

**Studying the intra and intermolecular
processes of biomolecules using two
dimensional infrared spectroscopy**



**University of
Nottingham**

UK | CHINA | MALAYSIA

David Tiemessen

School of Chemistry

University of Nottingham

PhD Supervisor: Professor Michael George

Doctor of Philosophy

Student ID: 4281946

June 2021

Declaration

I hereby declare that this Thesis and the work presented in it are my own, unless acknowledged otherwise.

D. Tiemessen

David Tiemessen

June 2021

Acknowledgements

I would firstly like to thank my supervisor, Mike George for his excellent support and guidance during my PhD. I would also like to thank my industrial supervisor, Paul Pudney for hosting me at Unilever during visits as well as the advice and support he also provided.

This PhD would not have been possible without Xue-Sun Zhong and a special thanks must go to him, who has taught me everything I know about lasers and ultrafast spectroscopy. I would also like to thank past and present members of the George Group for their help during my PhD. In particular, Ashley, Calum, Charlotte, Katie, Surajit and Raph have all made this process more enjoyable and have given invaluable advice.

Some of the experiments were performed at the Central Laser Facility. I would like to thank Greg Geetham and Igor Sazanovich for their assistance in setting up and performing those experiments. I would also like to thank Elena Owen for helpful discussions and providing mucin samples.

Finally, I would like to thank my friends and family for their support during my PhD.

Abstract

This Thesis focuses on using ultrafast time-resolved infrared spectroscopy and two-dimensional infrared spectroscopy (2D-IR) for determining the structure and dynamics of a system.

Chapter 1 gives an introduction to 2D-IR spectroscopy which is a technique that involves vibrationally exciting a sample, with an initial IR_{pump} pulse, then after a waiting time the sample is probed using an IR_{probe} pulse. A spectrum is obtained by scanning the frequency of the IR_{pump} pulse at a fixed waiting time. This technique allows the multiple transitions to be probed and valuable information (e.g. the coupling of vibrational modes) such as how energy is transferred, both within a molecule and to its environment, to be elucidated.

Chapter 2 describes investigations of the photodissociation of $\text{Fe}(\text{CO})_5$. This was chosen as a model system to investigate controlling the process of photodissociation using a vibrational excitation in solution. Transient 2D-IR spectroscopy (T-2D-IR) is a powerful tool, as electronic and vibrational pump pulses can be combined together with a sensitive IR probe to interrogate such processes on the ultrafast timescale. The T-2D-IR spectra of $\text{Fe}(\text{CO})_5$ in heptane and CH_2Cl_2 are presented and we demonstrate that photolysis of $\text{Fe}(\text{CO})_5$ can be controlled using a vibrational excitation such that the yield of the photoproducts can be

altered. We find that for $\text{Fe}(\text{CO})_5$ in heptane, exciting either vibrational mode 2 ps before photolysis by a 266 nm UV pulse results in a small increase in the formation of $\text{Fe}(\text{CO})_3$. Interestingly for CH_2Cl_2 , exciting only the lower vibrational mode 2 ps before photolysis by a 266 nm UV pulse results in decrease in the formation $\text{Fe}(\text{CO})_3$, no change is observed if the higher vibrational mode is excited. No change in the formation of $\text{Fe}(\text{CO})_3$ was also observed when the sample was immediately excited with an IR_{pump} pulse after the UV_{pump} pulse. These observations are discussed and the process of vibrationally controlling the photodissociation of $\text{Fe}(\text{CO})_5$ is clearly a very complex problem which requires further work to understand the underlying mechanism.

Chapter 3 discusses the use of 2D-IR spectroscopy to investigate the gelation mechanism of carrageenan. Polysaccharide gels are a very important component in the food, pharmaceutical and cosmetic industries. In the food industry these gels act as thickeners and stabilisers. They help give products their structure and physical properties such as texture. Carrageenan is one such polysaccharide gel and contains organosulfate groups. It has been shown that it limits ice crystals growth in frozen foods which is crucial to the stability of the product. There are three different types: ι , κ and λ -carrageenan, each of these differ in the number and position of the organosulfate functional groups. The interactions in and around these functional groups are thought to be critical to the properties of the gel and the gelation mechanism. 2D-IR spectroscopy was used to directly probe these functional groups and how their interactions change as gelation occurs. For ι -carrageenan in the solution state, the organosulfate groups exist in a large broad range of environments. Upon gelation, cross peaks were observed in the 2D-IR spectra, indicating that the organosulfate groups exist in

three main conformations which all interact with each other. The different environments are thought to relate to different cation interactions. These distinct environments were not observed for κ -carrageenan. We found that the spectral slices remained similar at different temperatures and this was interpreted to be due to the structural differences between ι -carrageenan and κ -carrageenan, with the latter only having one organosulfate group per monomer. The solvent dynamics of carrageenan was also investigated using ferrocyanide as a probe molecule by measuring the spectral diffusion using 2D-IR. The energy of water H-bonds in carrageenan was determined to be approximately half that of bulk water. This was thought to be because of disrupted H-bonding networks. While the bond reformation time was estimated to be approximately four times that of bulk water and was assumed to be due to confined water bonding networks. There was no discernible difference between the two types of carrageenan.

Chapter 4 investigated the interaction between salivary mucins and tea polyphenols. 2D-IR spectroscopy was used to probe the interactions of the milk protein, β -lactoglobulin, and the green tea polyphenol, epigallocatechin gallate, specifically using the amide I band in order to observe any conformational changes of the protein. The N-terminal domain of MUC5B (NT5B) has been previously shown to interact the most with theaflavins, specifically theaflavin digallate. Again, using 2D-IR spectroscopy, the secondary structure of NT5B was determined to be formed of mainly β -sheet and some α -helix. When NT5B and theaflavin digallate interacted, conformational changes were observed. A reduction in β -sheet was observed while the α -helix conformation remained largely unchanged. There was also an

indication of formation of aggregates and the results were interpreted to indicate that the β -sheet conformation is potentially important in the process of mucins cross linking.

Chapter 5 outlines the details of the equipment and experimental approaches used in the Thesis, and two appendices contain other work I have been associated with during my PhD. Appendix B contains the investigation of tryptophan to heme electron transfer in myoglobin, using TRIR spectroscopy where the experimental data had been previously collected with my work undertaking all the analysis and interpretation of these results. Appendix C details an investigation of lignin formation, which is a complex biopolymer, and is important in the formation and function of cell walls. Confocal Raman microscopy was used to show chemically distinct lignin was produced when altering a cell signalling pathway, where, again the experimental Raman data had been previously collected with my work undertaking the multivariate analysis and interpretation of these results.

Abbreviations

1D-IR	Conventional linear infrared spectroscopy
2D-IR	Two-dimensional infrared spectroscopy
2DCOS	Two-dimensional correlation spectroscopy
AOM	Acousto-optic modulator
Arg	Arginine
ATR	Attenuated total reflectance
BS	Beamsplitter
CD	Circular dichroism
CH ₂ Cl ₂	Dichloromethane
CLS	Central line slope
CN	Cyano ligand
CO	Carbonyl ligand
CSS	Charge separated state
CT	Charge transfer
CT5B	C-terminal domain of MUC5B
DBA	Donor-bridge-acceptor
DFG	Difference frequency generator
DFT	Density functional theory
DMSO	Dimethyl sulfoxide
EC	Epicatechin
EGCG	Epigallocatechin gallate
FFCF	Frequency-frequency correlation function
FT	Fourier transform
FTIR	Fourier transform infrared spectroscopy
FWHM	Full width half maximum
H-bonding	Hydrogen bonding
IR	Infrared
IR _{Probe}	IR probe pulse
IR _{Pump}	IR pump pulse
ISC	Inter-strand coupling
ivCLS	Inverse central line slope
IVR	Intramolecular vibrational redistribution
LO	Local oscillator
MCR	Multivariate curve resolution

