* (2), 503-14.
207. Calabrese, V.; Muñoz-García, J. C.; Schmitt, J.; Da Silva, M. A.; Scott, J. L.; Angulo, J.; Khimyak, Y. Z.; Edler, K. J., Understanding heat driven gelation of anionic cellulose nanofibrils: Combining saturation transfer difference (STD) NMR, small angle X-ray scattering (SAXS) and rheology. *Journal of Colloid and Interface Science* **2019**, *535*, 205-213.
208. Langan, P.; Nishiyama, Y.; Chanzy, H., A Revised Structure and Hydrogen-Bonding System in Cellulose II from a Neutron Fiber Diffraction Analysis. **1999**, *121* (43), 9940-9946.
209. Dalvit, C.; Vulpetti, A., Weak Intermolecular Hydrogen Bonds with Fluorine: Detection and Implications for Enzymatic/Chemical Reactions, Chemical Properties, and Ligand/Protein Fluorine NMR Screening. **2016**, *22* (22), 7592-7601.
210. Andrade, J. D., Interfacial phenomena and biomaterials. *Med Instrum* **1973**, *7* (2), 110-9.
211. Fyfe, C. A.; Blazek, A. I., Investigation of Hydrogel Formation from Hydroxypropylmethylcellulose (HPMC) by NMR Spectroscopy and NMR Imaging Techniques. **1997**, *30* (20), 6230-6237.
212. Blinc, A.; Lahajnar, G.; Blinc, R.; Zidansek, A.; Sepe, A., Proton NMR study of the state of water in fibrin gels, plasma, and blood clots. *Magn Reson Med* **1990**, *14* (1), 105-22.
213. Tanaka, T., Kinetics of Phase-Transition in Polymer Gels. *Physica A* **1986**, *140* (1-2), 261-268.
214. McCrystal, C. B.; Ford, J. L.; Rajabi-Siahboomi, A. R., A study on the interaction of water and cellulose ethers using differential scanning calorimetry. **1997**, *294* (1), 91-98.
215. Saša Baumgartner, G. L., Ana Sepe, and Julijana Kristl, <Investigation of the State and Dynamics of Water in Hydrogels of Cellulose Ethers by <sup>1</sup>H NMR Spectroscopy.pdf>. *AAPS PharmSciTech* **2002**, (36), 1-8.
216. Jhon, M. S.; Andrade, J. D., Water and Hydrogels. *J Biomed Mater Res* **1973**, *7* (6), 509-522.
217. Krochta, R. S. a. M., Plasticizers in edible films and coatings In *Innovations in Food Packaging* Elsevier Ltd 2005.
218. Hatakeyama, H.; Hatakeyama, T., Interaction between water and hydrophilic polymers. *Thermochimica Acta* **1998**, *308* (1-2), 3-22.
219. McBrierty, V. J.; Martin, S. J.; Karasz, F. E., Understanding hydrated polymers: the perspective of NMR. **1999**, *80* (2-3), 179-205.
220. Matthews, J. F.; Skopec, C. E.; Mason, P. E.; Zuccato, P.; Torget, R. W.; Sugiyama, J.; Himmel, M. E.; Brady, J. W., Computer simulation studies of microcrystalline cellulose Iβ. **2006**, *341* (1), 138-152.
221. O'Neill, H.; Pingali, S. V.; Petridis, L.; He, J.; Mamontov, E.; Hong, L.; Urban, V.; Evans, B.; Langan, P.; Smith, J. C.; Davison, B. H., Dynamics of water bound to crystalline cellulose. *Scientific Reports* **2017**, *7* (1).

222. Yamane, C.; Aoyagi, T.; Ago, M.; Sato, K.; Okajima, K.; Takahashi, T., Two Different Surface Properties of Regenerated Cellulose due to Structural Anisotropy. **2006**, *38* (8), 819-826.
223. Okugawa, A.; Sakaino, M.; Yuguchi, Y.; Yamane, C., Relaxation phenomenon and swelling behavior of regenerated cellulose fibers affected by water. *Carbohydrate Polymers* **2020**, *231*, 115663.
224. Petridis, L.; O'Neill, H. M.; Johnsen, M.; Fan, B.; Schulz, R.; Mamontov, E.; Maranas, J.; Langan, P.; Smith, J. C., Hydration Control of the Mechanical and Dynamical Properties of Cellulose. *Biomacromolecules* **2014**, *15* (11), 4152-4159.
225. Matthew Wallace, K. L., Agne Kuraite and Yaroslav Z. Khimyak, Rapid determination of the acidity, alkalinity and carboxyl content of aqueous samples by 1H NMR with minimal sample quantity. *Anal. Chem.* **2020**.
226. Spaic, M.; Small, D. P.; Cook, J. R.; Wan, W., Characterization of anionic and cationic functionalized bacterial cellulose nanofibres for controlled release applications. *Cellulose* **2014**, *21* (3), 1529-1540.
227. Schmitt, J.; Calabrese, V.; da Silva, M. A.; Lindhoud, S.; Alfredsson, V.; Scott, J. L.; Edler, K. J., TEMPO-oxidised cellulose nanofibrils; probing the mechanisms of gelation via small angle X-ray scattering. *Phys Chem Chem Phys* **2018**, *20* (23), 16012-16020.
228. Bell, C. L.; Peppas, N. A., Water, solute and protein diffusion in physiologically responsive hydrogels of poly(methacrylic acid-g-ethylene glycol). **1996**, *17* (12), 1203-1218.
229. Madsen, F., Complexation graft copolymer networks: swelling properties, calcium binding and proteolytic enzyme inhibition. **1999**, *20* (18), 1701-1708.
230. Rorschach, H. E.; Hazlewood, C. F., Protein dynamics and the NMR relaxation time T1 of water in biological systems. **1986**, *70* (1), 79-88.
231. Zimmerman, J. R.; Brittin, W. E., Nuclear Magnetic Resonance Studies in Multiple Phase Systems: Lifetime of a Water Molecule in an Adsorbing Phase on Silica Gel. **1957**, *61* (10), 1328-1333.
232. Capitani, D.; Crescenzi, V.; Segre, A. L., Water in Hydrogels. An NMR Study of Water/Polymer Interactions in Weakly Cross-Linked Chitosan Networks. **2001**, *34* (12), 4136-4144.
233. Blinc, R.; Rutar, V.; Zupancic, I.; Zidansek, A.; Lahajnar, G.; Slak, J., Proton Nmr Relaxation of Adsorbed Water in Gelatin and Collagen. *Appl Magn Reson* **1995**, *9* (2), 193-216.
234. Stejskal, E. O.; Tanner, J. E., Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *The Journal of Chemical Physics* **1965**, *42* (1), 288-292.
235. Ali R. Rajabi-Siahboomi, R. W. B., Peter Mansfield, Martyn C. Davied, and Colin D. Melia, Structure and Behavior in Hydrophilic Matrix Sustained Release Dosage Forms: 4. Studies of Water Mobility and Diffusion Coefficients in the Gel Layer of HPMC Tablets Using NMR Imaging. *Pharmaceutical Research* **1996**, *13* (3).
236. Atsuko Yamada-Nosaka, K. I., Minoru Todoki and Hiroshi Tanzawa, 1H-NMR studies on water in methacrylate hydrogels. I. *J Appl Polym Sci* **1990**, *39*, 2443-2452.
237. Mahajan, S. S.; Paranj, R.; Mehta, R.; Lyon, R. P.; Atkins, W. M., A Glutathione-Based Hydrogel and Its Site-Selective Interactions with Water. **2005**, *16* (4), 1019-1026.
238. Segarra-Maset, M. D.; Escuder, B.; Miravet, J. F., Selective Interaction of Dopamine with the Self-Assembled Fibrillar Network of a Molecular Hydrogel Revealed by STD-NMR. **2015**, *21* (40), 13925-13929.
239. Wallace, M.; Iggo, J. A.; Adams, D. J., Probing the surface chemistry of self-assembled peptide hydrogels using solution-state NMR spectroscopy. **2017**, *13* (8), 1716-1727.
240. Ramalhete, S. M.; Nartowski, K. P.; Sarathchandra, N.; Foster, J. S.; Round, A. N.; Angulo, J.; Lloyd, G. O.; Khimyak, Y. Z., Supramolecular Amino Acid Based Hydrogels: Probing the Contribution of Additive Molecules using NMR Spectroscopy. *Chemistry - A European Journal* **2017**, *23* (33), 8014-8024.
241. Fatkullin, N. F.; Yatsenko, G. A.; Kimmich, R.; Fischer, E., Theory of spin diffusion in liquid-phase polymer systems. *Journal of Experimental and Theoretical Physics* **1998**, *87* (2), 294-302.