MCT	Mercury cadmium telluride
NLS	Nodal line slope
NMR	Nuclear magnetic resonance spectroscopy
NT5B	N-terminal domain of MUC5B
OPA	Optical parametric amplifier
PES	Potential energy surface
PFMCH	perfluoro(methylcyclohexane)
PLS	Partial least squares
ROA	Raman optical activity spectroscopy
SANS	Small-angle neutron scattering
SAXS	Small-angle X-ray scattering
T-2D-IR	Transient two-dimensional infrared spectroscopy
T-2D-IR-EXSY	Transient two-dimensional infrared exchange spectroscopy
TDC	Transition dipole coupling
TEM	Transmission Electron Microscope
TERS	Tip-enhanced Raman spectroscopy
TF	Theaflavin
TF1	Theaflavin-3-monogallate
TF2	Theaflavin-3'-monogallate
THDG	Theaflavin digallate
TRIR	Time resolved infrared spectroscopy
Trp	Tryptophan
UV	Ultraviolet
UV _{Pump}	UV pump pulse
Vis	Visible

Table of contents

1	Introduction	1
1.1	Infrared spectroscopy	2
1.2	Time-resolved infrared spectroscopy	3
1.3	2D-IR Spectroscopy	4
1.3.1	Double resonance technique	4
1.3.2	Single vibrational mode	5
1.3.3	Coupled oscillator	7
1.3.4	Inhomogeneous broadening and spectral diffusion	8
1.3.5	2D-IR cross peaks	9
1.4	Time-domain 2D-IR spectroscopy	12
1.4.1	Four wave mixing technique	13
1.4.2	Pulse-shaping technique	13
1.5	2D-IR spectroscopy of Proteins	18
1.5.1	Examples of 2D-IR spectroscopy	22
1.6	Transient 2D-IR	24

2	Vibrationally controlling the photodissociation of iron pentacarbonyl	29
2.1	Introduction	29
2.1.1	Vibrational control of photochemical reactions	32
2.1.2	Photochemistry of Fe(CO) ₅	33
2.1.3	2D-IR study of Fe(CO) ₅	41
2.1.4	Aims	41
2.2	Results	42
2.2.1	TRIR	44
2.2.2	Ground state 2D-IR	45
2.2.3	T-2D-IR spectroscopy	52
2.3	Discussion	62
2.3.1	Pulse sequence	62
2.3.2	Solvent dependence	64
2.3.3	Pump frequency dependence	64
2.4	Conclusion and future work	67
2.4.1	Conclusion	67
2.4.2	Future work	68
3	Gelation of Carrageenan	71
3.1	Introduction	71
3.1.1	Structure and Function	71
3.1.2	Gelation Process	72
3.1.3	Previous research	77

3.1.4	Vibrational Spectroscopy	79
3.1.5	Aims	83
3.2	Results	86
3.2.1	ι -Carrageenan	87
3.2.2	κ -Carrageenan	93
3.2.3	Solvent dynamics	99
3.3	Discussion	114
3.3.1	Carrageenan gelation	114
3.3.2	Solvent environment	117
3.4	Conclusion and future work	120
3.4.1	Conclusion	120
3.4.2	Future work	121
4	The interactions of mucins and theaflavins	125
4.1	Introduction	125
4.1.1	MUC5B	126
4.2	Tea polyphenols	127
4.3	Interactions between β -lactoglobulin and tea polyphenols	130
4.4	Evidence of interactions between polyphenols and mucins	131
4.4.1	Aims	133
4.5	Results	135
4.5.1	β -LG	136
4.5.2	NT5B	144

4.6	Discussion	151
4.6.1	β -LG with EGCG	151
4.6.2	NT5B	156
4.7	Conclusion and future work	158
4.7.1	Conclusion	158
4.7.2	Future work	159
5	Experimental	161
5.1	2D-IR spectroscopy setups	161
5.1.1	ULTRA spectrometer and laser system	162
5.1.2	Nottingham 2D-IR spectrometer and laser system	163
5.2	2D-IR spectroscopy data analysis and processing	165
5.2.1	Frequency domain	165
5.2.2	Time domain	166
5.3	FTIR spectroscopy	169
5.3.1	MCR analysis	171
5.3.2	Peak fitting	171
5.4	Sample preparation	172
5.4.1	Fe(CO) ₅	172
5.4.2	Carrageenan	172
5.4.3	Pure carrageenan samples	172
5.4.4	Carrageenan and ferrocyanide	173
5.4.5	Mucins and theaflavin	174

Table of contents	xvii
References	177
Appendix A 2D-IR Spectrometer operating procedures	187
A.1 Turning on the 2D-IR spectrometer	187
A.2 Turning off the 2D-IR spectrometer	188
A.3 Basic alignment of the 2D-IR spectrometer	188
A.4 Laser realignment of input beam	189
A.5 Calibration of 2D-IR pump beam	190
A.6 Collection of 2D-IR spectra	191
Appendix B Tryptophan-to-Heme Electron Transfer in Myoglobins	193
Appendix C Schengen-pathway controls spatially separated and chemically distinct lignin deposition in the endodermis	241

Chapter 1

Introduction

When investigating various systems, the structure and dynamics are of great importance. In order to determine the structure of the system, the probing technique needs to have good structural resolution i.e. can it distinguish between different conformations/bonds within the sample? NMR is a great example of a technique that has excellent structural resolution, as each nucleus (spin dependent) can emit a signal distinguishing itself.¹ Infrared (IR) spectroscopy also has good structural resolution which is bond specific.² The dynamics are the second important property in understanding various systems, as it allows non-equilibrium states to be studied. NMR can be used to study the dynamics of a system, however, even with modern advancements, the time resolution is limited to microseconds.¹ Two-dimensional infrared (2D-IR) spectroscopy provides excellent time resolution down to femtoseconds as well as great structural resolution where vibrational couplings can be probed.³

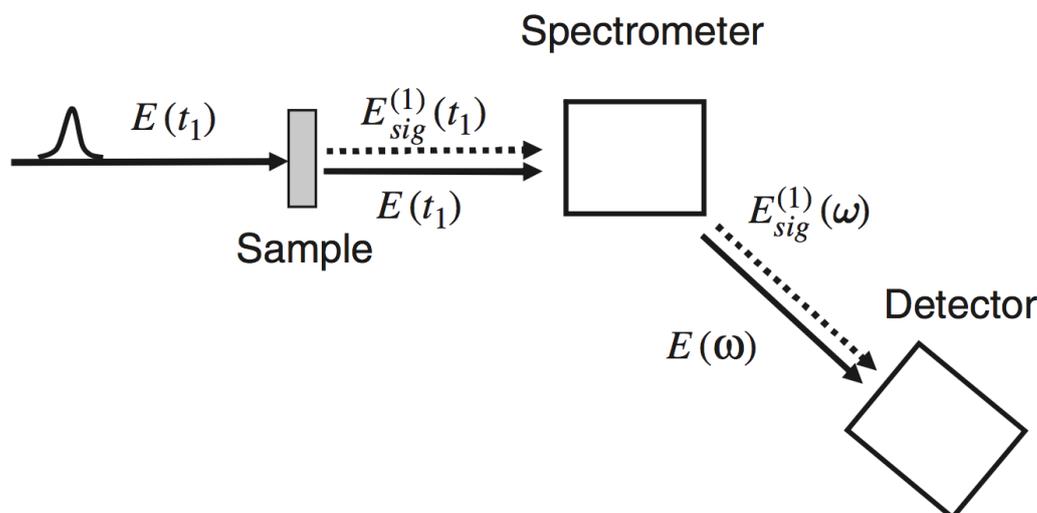


Fig. 1.1 Schematic of normal IR spectroscopy showing incident light on a sample with the signal emitted being measured using a spectrometer and detector. Reproduced from [4].

1.1 Infrared spectroscopy

Although conventional IR spectroscopy (1D-IR) provides a lot of information about a sample, for example bond sensitive peaks, no information is provided as to how each of the peaks in the spectra correlate with each other. For example, the peaks could be different vibrations from the same molecule or they could be vibrations for different types of molecules. This is indistinguishable using 1D-IR spectroscopy.

1D-IR essentially involves the sample interacting with an incident electric field, causing the electrons of a molecule to oscillate. These oscillating electrons emit an electric field collinear to the incident electric field, with a phase shift of π . This results in the emitted signal destructively interfering with the incident field, which appears as an absorption in the IR spectrum. For 1D-IR only one transition is probed at a time.²

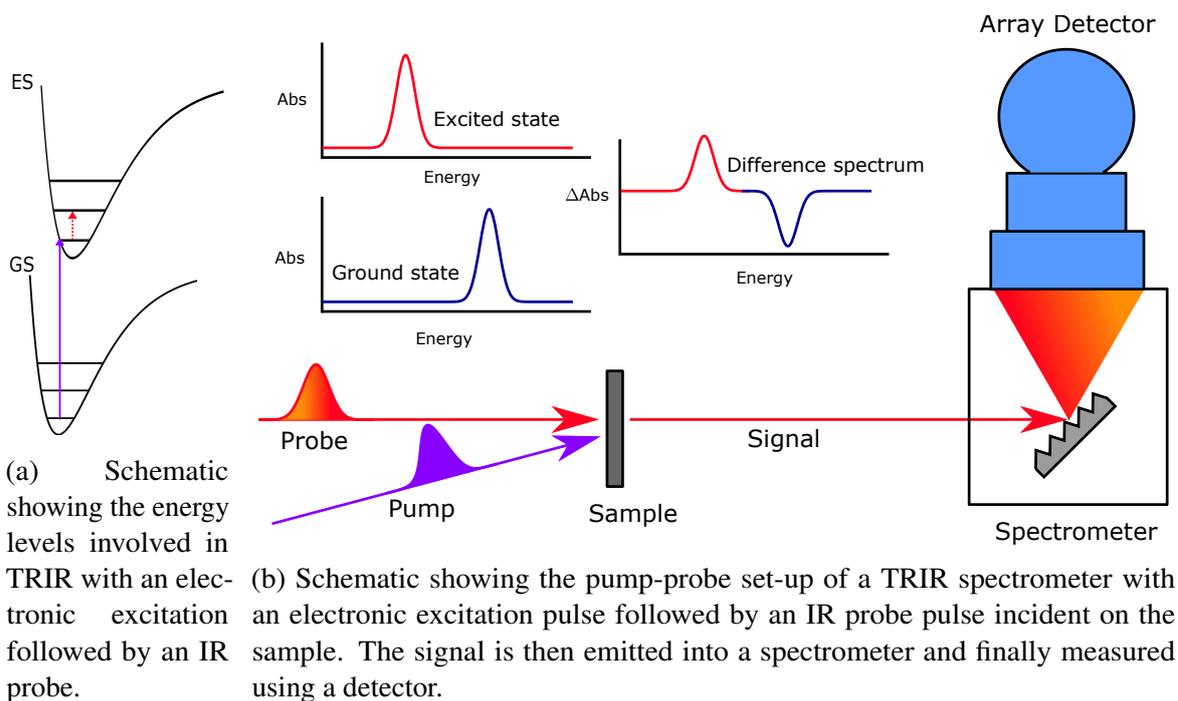


Fig. 1.2 Schematic of TRIR spectroscopy.

1.2 Time-resolved infrared spectroscopy

Femtosecond time-resolved infrared spectroscopy (TRIR) is a 3rd order non-linear spectroscopy technique. It involves a UV or visible pump pulse together with a broadband IR probe pulse. The pump pulse is first incident on a sample which electronically excites the sample. Then, after a delay time t a broadband probe pulse is incident on the sample which measures the spectrum with the sample in the excited state. A spectrum collected without the pump pulse (with the sample in electronic ground state) is then subtracted from the excited state spectrum to create a TRIR difference spectrum. The time delay can then be varied to ‘watch’ how the excited state changes with time. This is summarised in Figure 1.2.⁵

1.3 2D-IR Spectroscopy

2D-IR allows multiple vibrational transitions to be probed simultaneously.³ This provides a more significant amount of information about how various vibrational modes correlate and couple. 2D-IR spectra can be collected in either the frequency or time domain.⁴ The frequency domain method called double resonance spectroscopy is similar, conceptually, to time resolved IR spectroscopy (TRIR) and is the easiest to understand. The following section first describes the basic set-up of the double resonance technique and how spectra are obtained, followed by what information can be extracted from them. Finally, the set-up of the time domain pulse shaping method is described. 2D-IR spectroscopy was used throughout this work to investigate intra and intermolecular processes. Spectra were obtained in both the frequency and time domain technique. Therefore the remainder of this chapter is dedicated to the details of this technique.

1.3.1 Double resonance technique

The double resonance technique essentially involves a narrowband pump pulse which excites the sample, then after a waiting time T a broadband probe pulse is incident on the sample (the pulse train is shown in Figure 1.5). A spectrum is then recorded by dispersing the signal using a spectrometer and the intensity is then measured using a Mercury Cadmium Telluride (MCT) array detector. The difference spectrum is obtained by measuring the signal produced from only the probe pulse and comparing it with the signal produced from the pump pulse followed by the probe pulse.⁶ The pump pulse frequency is then scanned over a range at

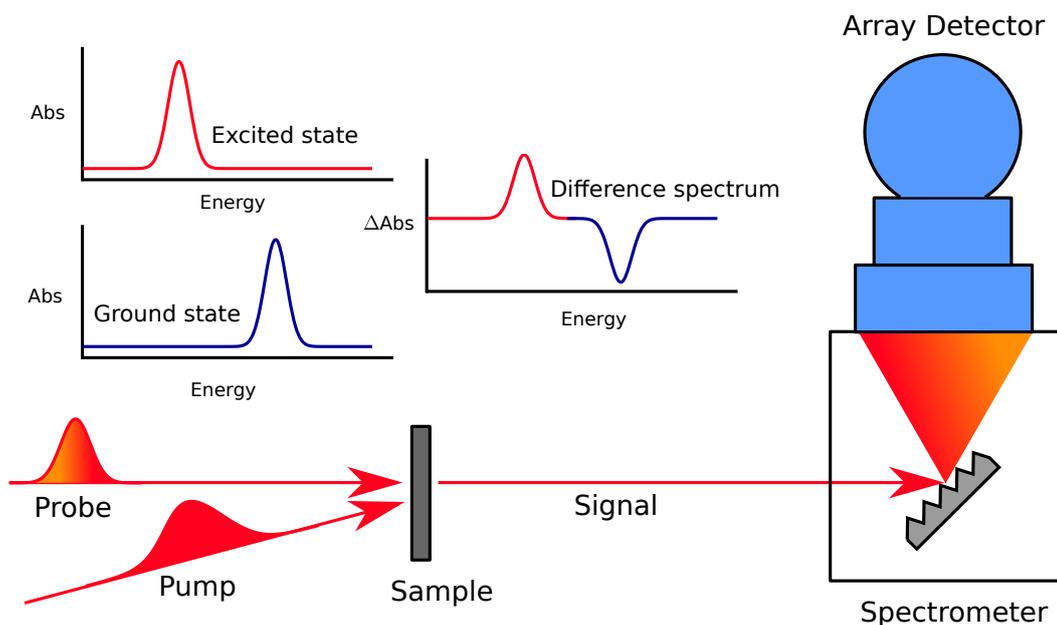


Fig. 1.3 Schematic showing the pump-probe set-up of a 2D-IR spectrometer which is similar to that of TRIR. A narrowband IR pump pulse is incident on the sample followed a broadband IR probe pulse. The signal is then emitted into a spectrometer and finally measured by a detector. Above, the vibrational ground state and excited spectra are shown alongside the calculated difference spectrum.

a fixed waiting time to collect a complete 2D-IR spectrum (shown in Figure 1.4). This allows the resonance of the vibrational modes to be probed. Figure 1.3 shows a schematic representation of the pump-probe geometry set-up.