242. Komlosh, M. E.; Callaghan, P. T., Spin Diffusion in Semidilute Random Coil Polymers Studied by Pulsed Gradient Spin-Echo NMR. **2000**, *33* (18), 6824-6827.
243. Kalk, A.; Berendsen, H. J. C., Proton magnetic relaxation and spin diffusion in proteins. **1976**, *24* (3), 343-366.
244. Kubo, A.; McDowell, C. A., Spectral spin diffusion in polycrystalline solids under magic-angle spinning. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases* **1988**, *84* (11), 3713.
245. Meier, B. H., Polarization transfer and spin diffusion in solid-state NMR. *Adv. Magn. Opt. Reson.* **1994**, *18*, 1-116.
246. Fischer, E.; Kimmich, R.; Fatkullin, N.; Yatsenko, G., Segment diffusion and flip-flop spin diffusion in entangled polyethyleneoxide melts: A field-gradient NMR diffusometry study. **2000**, *62* (1), 775-782.
247. Fischer, E.; Kimmich, R.; Fatkullin, N., Spin diffusion in melts of entangled polymers. **1997**, *106* (23), 9883.
248. Emsley, L., Spin Diffusion for NMR Crystallography. In *Encyclopedia of Magnetic Resonance*, 2009.
249. Kumashiro, K. K.; Schmidt-Rohr, K.; Murphy, O. J.; Ouellette, K. L.; Cramer, W. A.; Thompson, L. K., A Novel Tool for Probing Membrane Protein Structure: Solid-State NMR with Proton Spin Diffusion and X-Nucleus Detection. **1998**, *120* (20), 5043-5051.
250. Huster, D.; Yao, X.; Hong, M., Membrane Protein Topology Probed by  $^1\text{H}$  Spin Diffusion from Lipids Using Solid-State NMR Spectroscopy. **2002**, *124* (5), 874-883.
251. Zhang, W.; Cory, D. G., First Direct Measurement of the Spin Diffusion Rate in a Homogenous Solid. **1998**, *80* (6), 1324-1327.
252. Clauss, J.; Schmidt-Rohr, K.; Spiess, H. W., Determination of Domain Sizes in Heterogeneous Polymers by Solid-State Nmr. *Acta Polym* **1993**, *44* (1), 1-17.
253. Wang, J., On the determination of domain sizes in polymers by spin diffusion. **1996**, *104* (12), 4850.
254. da Silva, M. A.; Calabrese, V.; Schmitt, J.; Celebi, D.; Scott, J. L.; Edler, K. J., Alcohol induced gelation of TEMPO-oxidized cellulose nanofibril dispersions. *Soft Matter* **2018**, *14* (45), 9243-9249.
255. Kimura, Y.; Haraguchi, K., Clay-Alcohol-Water Dispersions: Anomalous Viscosity Changes Due to Network Formation of Clay Nanosheets Induced by Alcohol Clustering. *Langmuir* **2017**, *33* (19), 4758-4768.
256. Duus, J. Ø.; Gotfredsen, C. H.; Bock, K., Carbohydrate Structural Determination by NMR Spectroscopy: Modern Methods and Limitations†. *Chemical Reviews* **2000**, *100* (12), 4589-4614.
257. Martineau, C.; Senker, J.; Taulelle, F., NMR Crystallography. 2014; pp 1-57.
258. Nonappa, N.; Kolehmainen, E., Solid state NMR studies of gels derived from low molecular mass gelators. *Soft Matter* **2016**, *12* (28), 6015-6026.
259. Nurisso, A.; Kozmon, S.; Imberty, A., Comparison of docking methods for carbohydrate binding in calcium-dependent lectins and prediction of the carbohydrate binding mode to sea cucumber lectin CEL-III. *Molecular Simulation* **2008**, *34* (4), 469-479.
260. Lan, Y.; Corradini, M. G.; Weiss, R. G.; Raghavan, S. R.; Rogers, M. A., To gel or not to gel: correlating molecular gelation with solvent parameters. *Chemical Society Reviews* **2015**, *44* (17), 6035-6058.
261. Kametani, S.; Sekine, S.; Ohkubo, T.; Hirano, T.; Ute, K.; Cheng, H. N.; Asakura, T., NMR studies of water dynamics during sol-to-gel transition of poly (N-isopropylacrylamide) in concentrated aqueous solution. *Polymer* **2017**, *109*, 287-296.
262. Segredo-Morales, E.; Martin-Pastor, M.; Salas, A.; Évora, C.; Concheiro, A.; Alvarez-Lorenzo, C.; Delgado, A., Mobility of Water and Polymer Species and Rheological Properties of Supramolecular

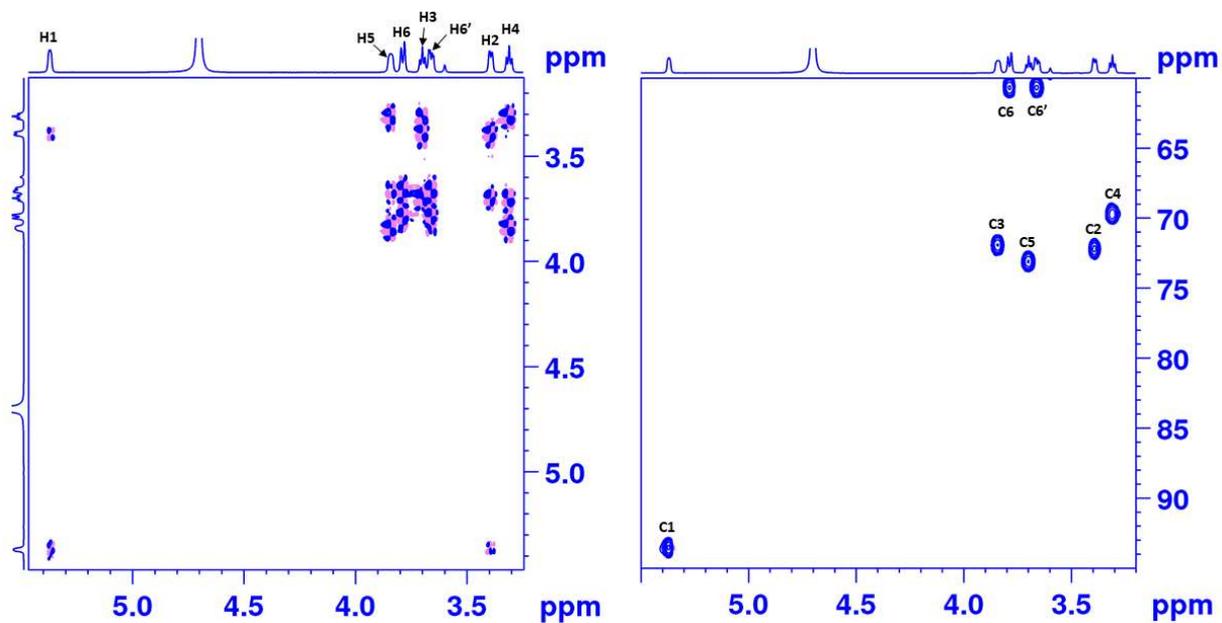
Polypseudorotaxane Gels Suitable for Bone Regeneration. *Bioconjugate Chemistry* **2018**, 29 (2), 503-516.

## Appendix

## Chapter 3

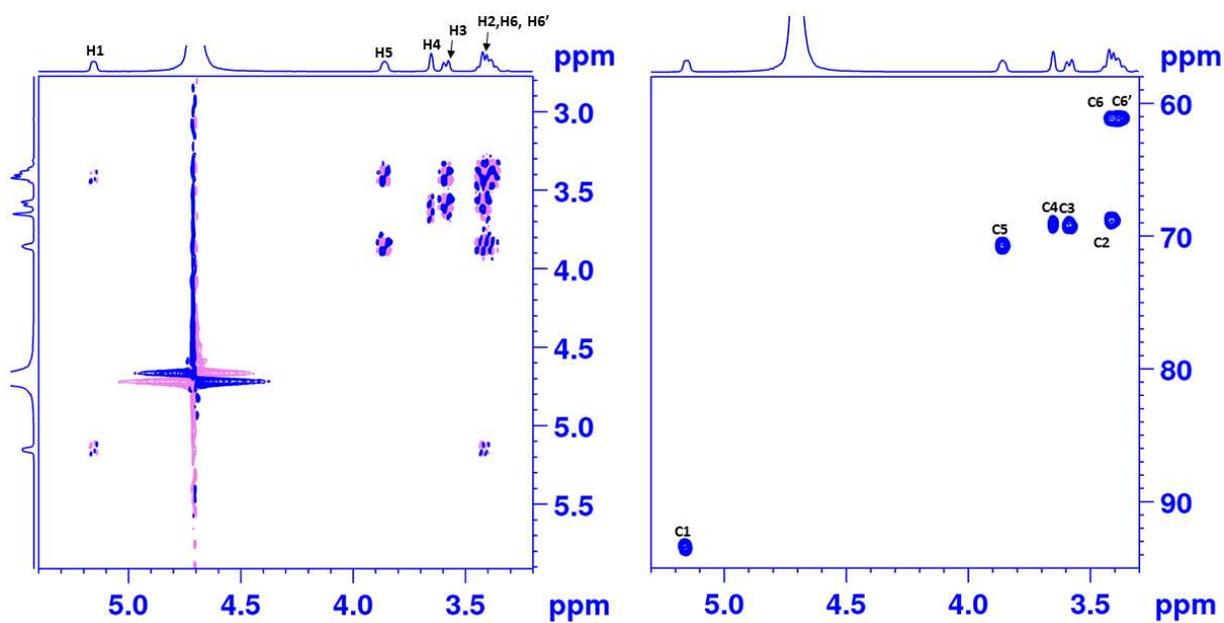
Molecular recognition of donor and donor-like substrates by CDP

### Glc-1-P

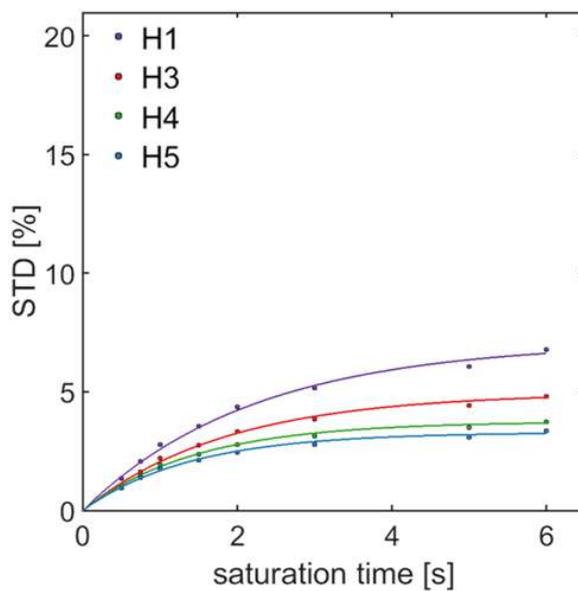


**Figure A.1:** (left) COSY and (right) HSQC spectra Glc-1-P at 278 K

## Gal-1-P



**Figure A.2:** (left) COSY and (right) HSQC spectra Gal-1-P at 278 K.

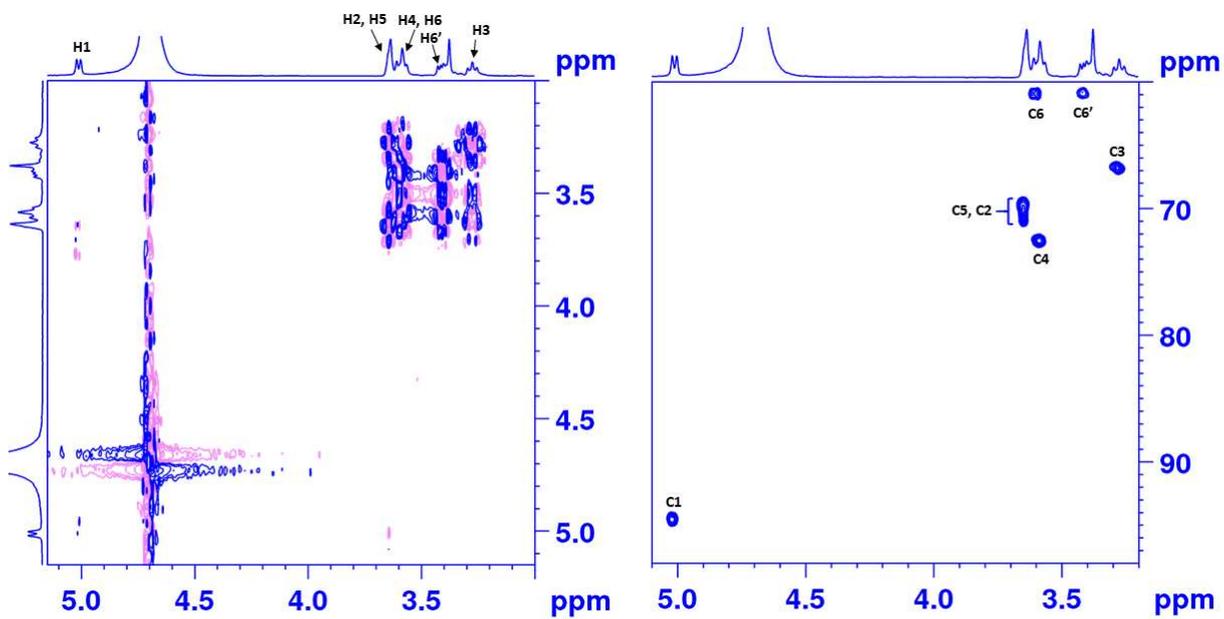


**Figure A.3:** Gal-1-P STD NMR build-up curves recorded at increasing saturation time at 278 K. 50  $\mu$ M binding unit was used for a ligand concentration of 5 mM.

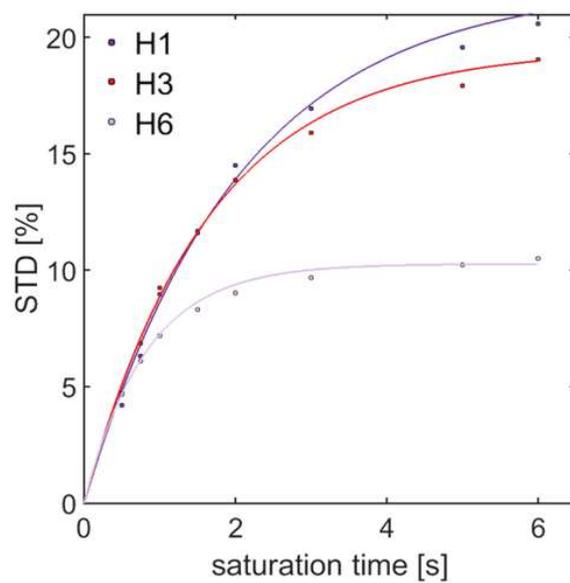
**Table A.1:**  $STD_{max}$ ,  $k_{sat}$  and  $STD_0$  for Gal-1-P. Relative STD (%) were obtained through normalisation against the largest ligand STD initial slope (H1; 100%)

	$STD_{max}$	$k_{sat}$	$STD_0$	STD (%)
<b>H1</b>	7.09	0.45	3.21	100
<b>H3</b>	4.96	0.54	2.65	83
<b>H4</b>	3.75	0.68	2.54	79
<b>H5</b>	3.29	0.71	2.35	73

## Man-1-P



**Figure A.4:** (left) COSY and (right) HSQC spectra Man-1-P at 278 K.

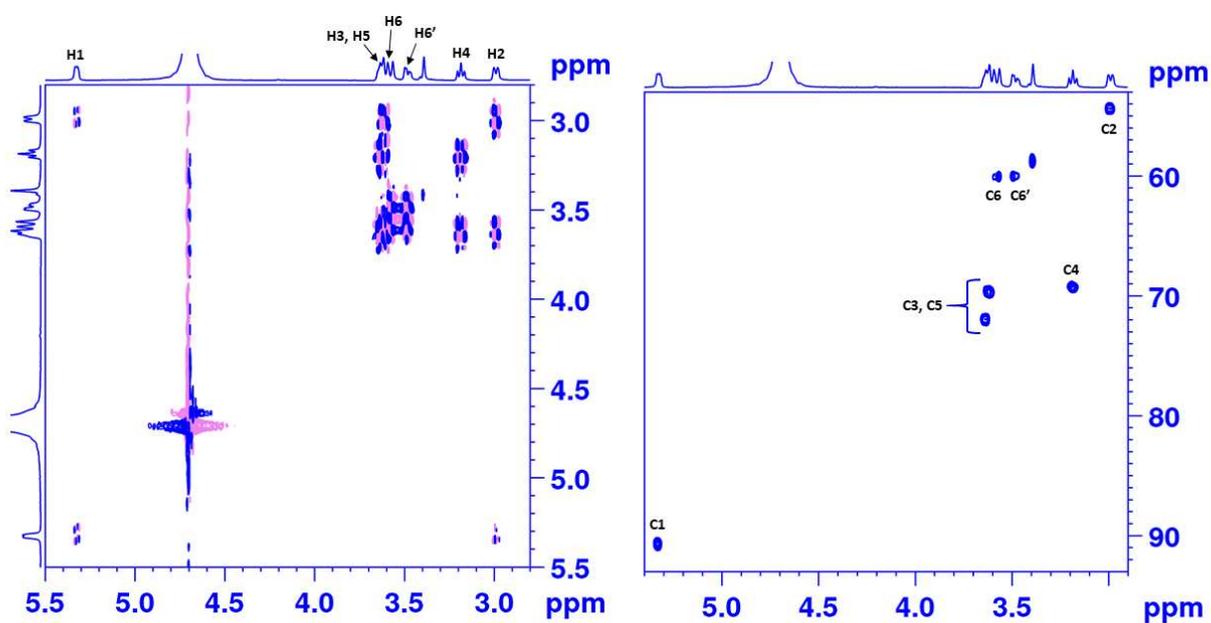


**Figure A.5:** Gal-1-P STD NMR build-up curves recorded at increasing saturation time at 278 K. 50  $\mu$ M binding unit was used for a ligand concentration of 5 mM.

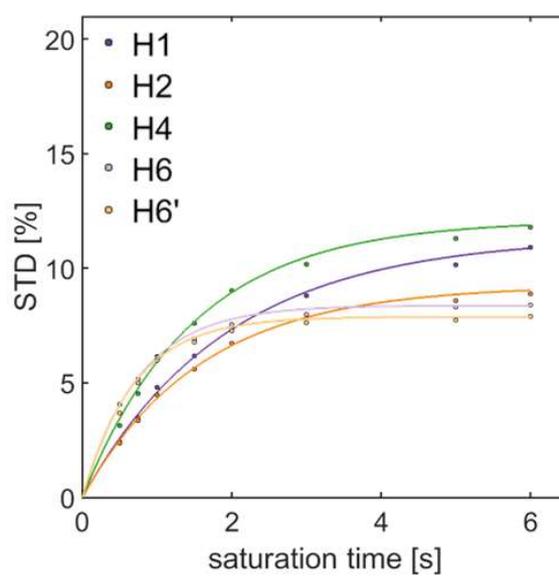
**Table A.2:**  $STD_{max}$ ,  $k_{sat}$  and  $STD_0$  for Man-1-P. Relative STD (%) were obtained through normalisation against the largest ligand STD initial slope (H6'; 100%)

	$STD_{max}$	$k_{sat}$	$STD_0$	STD (%)
<b>H1</b>	22.22	0.49	10.91	87
<b>H3</b>	19.49	0.61	11.81	94
<b>H6'</b>	10.27	1.22	12.55	100

## GIN-1-P



**Figure A.6:** (left) COSY and (right) HSQC spectra GIN-1-P at 278 K.

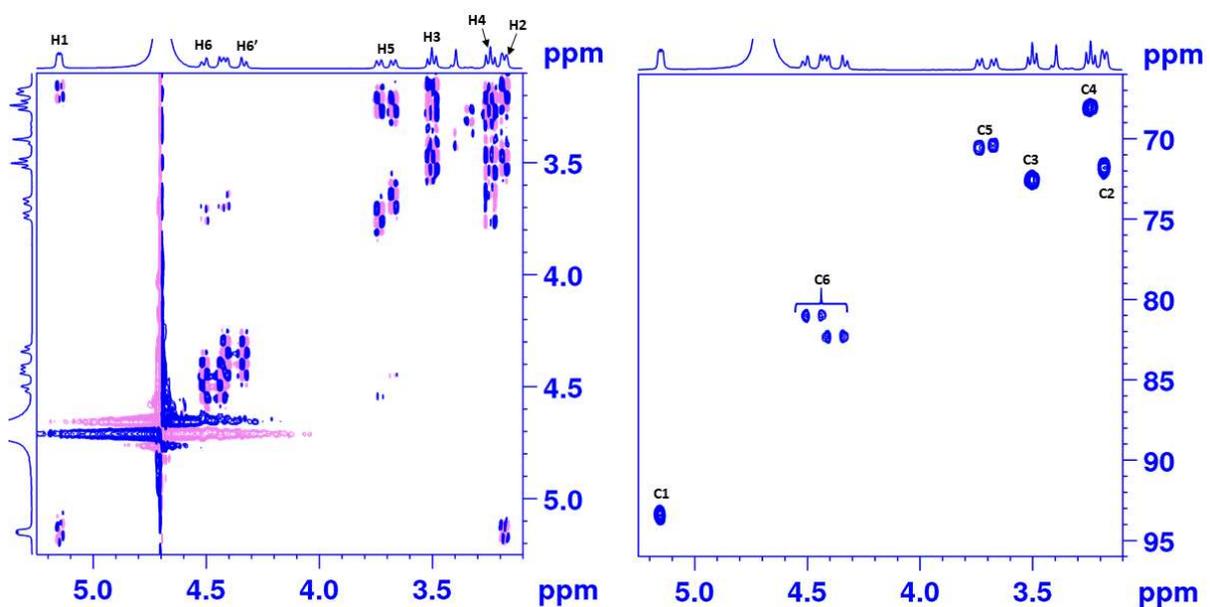


**Figure A.7:** GIN-1-P STD NMR build-up curves recorded at increasing saturation time at 278 K. 50  $\mu$ M binding unit was used for a ligand concentration of 5 mM.