1.3.2 Single vibrational mode

Figure 1.6 displays the vibrational energy levels for a single vibrational mode along with the corresponding 2D-IR spectrum. Initially molecules are in the ground state ($|0\rangle$), the pump pulse then excites the molecules into the first excited state ($|1\rangle$) undergoing the fundamental transition (where ω_{01} is the transition frequency). After a waiting time, the probe pulse then interacts with the molecules. This can cause two things to happen: stimulated emission

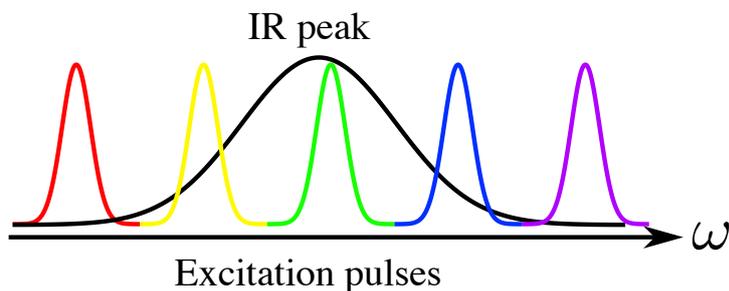


Fig. 1.4 The black peak represents a normal IR spectral band in the frequency domain. The narrower peaks represent the pump pulses at different wavelengths along the spectral band.

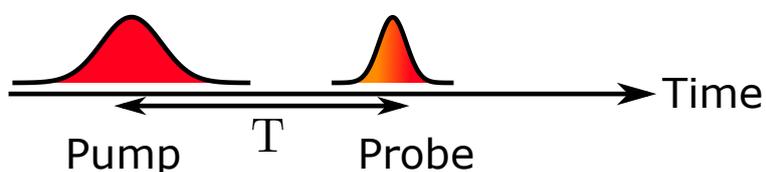


Fig. 1.5 The pulse train used to obtain 2D-IR spectra in the frequency domain. A long narrowband pump pulse is incident on the sample, then after a waiting time T a short broadband probe pulse is incident.

occurs in which the molecule returns back to the ground state (ω_{01}) or excited state absorption (ω_{12}) occurs where the molecule is excited to the second state ($|2\rangle$).⁷ This results in a 2D-IR spectrum containing one positive (red) and negative peak (blue). The negative peak located on the diagonal axis represents the fundamental transition (bleach or parent band) and occurs because there is a loss in the ground state population of molecules. The positive peak located at a slightly lower frequency in the probe axis, represents the first hot band transition (ω_{12}) and is redshifted due to the anharmonicity of the potential. This occurs because there is an increase in the population of $|1\rangle$. The anharmonicity (Δ) of the potential can easily be calculated by simply measuring the distance between the two peaks.⁴

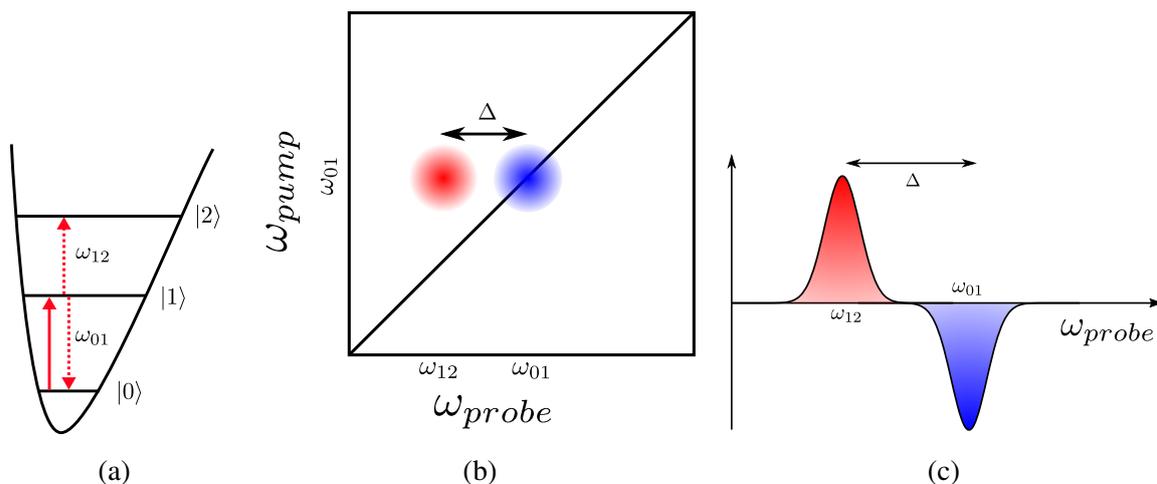


Fig. 1.6 (a) Vibrational energy levels of a single vibrational mode, showing the fundamental ω_{01} and the first hot band ω_{12} transitions. The solid arrow represents the pump pulse action, the dotted arrows represent the probe pulse action (stimulated emission and first hot band transition). (b) The corresponding 2D-IR spectrum, the blue and red circles represent negative and positive peaks respectively. $\Delta = \omega_{01} - \omega_{12}$ is the anharmonicity of the vibrational mode. (c) Is a pump spectral slice of the 2D-IR spectrum showing a difference spectrum taken at ω_{01} , showing the two peaks and anharmonicity.

1.3.3 Coupled oscillator

While the 2D-IR spectrum of a single vibrational mode provides a lot of information, the spectrum of a relatively simple coupled oscillator system contains much more information, but as can be seen from Figure 1.7, becomes very complex. This is because the system can evolve using many more different energy pathways. The solid arrows represent processes caused by the pump pulse and the dashed arrows represent the processes caused by the probe pulse. The cross peaks in this case can be exploited to calculate the coupling angle and distance.⁴

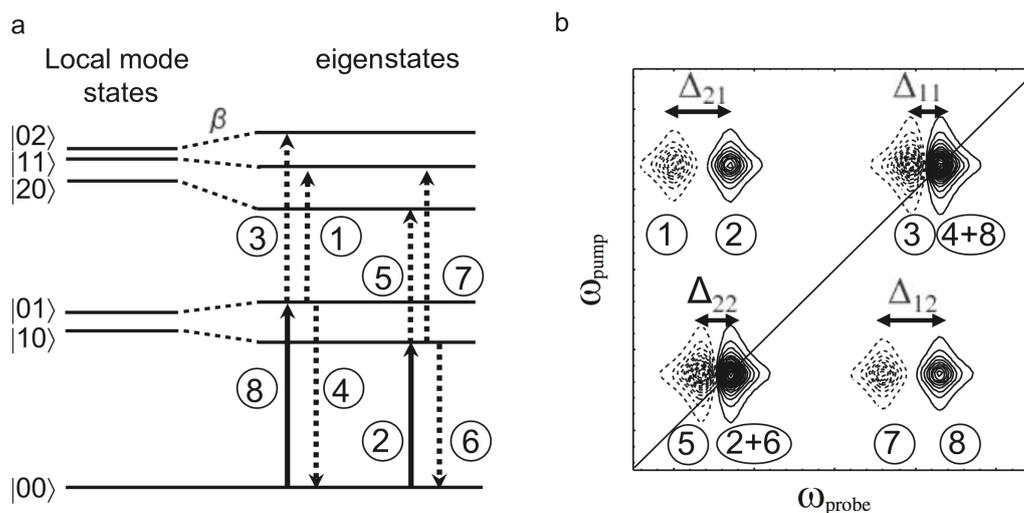


Fig. 1.7 (a) The vibrational energy levels of a two-level coupled oscillator, the solid arrows show processes performed by the pump pulse, the dashed arrows shows processes caused by the probe pulse. (b) The corresponding 2D-IR spectrum of the coupled oscillator system. The diagonal peaks correspond to spectral bands from normal 1D-IR, whereas the cross peaks correspond to the coupling between the vibrational modes. Reproduced from [4].

1.3.4 Inhomogeneous broadening and spectral diffusion

Peaks in a 1D-IR spectrum are a convolution between the homogeneous and inhomogeneous broadening of signals. The homogeneous broadening is caused by the uncertainty in the vibrational lifetime of a state. Whereas, inhomogeneous broadening results from molecules being contained within slightly different environments (micro-ensembles). It is almost impossible separate the two types of line broadening using 1D-IR. The two different types of broadening can be separated using 2D-IR spectroscopy by varying the waiting time between pump and probe pulses. Figure 1.8 shows two 1D-IR and two 2D-IR spectra corresponding to a small and large waiting time.⁸

When there is a small waiting time, the narrowband pump pulse only excites the micro-ensemble within in the spectral band, the probe pulse then immediately measures this excita-

tion. However, the system has not had any time to evolve, therefore the excited molecules remain in the same micro-ensemble. This results in two peaks in the 2D-IR spectrum which are broad in the diagonal axis but narrow in the anti-diagonal axis. The conventional lineshape (1D-IR) is replicated along the diagonal axis whereas the homogeneous lineshape is sometimes represented in the anti-diagonal axis. If the waiting time is then increased to allow the system to evolve, then the excited molecules in the micro-ensemble redistribute among the range of environments, transferring energy. This results in the peaks becoming symmetrical and broadening in the anti-diagonal axis, this is due to inhomogeneous broadening. The inhomogeneous broadening can be measured as function of the waiting time, this gives great insights into the dynamics of the system.⁹

Spectral diffusion essentially utilises this process to measure the spatial vibrational coupling. For example you can excite one vibrational mode and measure how the vibrational energy is transferred to the surrounding environment. The changes can then be viewed as a function of time. This is typically utilised to investigate local solvent fluctuations.⁷ In Chapter 3, the spectral diffusion of the probe molecule, ferrocyanide, in carrageenan solutions was measured to investigate the effect the gel had on the solvent environment.

1.3.5 2D-IR cross peaks

Cross peaks are spectral features located on the off diagonal axis of spectra, they represent how the various diagonal peaks correlate with each other. These peaks contain a lot of

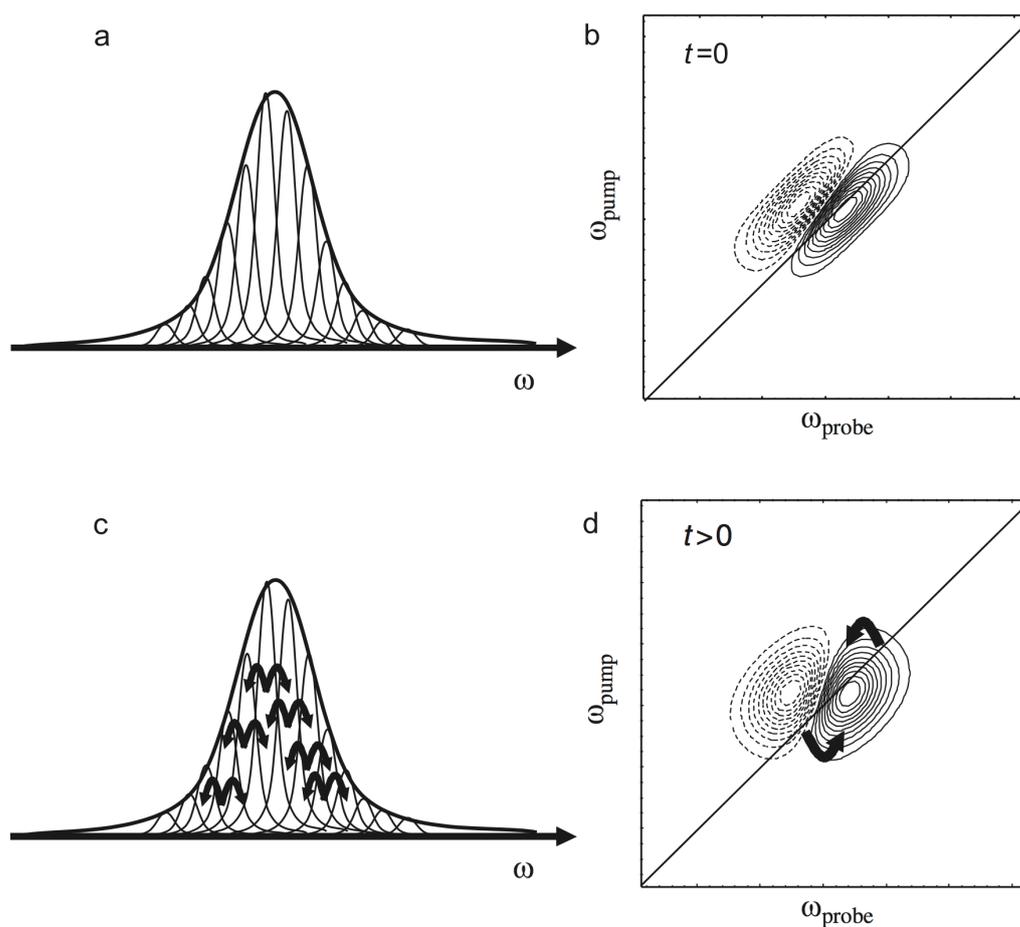


Fig. 1.8 (a) 1D-IR spectrum of the sample with the micro-ensembles represented by the narrower peaks. (b) The corresponding 2D-IR spectrum of the sample with only a small waiting time between pump and probe pulses. (c) 1D-IR spectrum of the sample, here the molecules have evolved in time and changed environments (micro-ensembles). (d) The corresponding 2D-IR spectrum of the sample with a larger waiting time between the pump and probe pulses. This allows the system to evolve leading to broader lineshape in the anti-diagonal axis. Reproduced from [4].

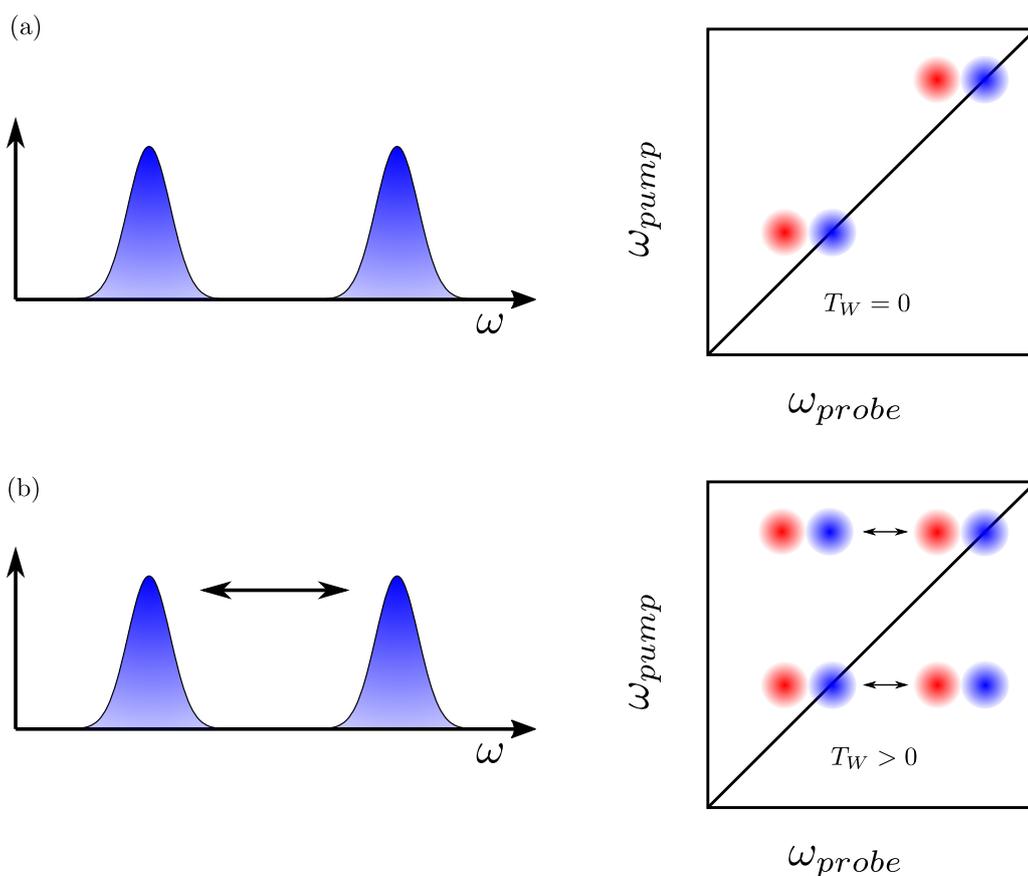


Fig. 1.9 (a) 1D-IR spectrum (left) of a sample with two conformations A and B with the corresponding 2D-IR spectrum with small waiting time between pump and probe pulses - a bimodal distribution is observed (only diagonal peaks). (b) 1D-IR (left) and 2D-IR (right) spectra for the sample at larger waiting times, cross peaks are observed between diagonal peaks demonstrating that molecules are ‘pumped’ in one conformation but are then ‘probed’ as the other conformation.