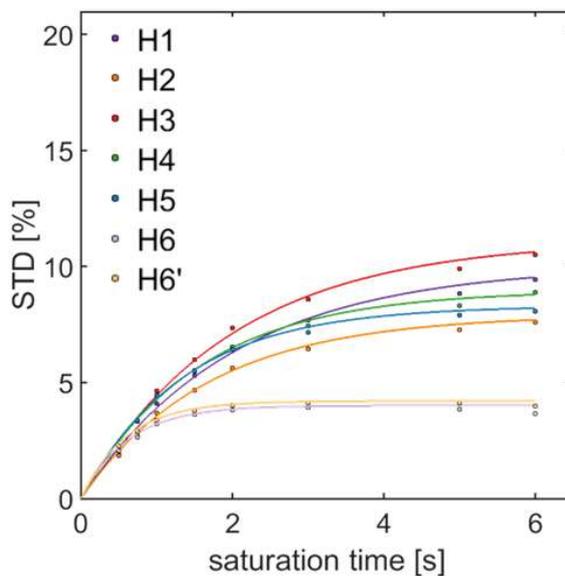
**Table A.3:**  $STD^{max}$ ,  $k_{sat}$  and  $STD_0$  for GlcN-1-P. Relative STD (%) were obtained through normalisation against the largest ligand STD initial slope (H6'; 100%)

	$STD^{max}$	$k_{sat}$	$STD_0$	STD (%)
H1	11.39	0.52	5.88	52
H2	9.25	0.63	5.83	51
H4	12.09	0.67	8.12	71
H6	8.38	1.32	11.02	97
H6'	7.88	1.44	11.36	100

### 6F-Glc-1-P



**Figure A.8:** (left) COSY and (right) HSQC spectra 6F-Glc-1-P at 278 K.



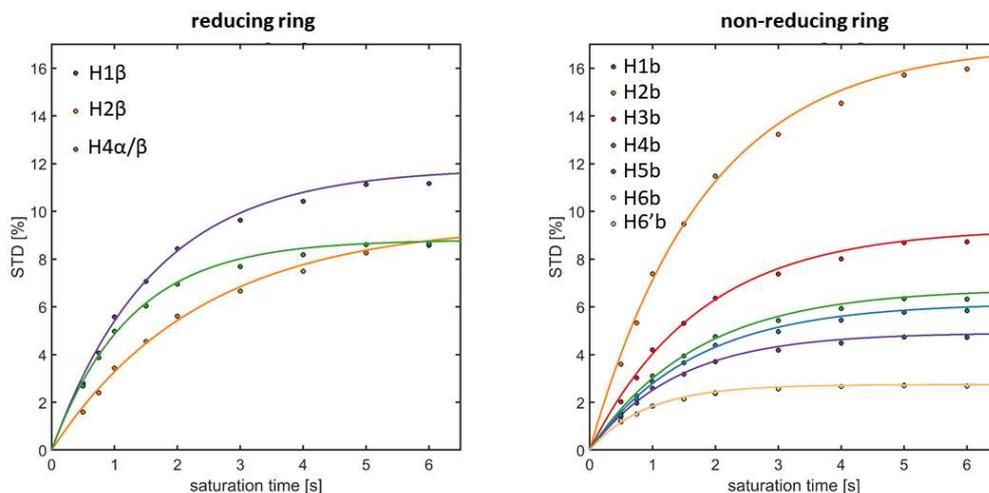
**Figure A.9:** 6F-Glc-1-P STD NMR build-up curves recorded at increasing saturation time at 278 K. 35  $\mu$ M binding unit was used for a ligand concentration of 3.5 mM.

**Table A.4:**  $STD^{max}$ ,  $k_{sat}$  and  $STD_0$  for 6F-Glc-1-P. Relative STD (%) were obtained through normalisation against the largest ligand STD initial slope (H6'; 100%)

	$STD^{max}$	$k_{sat}$	$STD_0$	STD (%)
<b>H1</b>	10.06	0.49	4.94	70
<b>H2</b>	7.92	0.60	4.72	67
<b>H3</b>	11.16	0.51	5.66	80
<b>H4</b>	8.97	0.64	5.78	82
<b>H5</b>	8.30	0.74	6.10	86
<b>H6</b>	4.01	1.60	6.42	91
<b>H6'</b>	4.20	1.69	7.07	100

## Molecular recognition of acceptor and acceptor-like substrates by CDP

### D-cellobiose

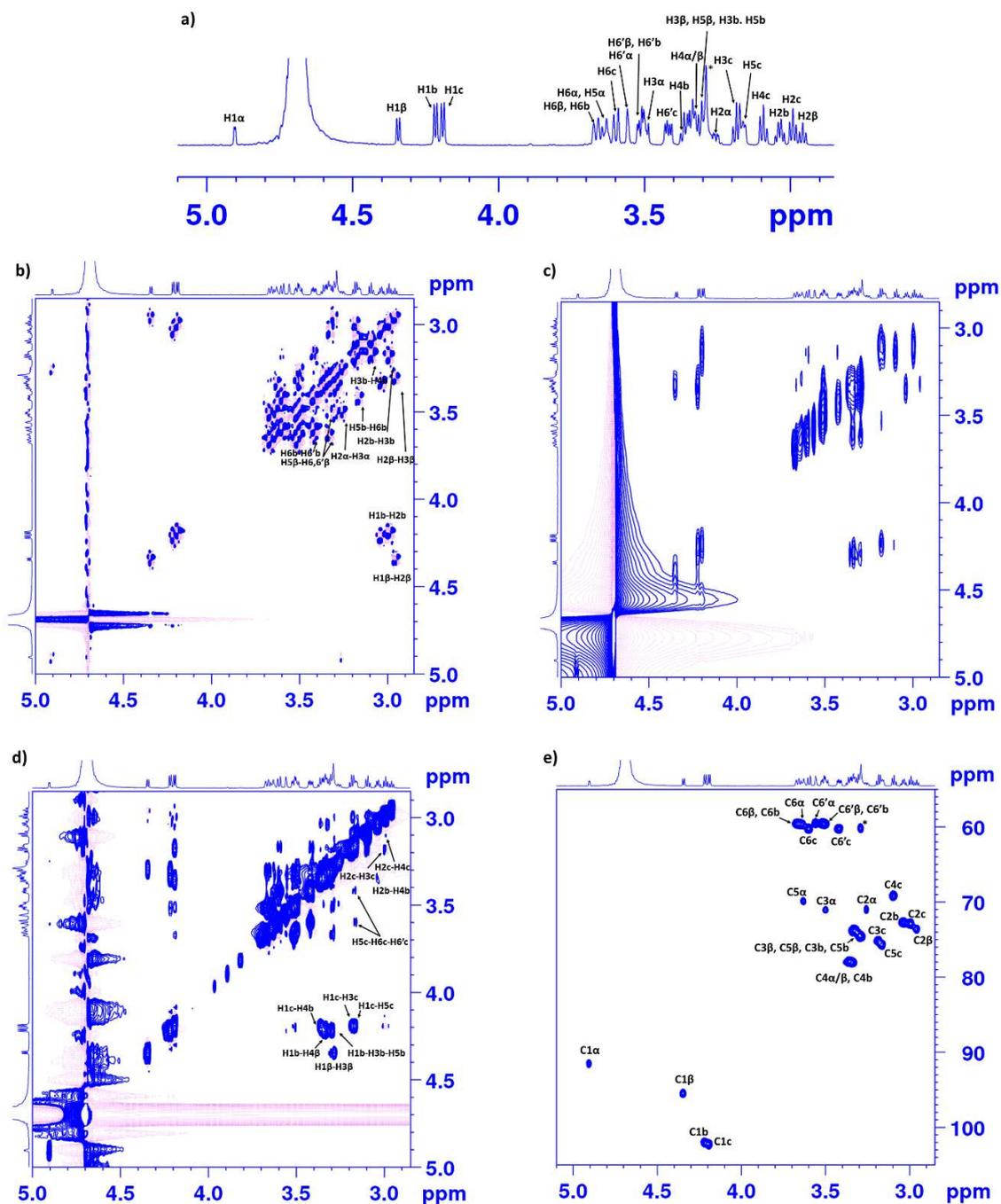


**Figure A.10:** D-cellobiose NMR build-up curves recorded at increasing saturation time (from 0.5 to 6 seconds) in  $[D_{11}]$ Tris buffer 25 mM, pH 7.4 at 278 K. 15  $\mu$ M binding unit was used for a ligand concentration of 3 mM.