information about the system and could be a result of spectral diffusion, coupling, chemical exchange or fermi resonance.⁷

Chemical exchange

NMR has been used for many years to measure chemical exchange of molecules, however this process only has a time resolution of microseconds.¹ 2D-IR spectroscopy can also be used to

measure chemical exchange but down to a femtosecond timescale, which allows processes such as *cis* to *trans* interconversion to be measured. Take for example a mixture of A and B conformations separated by a small energy barrier. Figure 1.9 displays the 2D-IR spectra for small and long waiting times. At small waiting times a bimodal distribution is observed (two sets of diagonal peaks), because at early times molecules are in either conformation and there is a very low transition probability. If the waiting time is increased, then cross peaks are observed in the off diagonal axis. This is because there is greater transition probability due to energy fluctuations overcoming the energy barrier – i.e. a molecule is ‘pumped’ while in the A conformation but then ‘probed’ as the B conformation.⁴

1.4 Time-domain 2D-IR spectroscopy

The double resonance technique for obtaining 2D-IR spectra has some limitations. For example the time resolution is limited to the duration of the narrowband pump pulse, which is *ca.* 1-2 ps due to its narrowband nature. The lineshape of the peaks in the spectra are also limited to the bandwidth of the pump pulse which is usually *ca.* 10 cm^{-1} .⁶ Two separate optical parametric amplifiers (OPA) and difference frequency generators (DFG) are also required to produce individually tuned mid-IR pulses for the pump and probe. The method for collecting 2D-IR spectra in the time-domain is more sophisticated but allows some of the limitations stated above to be resolved.^{10,11}

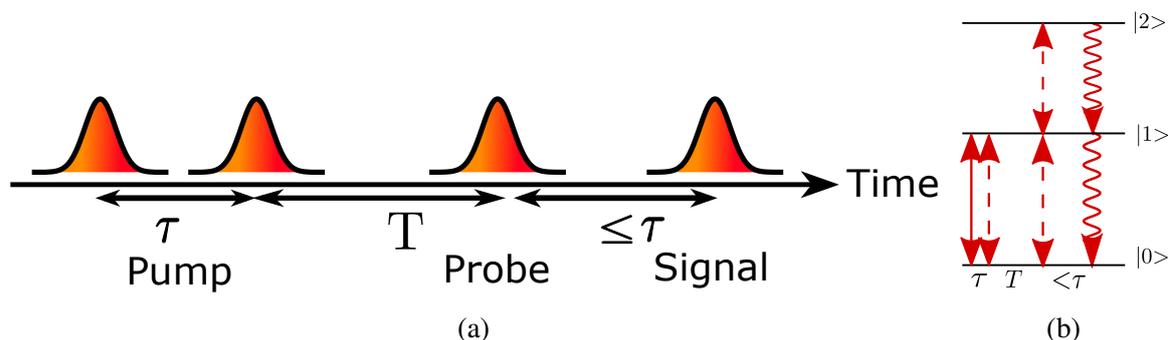


Fig. 1.10 (a) The pulse train used to obtain 2D-IR spectra in the time domain, all pulses are broadband. The first two pulses are pump pulses separated by a time τ . Following this a probe pulse is incident on the sample after a waiting time T . The signal is then emitted from the sample after a time delay. (b) The energy level diagram on the right represents each pulse interaction with the system.

1.4.1 Four wave mixing technique

The original method for time-domain 2D-IR involved a technique called four-wave mixing.¹²

Simply put, three incident pulses from different directions were spatially overlapped on the sample. A signal was then emitted in a separate direction, this was then combined with a local oscillator pulse (LO) to measure the signal in the time domain. This technique is however, extremely difficult to perform experimentally.^{4,7}

1.4.2 Pulse-shaping technique

The pulse-shaping technique is the most recent method developed to collect 2D-IR spectra.¹³ It combines the relative ease of a pump-probe set-up but collects spectra in the time-domain.¹⁰

Figure 1.10 shows the pulse train used to collect 2D-IR spectra in the time domain. There are three pulses incident on the sample, two pump pulses and one probe pulse. All three

pulses typically have the same large bandwidth (providing excellent time resolution).⁴ The first pump pulse creates an initial coherence in the system, then after a time τ the second pump pulse creates a population in the excited state. The probe pulse is then incident on the sample after a waiting time T , this creates a second coherence in the system. A signal is then emitted from the system within τ . A complete 2D-IR spectrum is then obtained by scanning through τ . This causes the different vibrational modes to go in and out of resonance which is reflected in the spectrum. The pump axis of the spectrum arises from Fourier transforming (FT) the scan of τ values.⁷

Pulse shaper

The pulse shaper is the key part of the apparatus that controls the frequency, phase and time delay of the pump pulses.¹³ Figure 1.11 is a schematic of the basic pulse shaper design which contains various gratings, cylindrical mirrors and an acousto-optic modulator (AOM). An initial laser pulse is first dispersed by a grating onto a cylindrical mirror, this aligns the range of frequencies spatially along the face of the AOM. A computer is used to generate arbitrary waveforms which are then transmitted to the AOM crystal, standing acoustic waves are then generated in the crystal itself which sets the amplitude and phase of each IR frequency diffracted by the AOM.^{14,15} The emitted frequencies are then focused onto another grating using a cylindrical mirror, this re-collimates the frequencies into one collinear path. The pulse-shaper, with the time-domain method, is specifically used to split one incident pump pulse into two pulses separated by τ .¹³

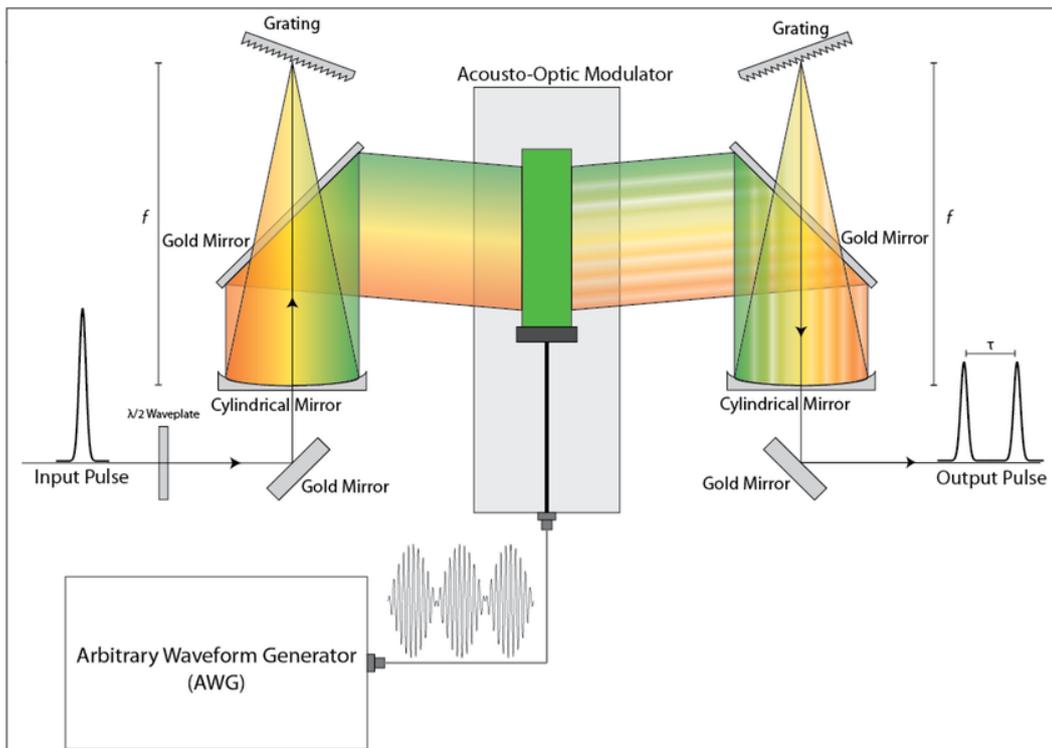


Fig. 1.11 Schematic showing the internal set-up of a pulse shaper. An incident laser pulse is split by frequency onto a cylindrical mirror, this aligns frequencies of light spatially along the face of the AOM. The AOM can then control the phase, frequency and pulse duration. Another cylindrical mirror then focuses the frequencies of light onto another grating which collimates the frequencies of light onto one beam path. Reproduced from [16].

Experimental set-up

A representation of the pulse-shaping 2D-IR spectrometer used is shown in Figure 1.12. A Ti:Sapphire laser emits 800 nm, *ca.* 100 fs pulses at a repetition rate of 1 kHz. This enters an OPA which converts the frequency of these pulses into two near-IR frequencies. A DFG then combines these two frequencies together to create tunable mid-IR pulses. The mid-IR pulse is then split into pump and probe pulses using a beam splitter. The pump pulse enters the pulse shaper, two pump pulses separated by τ are then emitted by pulse shaper and incident on the sample. After a waiting time T_W the probe pulse is then incident on the sample, following this a signal is then emitted which is collinear with the probe pulse. This allows the probe pulse to act as a LO with the signal which is dispersed by a spectrometer and the time varying signal is measured using a MCT array detector. A 2D-IR spectrum is collected by scanning the delay time between the two pump pulses while the waiting time is fixed. The pump axis of the spectrum is from the FT of τ whereas the probe axis is from the FT signal by the spectrometer.

The pulse-shaping method has many advantages over the double resonance technique. Mainly, it has superior time resolution *ca.* 100 fs compared with *ca.* 1-2 ps for the frequency domain method, due to the broadband nature of the pulses. The phase of each pump pulse can be controlled using the pulse shaper, which allows the phases to be progressively cycled. This reduces the scatter and background from signals significantly, which results in a superior signal to noise ratio.^{10,13–15}

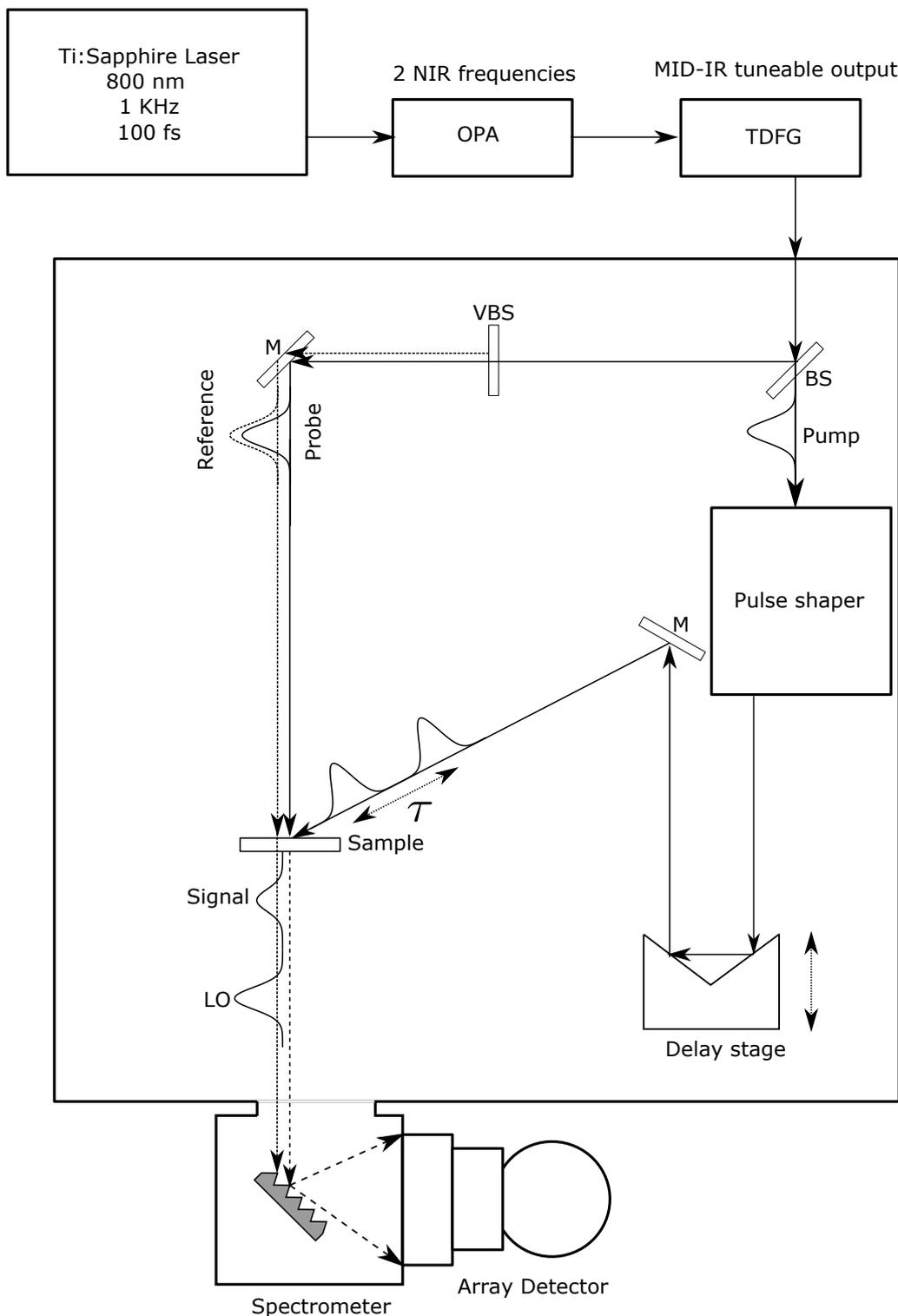


Fig. 1.12 Schematic showing the PHASETECH 2D-IR spectrometer set-up. Tuneable mid-IR pulses are produced by a Ti:Sapphire laser in conjunction with an OPA and DFG. It is then split into pump, probe and reference pulses using beam splitters (BS and VBS), the pump pulse is split into two pulses by the pulse shaper. Finally, all three pulses interact with the sample producing a signal which is measured using a spectrometer and MCT array detector. The reference pulse is used to account for shot-to-shot variations.

1.5 2D-IR spectroscopy of Proteins

2D-IR spectroscopy is ideally suited to investigating biomolecular systems because they have many IR active vibrational modes and the virtually limitless time resolution available allows aspects such as protein folding, ligand binding, protein-solvent interactions and tertiary structures to be probed.^{17,18} Interactions between a salivary mucin protein and tea polyphenols have been investigated in Chapter 4. Below is a brief introduction into the structure and IR spectroscopy of proteins.