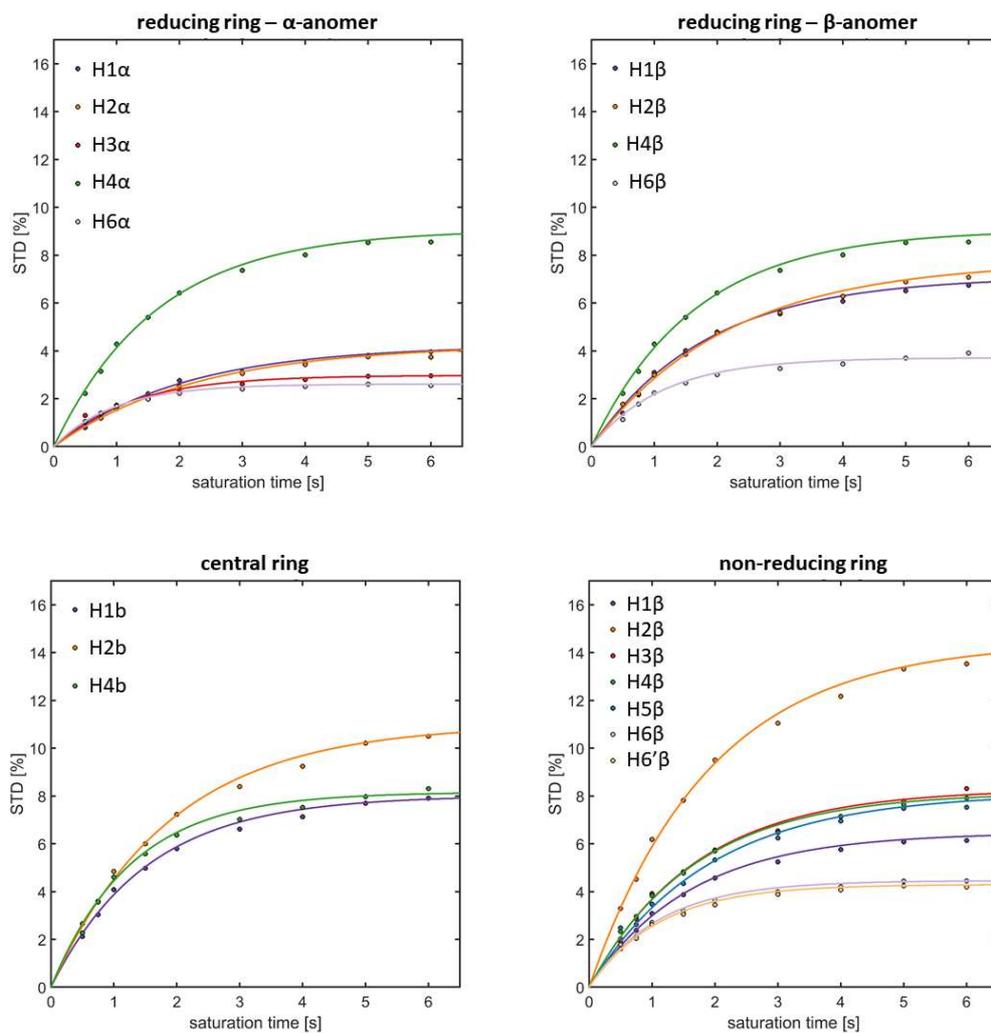
**Table A.5:**  $STD_{max}$ ,  $k_{sat}$  and  $STD_0$  for D-cellobiose in  $[D_{11}]$ Tris buffer 25 mM, pH 7.4 The assigned STD % were obtained through normalisation against the maximum ligand STD initial slope (H2 terminal non-reducing ring; 100%)

	$STD_{max}$	$k_{sat}$	$STD_0$	STD (%)
<b>H1<math>\beta</math></b>	11.82	0.61	7.24	78.52
<b>H2<math>\beta</math></b>	9.52	0.42	4.02	43.68
<b>H4<math>\alpha/\beta</math></b>	8.81	0.80	7.05	76.55
<b>H1b</b>	4.92	0.71	3.51	38.04
<b>H2b</b>	17.03	0.54	9.21	100.00
<b>H3b</b>	9.30	0.57	5.28	57.27
<b>H4b</b>	6.76	0.59	3.97	43.05
<b>H5b</b>	6.16	0.60	3.71	40.24
<b>H6b</b>	2.74	1.10	3.01	32.65
<b>H6'b</b>	2.74	1.11	3.05	33.05

D-celotriose



**Figure A.11:** a) COSY, b) TOCSY, c) NOESY (mixing time 300 ms) and d) HSQC spectra recorded for D-celotriose (1.5 mM, in [D<sub>11</sub>]Tris 25 mM) at 278 K.

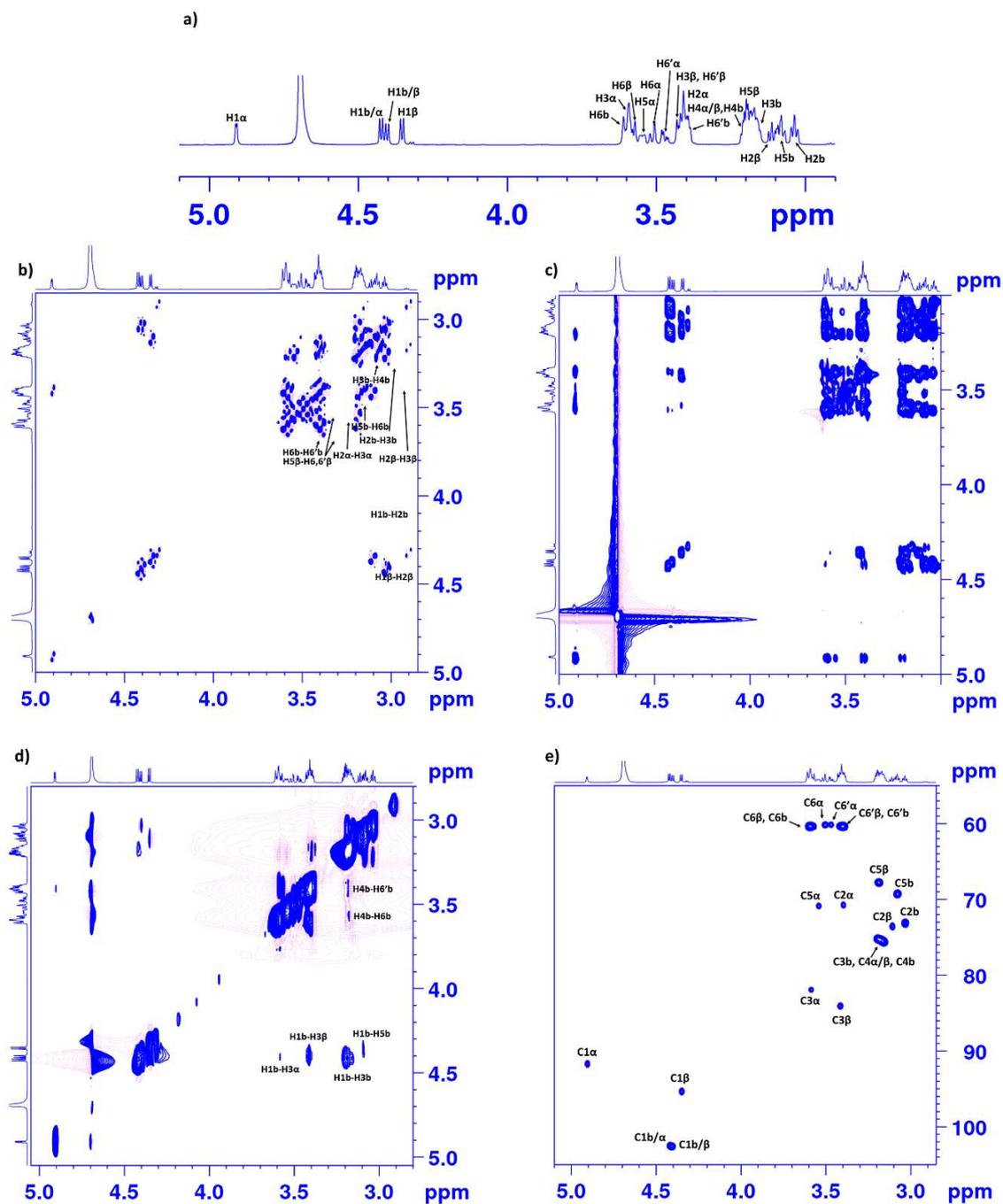


**Figure A.12:** D-celotriose NMR build-up curves recorded at increasing saturation time (from 0.5 to 6 seconds) in  $[D_{11}]$ Tris buffer 25 mM, pH 7.4 at 278 K. 15  $\mu$ M binding unit was used for a ligand concentration of 3 mM.

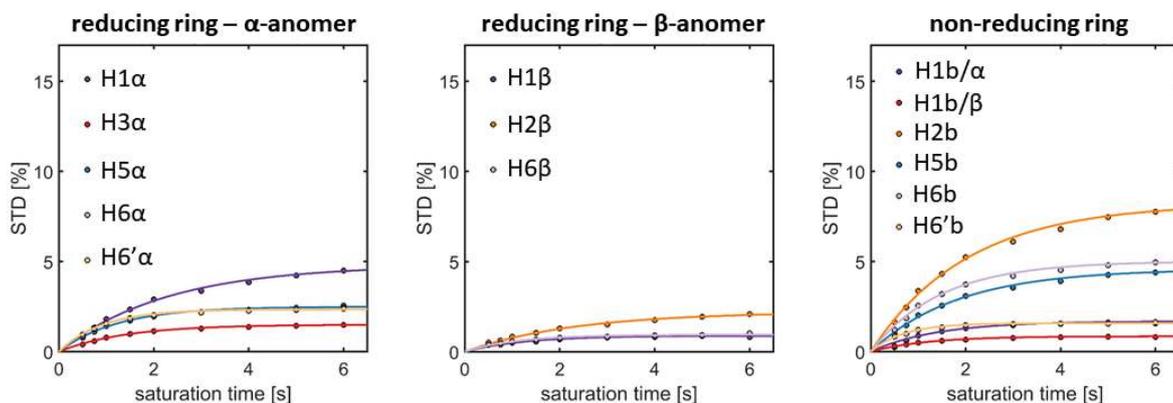
**Table A.6:**  $STD_{max}$ ,  $k_{sat}$  and  $STD_0$  for D-celotriose in [D<sub>11</sub>]Tris buffer 25 mM, pH 7.4 The assigned STD % were obtained through normalisation against the maximum ligand STD initial slope (H2 terminal non-reducing ring; 100%)

	$STD_{max}$	$k_{sat}$	$STD_0$	STD (%)
<b>H1<math>\alpha</math></b>	4.24	0.49	2.06	27.25
<b>H2<math>\alpha</math></b>	4.26	0.44	1.88	24.93
<b>H3<math>\alpha</math></b>	2.98	0.82	2.44	32.33
<b>H4<math>\alpha</math></b>	9.06	0.61	5.52	73.09
<b>H6<math>\alpha</math></b>	2.61	1.04	2.71	35.81
<b>H1<math>\beta</math></b>	7.11	0.54	3.86	51.13
<b>H2<math>\beta</math></b>	7.75	0.46	3.55	47.03
<b>H4<math>\beta</math></b>	9.06	0.61	5.52	73.09
<b>H6<math>\beta</math></b>	3.71	0.89	3.30	43.66
<b>H1b</b>	8.02	0.66	5.29	70.01
<b>H2b</b>	11.01	0.53	5.84	77.25
<b>H4b</b>	8.16	0.79	6.43	85.14
<b>H1c</b>	6.47	0.62	4.00	52.99
<b>H2c</b>	14.46	0.52	7.56	100.00
<b>H3c</b>	8.29	0.59	4.87	64.44
<b>H4c</b>	8.15	0.60	4.87	64.42
<b>H5c</b>	8.13	0.53	4.29	56.71
<b>H6c</b>	4.46	0.90	4.03	53.30
<b>H6'c</b>	4.30	0.90	3.87	51.21

## D-laminaribiose



**Figure A.13:** a) COSY, b) TOCSY, c) NOESY (mixing time 300 ms) and d) HSQC spectra recorded for D-laminaribiose (3 mM, in Tris-d<sub>11</sub> 25 Mm) at 278 K.

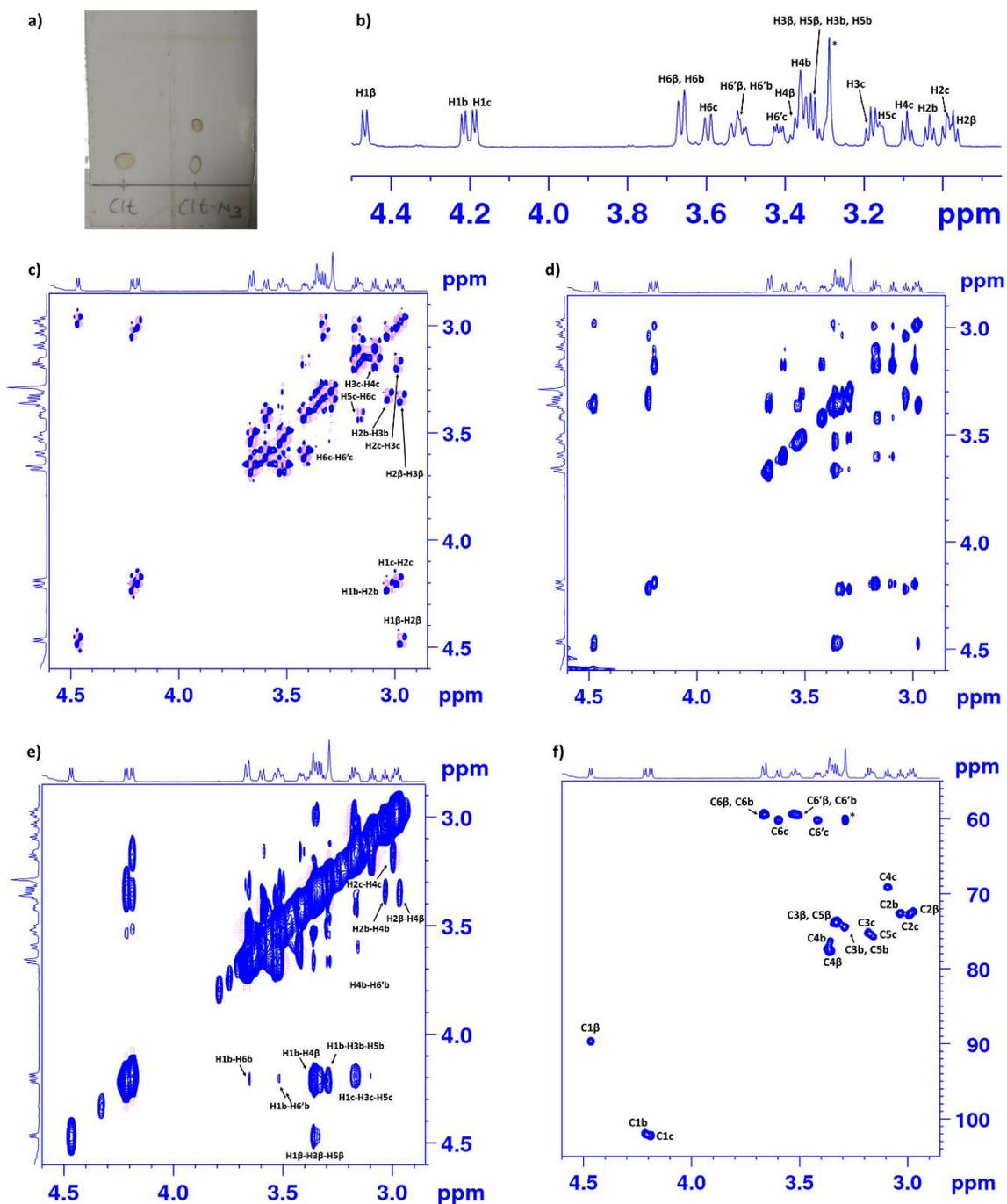


**Figure A.14:** D-laminaribiose NMR build-up curves recorded at increasing saturation time (from 0.5 to 6 seconds) in  $[D_{11}]$ Tris buffer 25 mM, pH 7.4 at 278 K. 10  $\mu$ M binding unit was used for a ligand concentration of 2 mM.

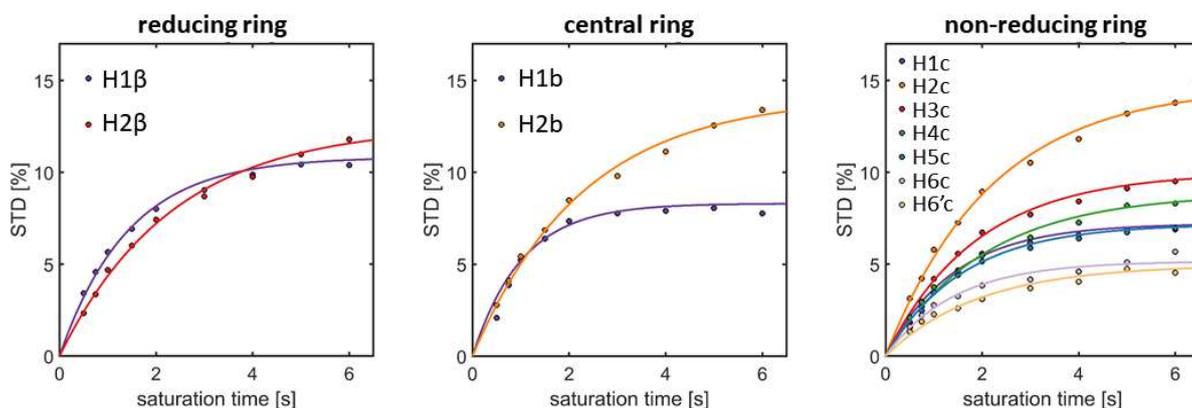
**Table A.7:**  $STD_{max}$ ,  $k_{sat}$  and  $STD_0$  for D-cellobiose in  $[D_{11}]$ Tris buffer 25 mM, pH 7.4 The assigned STD % were obtained through normalisation against the maximum ligand STD initial slope (H2 terminal non-reducing ring; 100%)

	$STD_{max}$	$k_{sat}$	$STD_0$	STD (%)
H1 $\alpha$	5.41	0.44	2.38	58.87
H3 $\alpha$	1.69	0.71	1.20	29.58
H5 $\alpha$	2.81	0.80	2.25	55.66
H6 $\alpha$	2.65	1.08	2.87	70.85
H6' $\alpha$	2.64	1.10	2.90	71.49
H1 $\beta$	0.81	0.79	0.64	15.89
H2 $\beta$	2.03	0.43	0.87	21.38
H6 $\beta$	0.87	0.90	0.79	19.41
H1b/ $\alpha$	1.86	0.72	1.34	33.03
H1b/ $\beta$	0.78	0.85	0.66	16.34
H2b	8.23	0.49	4.05	100.00
H5b	4.59	0.54	2.50	61.67
H6b	5.02	0.68	3.40	83.98
H6'b	1.58	1.44	2.28	56.28

## D-cellootriosyl-azide



**Figure A.15:** a) TLC showing partial conversion of D-cellootriosyl-azide into D-cellootriosyl-azide. b) 1D <sup>1</sup>H NMR spectra c) COSY, d) TOCSY, e) NOESY (mixing time 300 ms) and f) <sup>13</sup>C-<sup>1</sup>H HSQC spectra with peaks assignment reported.



**Figure A.16:** D-celotriosyl-azide NMR build-up curves recorded at increasing saturation time (from 0.5 to 6 seconds) in  $[D_{11}]$ Tris buffer 25 mM, pH 7.4 at 278 K. 10  $\mu$ M binding unit was used for a ligand concentration of 2 Mm.

**Table A.8:**  $STD_{max}$ ,  $k_{sat}$  and  $STD_0$  for D-celotriosyl-azide in  $[D_{11}]$ Tris buffer 25 mM, pH 7.4. The assigned STD % were obtained through normalisation against the maximum ligand STD initial slope (H1 central ring; 100%)

	$STD_{max}$	$k_{sat}$	$STD_0$	STD (%)
H1β	10.83	0.70	7.60	93.37
H2β	12.54	0.43	5.37	65.97
H1b	8.30	0.98	8.14	100.00
H2b	14.13	0.44	6.21	76.25
H1c	7.22	0.70	5.08	62.37
H2c	14.73	0.45	6.70	82.28
H3c	10.00	0.53	5.29	65.04
H4c	8.86	0.49	4.32	53.10
H5c	7.18	0.63	4.50	55.26
H6c	5.16	0.71	3.64	44.72
H6'c	4.97	0.52	2.58	31.64

## Orientation of d-cellobiose in the binding pocket – DEEP STD experiments

**Table A.9:** Amino acid residues of CDP chain A within 5 Å from the ligand. <sup>1</sup>H chemical shifts have been simulated by shiftx2.ca (ref. <http://shiftx2.ca>), and resonances within 0.6 ppm from the directly irradiated frequencies (1.5, 6.7 and 7.3 ppm) are highlighted in the same colour scale as in Figure 3.10.

<b>297</b>	<b>ASP</b>	H	8.12
		HA	4.63
		HB2	2.71
		HB3	2.45
<b>300</b>	<b>TYR</b>	<b>H</b>	<b>7.85</b>
		HA	4.08
		HB2	3.16
		HB3	2.96
		<b>HD1</b>	<b>7.09</b>
		<b>HD2</b>	<b>7.01</b>
		<b>HE1</b>	<b>6.72</b>
		<b>HE2</b>	<b>7.23</b>

**Table A.10:** Amino acid residues of CDP chain B within 5 Å from the ligand. <sup>1</sup>H chemical shifts have been simulated by shiftx2.ca (ref. <http://shiftx2.ca>), and resonances within 0.6 ppm from the directly irradiated frequencies (1.5, 6.7 and 7.3 ppm) are highlighted in the same colour scale as in Figure 3.10.

486	ARG	H	8.44	502	GLU	H	7.82	874	GLN	HA	4.63	HD2	6.93				
		HA	4.30			HA	4.06			HB2	2.79	HE1	7.77				
		HB2	1.79			HB2	2.27			HB3	2.63	H	7.98				
		HB3	1.69			HB3	2.13			HD1	7.04	HA	4.18				
		HD2	3.15			HG2	2.39			HD2	6.92	HB2	2.12				
		HD3	3.19			HG3	2.31			HE1	6.78	HB3	1.85				
		HE	7.29			622	TRP			H	8.89	HE2	7.18	HE21	7.35		
		HG2	1.59							HA	4.51	809	ARG	H	6.27	HE22	7.23
		HG3	1.68							HB2	3.20			HA	2.74	HG2	2.43
492	GLN	H	7.55	623	ASN	HB3	3.11	HB2	1.10	886	PRO	HA	4.324				
		HA	4.30			HD1	7.30	HB3	1.29			HB2	2.07				
		HB2	2.04			HE1	10.13	HE	7.26			HB3	1.98				
		HB3	1.97			HE3	7.47	HD2	2.68			HD2	3.65				
		HE21	7.18			H	9.14	HD3	2.74			HD3	3.55				
		HE22	7.01			HA	4.55	HE	7.26			HG2	1.94				
		HG2	2.25			HB2	2.86	HG2	1.20			HG3	0.93				
HG3	2.26	HB3	2.84	HG3	0.93	H	8.60	HG2	1.86								
496	ARG	H	7.76	624	ASP	HD21	7.66	810	GLU	HA	4.19	889	SER	H	8.07		
		HA	4.20			HD22	7.07			HB2	2.10			HA	4.69		
		HB2	1.68			H	8.45			HB3	2.09			HB2	3.90		
		HB3	1.65			HA	4.44			HG2	2.52			HB3	3.74		
		HD2	3.15			HB2	2.58			HG3	2.43			H	8.85		
		HD3	3.22			HB3	2.59			815	PHE			H	8.50	HA	3.72
		HE	7.21			625	CYS							HA	4.37	HA2	3.92
		HG2	1.53											H	8.83	HB2	2.85
		HG3	1.50							HA	4.36			HB3	2.87		
501	ARG	H	8.23	786	LEU	HB2	2.85	HD1	7.12	817	HIS	H	8.42				
		HA	3.94			HB3	2.88	HD2	6.86			HA	4.29				
		HB2	1.78			H	8.05	HE1	6.98			HB2	3.31				
		HB3	1.39			HA	4.08	HE2	7.15			HB3	3.02				
		HD2	3.07			HB2	1.64	HZ	6.88								
		HD3	3.32			HB3	1.50	H	8.42								
		HE	8.00			HD1	0.61	HD2	0.38								
		HG2	1.70			HD2	0.38	HG	1.38								
		HG3	1.82			804	TYR	H	7.09								

**Table A.11:**  $\Delta$ -STD factors calculated for 1.5/6.7 Differential Irradiation of the CDP/D-cellobiose complex.

Protons	$^1\text{H } \delta$ (ppm) <sup>[a]</sup>	STD % 1.5 ppm	STD % 6.7 ppm	Ratio STD (1.5/6.7 ppm)	$\Delta$ STD
H1 $\beta$	4.36	3.46	0.36	0.75	0.05
H2 $\beta$	2.98	2.05	2.71	0.78	0.08
H4 $\beta$	3.36	2.71	1.43	0.70	0.00
H1b	4.21	1.58	2.02	0.75	0.05
H2b	3	4.28	1.20	0.76	0.06
H3b	3.2	2.60	3.25	0.76	0.06
H4b	3.11	1.98	1.74	0.67	-0.03
H5b	3.17	1.92	1.26	0.64	-0.06
H6b	3.61	1.45	1.18	0.61	-0.08
H6'b	3.43	1.51	0.93	0.64	-0.06
		<b>SUM</b>	<b>SUM</b>	<b>STD average</b>	
		24.02	17.04	0.70	

<sup>[a]</sup> Spectra for assignment acquired at 278 K.

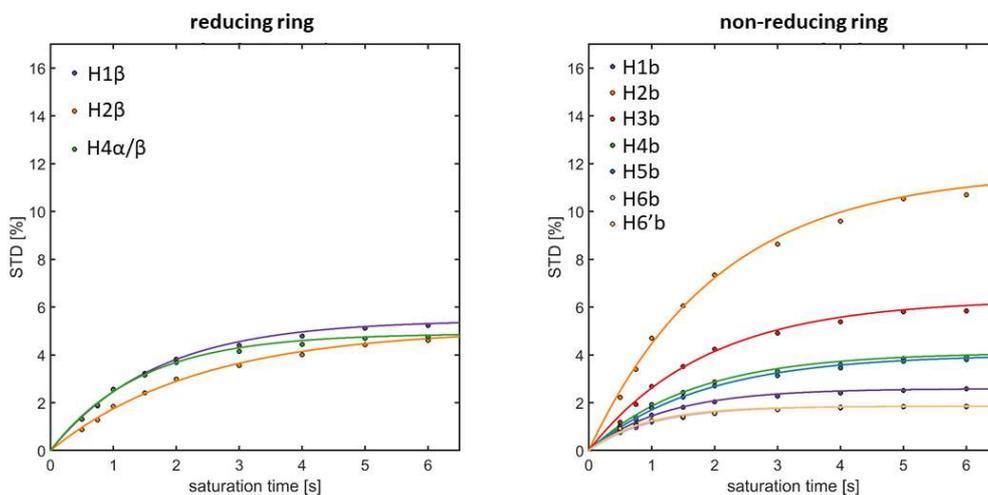
**Table A.12:**  $\Delta$ -STD factors calculated for 1.5/7.3 Differential Irradiation of the CDP/D-cellobiose complex.

Protons	$^1\text{H } \delta$ (ppm) <sup>[a]</sup>	STD % 1.5 ppm	STD % 7.3 ppm	Ratio STD (1.5/7.3 ppm)	$\Delta$ STD
H1 $\beta$	4.36	3.46	0.33	0.68	0.04
H2 $\beta$	2.98	2.05	2.51	0.72	0.08
H4 $\beta$	3.36	2.71	1.34	0.66	0.02
H1b	4.21	1.58	1.85	0.68	0.04
H2b	3	4.28	1.04	0.66	0.02
H3b	3.2	2.60	3.16	0.74	0.10
H4b	3.11	1.98	1.54	0.59	-0.05
H5b	3.17	1.92	1.16	0.59	-0.05
H6b	3.61	1.45	1.06	0.55	-0.09
H6'b	3.43	1.51	0.84	0.58	-0.06
		<b>SUM</b>	<b>SUM</b>	<b>STD average</b>	
		24.02	15.72	0.64	

<sup>[a]</sup> Spectra for assignment acquired at 278 K.

## Impact of inorganic phosphate in acceptor binding

### CDP/D-cellobiose complex with small excess of inorganic phosphate

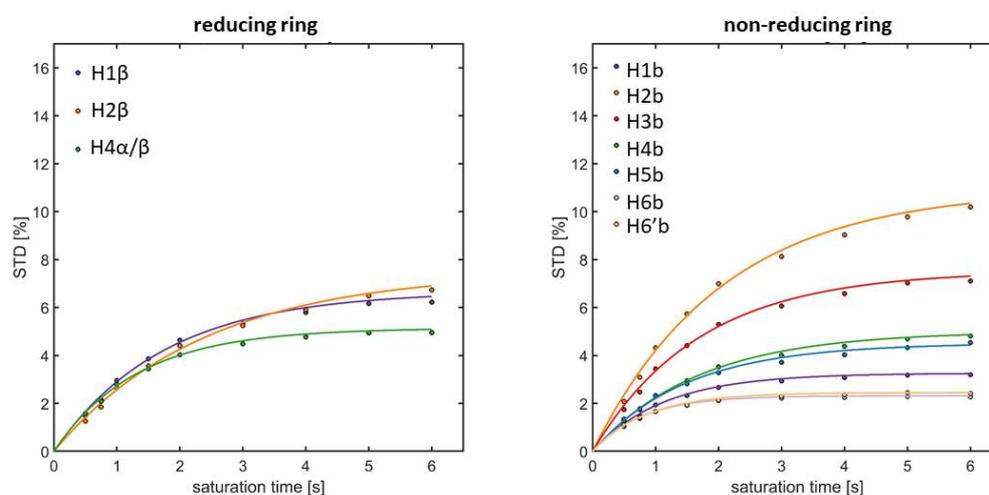


**Figure A.17:** D-cellobiose NMR build-up curves recorded at increasing saturation time (from 0.5 to 6 seconds) in  $[D_{11}]$ Tris buffer 25 mM, pH 7.4 at 278 K. 15  $\mu$ M binding unit was used for a ligand concentration of 3 mM and an inorganic phosphate concentration of 100  $\mu$ M.

**Table A.13:**  $STD_{max}$ ,  $k_{sat}$  and  $STD_0$  for D-cellobiose in  $[D_{11}]$ Tris buffer 25 mM, pH 7.4 with K<sub>3</sub>PO<sub>4</sub> 100  $\mu$ M. The assigned STD % were obtained through normalisation against the maximum ligand STD initial slope (H2 terminal non-reducing ring; 100%)

	$STD_{max}$	$k_{sat}$	$STD_0$	STD (%)
<b>H1 <math>\beta</math></b>	5.46	0.60	3.27	58.09
<b>H2 <math>\beta</math></b>	5.09	0.42	2.14	37.98
<b>H4 <math>\alpha/\beta</math></b>	4.90	0.70	3.43	60.85
<b>H1 b</b>	2.59	0.82	2.11	37.44
<b>H2 b</b>	11.65	0.48	5.63	100.00
<b>H3 b</b>	6.32	0.53	3.37	59.89
<b>H4 b</b>	4.10	0.60	2.47	43.77
<b>H5 b</b>	4.02	0.55	2.21	39.16
<b>H6 b</b>	1.85	1.00	1.86	33.05
<b>H6' b</b>	1.85	1.10	2.04	36.19

### CDP/D-cellobiose complex with large excess of inorganic phosphate



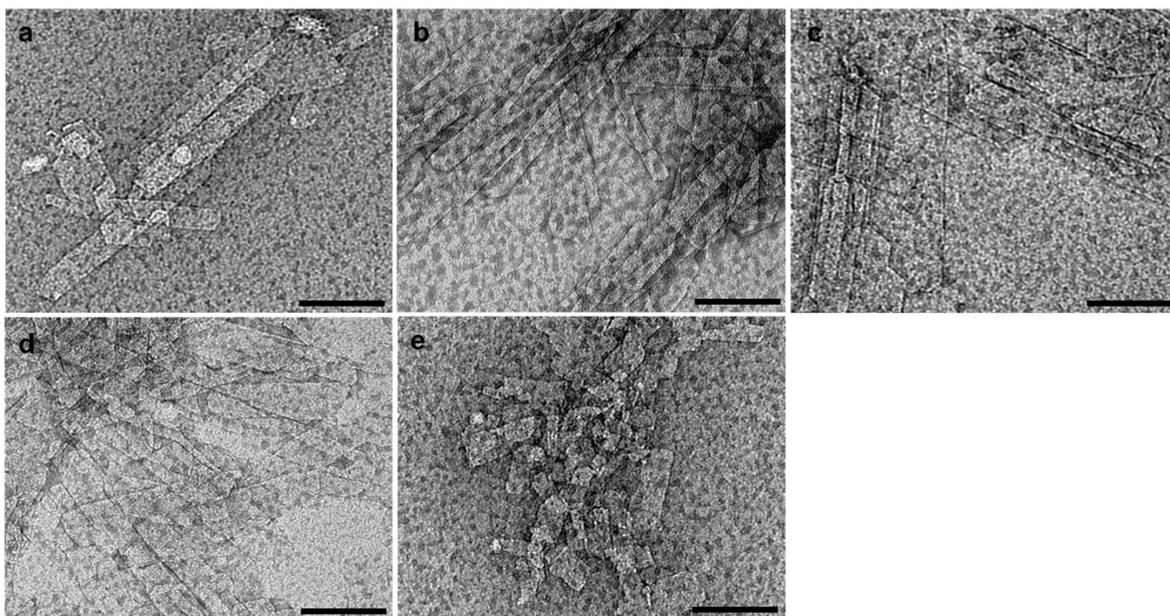
**Figure A.18:** D-cellobiose NMR build-up curves recorded at increasing saturation time (from 0.5 to 6 seconds) in PBS 25 mM, pH 7.4 at 278 K. 15  $\mu$ M binding unit was used for a ligand concentration of 3 mM.

**Table A.14:**  $STD_{max}$ ,  $k_{sat}$  and  $STD_0$  for D-cellobiose in PBS 25 mM, pH 7.4. The assigned STD % were obtained through normalisation against the maximum ligand STD initial slope (H2 terminal non-reducing ring; 100%)

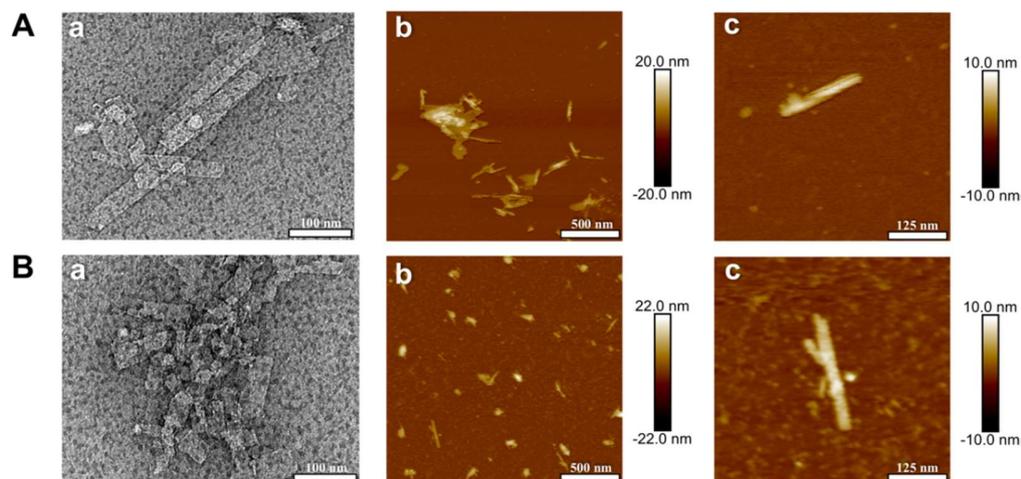
	$STD_{max}$	$k_{sat}$	$STD_0$	STD (%)
H1 $\beta$	6.67	0.57	3.81	71.80
H2 $\beta$	7.51	0.42	3.14	59.18
H4 $\alpha/\beta$	5.14	0.76	3.90	73.33
H1 b	3.26	0.88	2.87	53.95
H2 b	10.94	0.49	5.31	100.00
H3 b	7.52	0.59	4.43	83.47
H4 b	4.85	0.59	2.86	53.76
H5 b	4.51	0.67	3.03	57.00
H6 b	2.31	1.28	2.97	55.95
H6' b	2.45	1.16	2.83	53.37

## Chapter 4

### Morphological characterisation of deoxy-fluorinated cellodextrin derivatives – TEM and AFM

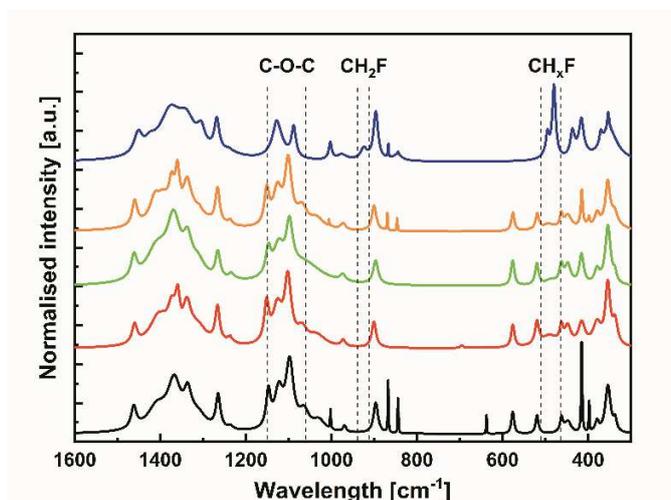


**Figure A.19:** TEM images of enzymatically produced cellodextrin EpC (a) and enzymatically produced fluorinated cellodextrins 2F-EpC (b), 3F-EpC (c), 6F-EpC (d) multi-6F-EpC (e) negatively stained with 2% uranyl acetate. Scale bars correspond to 100 nm.



**Figure A.20:** TEM (a) and AFM (b and c) images of EpC (5, row A) and multi-6F-EpC (4, row B). Scale bars are shown at the bottom of each image. The gradient bars next to b and c correspond to height measurements.

### Short range characterisation of deoxy-fluorinated cellodextrin derivatives – Raman Spectroscopy



**Figure A.21:** Deconvoluted and normalised Raman spectra for EpC (black), 2F-EpC (red), 3F-EpC (green), 6F-EpC (orange) and multi-6F-EpC (blue). Dashed lines correspond to boundaries of bands associated with C-O-C stretching (C-O-C) and presence of fluorinated carbon groups (CH<sub>2</sub>F and CH<sub>x</sub>F). Each Raman spectrum represents the average of three Lorentzian-deconvoluted spectra upon noise removal.

## Chapter 5

### Fit of the SDTD curves to the 1D equation

**Script A.1:** Matlab script used in this work to plot the SDTD data vs the square root of saturation time (SQRT\_tsat) and carry of the fit to Equation 5.15.

```
function [fitresult, gof] = createFit(SQRT_tsat, SDTD)
%CREATEFIT(SQRT_TSAT,SDTD)
% Create a fit.
%
% Data for 'SDTD_fit' fit:
%   X Input : SQRT_tsat
%   Y Output: SDTD
% Output:
%   fitresult : a fit object representing the fit.
%   gof : structure with goodness-of fit info.
%
% See also FIT, CFIT, SFIT.

% export data, introduce path to your xlsx file in fname.
% SQRT_tsat and SDTD data must be in the first and second column, respectively, of fname
fname = 'PATH/filename.xlsx';
data = readmatrix(fname);
SQRT_tsat = data(:,1);
SDTD = data(:,2);
scatter(SQRT_tsat,SDTD);

%% Fit: 'untitled fit 1'.
[xData, yData] = prepareCurveData( SQRT_tsat, SDTD );

% Set up fitype and options.
```

```

ft = fitype( 'C*erfc(r/(sqrt(4*D*x))-b)', 'independent', 'x', 'dependent', 'y' );
opts = fitoptions( 'Method', 'NonlinearLeastSquares' );
opts.Display = 'Off';
opts.Lower = [-Inf 0 1 0.2];
opts.StartPoint = [1 0.000125 1 0.2];
opts.Upper = [Inf 0.1 1 0.2];

% Fit model to data.
[fitresult, gof] = fit( xData, yData, ft, opts );

% Plot fit with data.
figure( 'Name', 'SDTD_fit' );
h = plot( fitresult, xData, yData );
legend( h, 'SDTD vs. SQRT_tsat', 'SDTD_fit', 'Location', 'NorthEast', 'Interpreter', 'none' );
% Label axes
xlabel( 'SQRT_tsat', 'Interpreter', 'none' );
ylabel( 'SDTD', 'Interpreter', 'none' );
grid on

```