Proteins are biomolecules formed of units called amino acids and are critical to life. The function of proteins are fundamentally linked to their structure. These polypeptide chains consist of repeating core units called a peptide, with each peptide containing a specific side chain which defines the amino acid. The sequence of the amino acids forming the protein is known as its primary structure. The repeating units of the peptide interact by forming H-bonds between them, these result in local structural elements which are defined as the secondary structure of the protein. β -sheet and α -helix are the two main types of secondary structure (shown in Figure 1.13) with others including random coil and PPII conformations. The overall 3D structure of the protein is referred to as the tertiary structure. Finally, the quaternary structure is defined as how the monomeric protein units interact to form larger protein complexes.¹⁹

Vibrational spectroscopy has been used extensively to probe protein structure.²¹ Changes in side-chain environments can result in frequency shifts, however, IR spectroscopy is particularly sensitive to the amide vibrational modes. There are three main modes: the amide

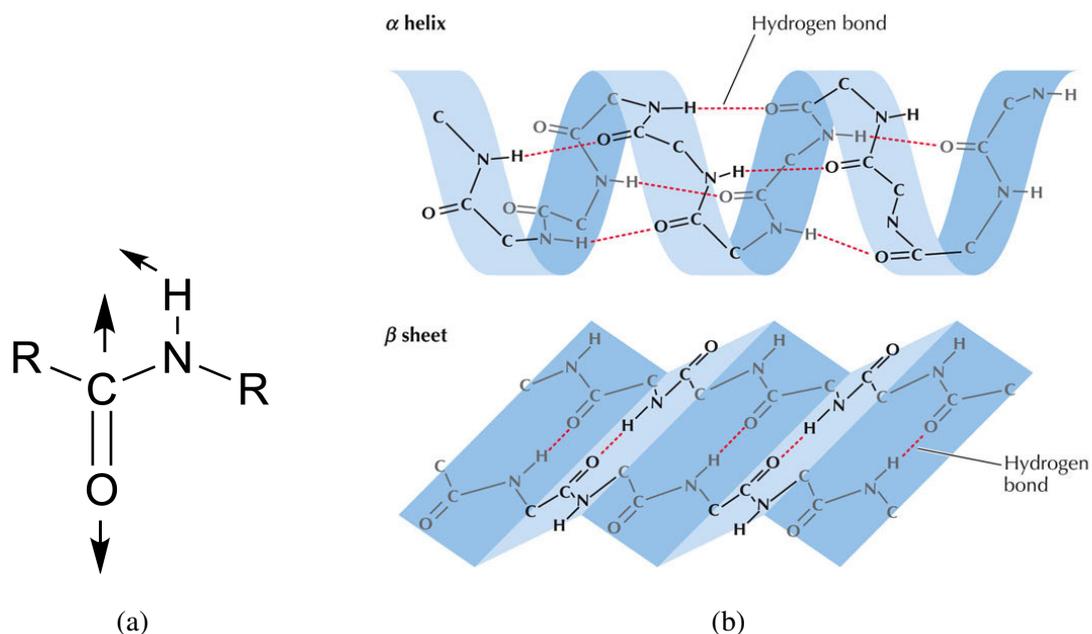


Fig. 1.13 (a) Peptide unit showing the amide I vibrational mode. (b) α -helix structure above and β -sheet structure below showing the H-bonding between peptide units. Reproduced from [20].

I mode (*ca.* 1650 cm^{-1}) is mainly due to the stretching mode of C=O and in-plane bending mode of N-H and the amide II (*ca.* 1550 cm^{-1}) and III (*ca.* 1350 cm^{-1}) modes are both from different combinations of the N-H in plane bending and of the C-N stretching mode.^{2,22} The amide I band is particularly sensitive to the H-bonding and dipole-dipole interactions of the backbone but not significantly influenced by the side chain groups. The interactions of the individual vibrational modes result in what is called transition dipole coupling (TDC) which is the process by which the local oscillators couple with nearest neighbours resulting in de-localisation of the amide I mode. This results in it being an excellent probe of the protein secondary structure,²³ There are characteristic frequencies for different types of secondary structure for example, α -helix – *ca.* 1650 cm^{-1} , β -sheet – *ca.* 1630 cm^{-1} , β -turn – *ca.* 1670 cm^{-1} and random coil/disordered - *ca.* 1645 cm^{-1} .²⁴ In IR spectroscopy, these characteristic

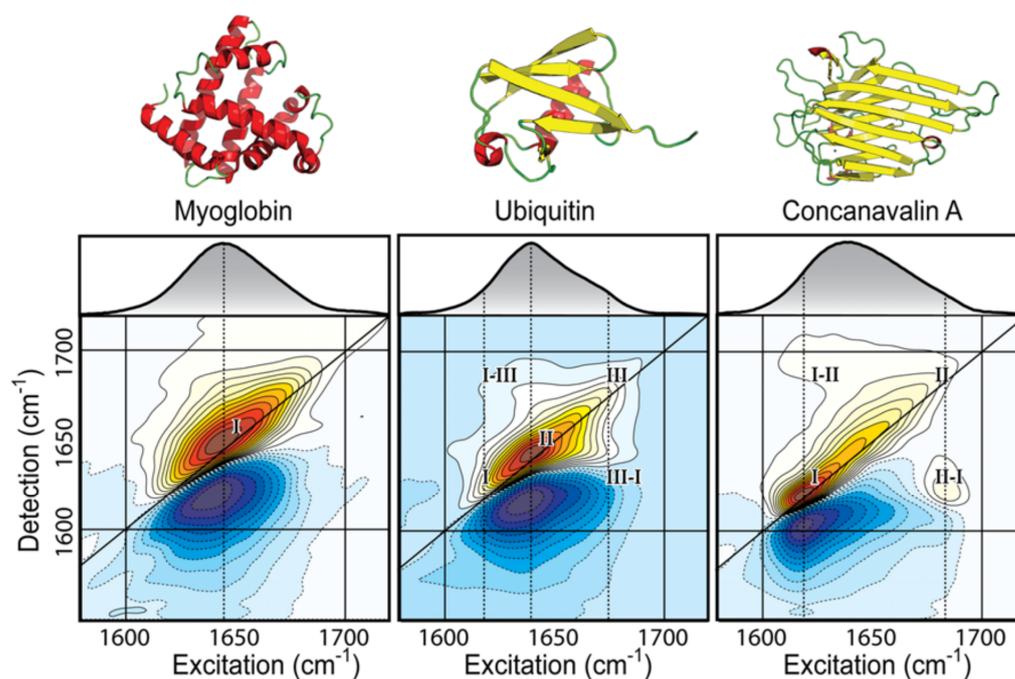


Fig. 1.14 2D-IR spectra of Myoglobin (α -helix), Ubiquitin (α -helix and β -sheet), and Concanavalin A (β -sheet) with the corresponding structures above. Each protein has a different spectrum due to the different structures. Reproduced from [7].

frequencies have been exploited to estimate the relative secondary structure composition by peak fitting the large broad amide I band.²⁵ However, this method does have to be used cautiously as the subjective nature of it can result in spurious results.²⁶

2D-IR

2D-IR spectroscopy is particularly sensitive to different protein conformations due to delocalisation of the vibrational modes and TDC. This results in complex but unique spectral lineshapes for α -helix and β -sheet secondary structures. Figure 1.14 illustrates the different 2D-IR spectra obtained for proteins in different conformations (Myoglobin, Ubiquitin, and

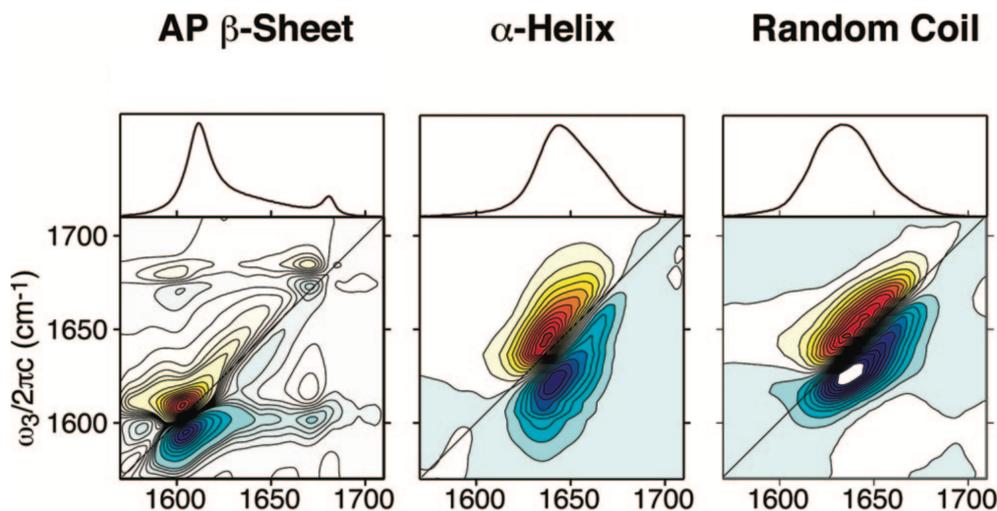


Fig. 1.15 FTIR (above) and 2D-IR (below) spectra of idealised secondary structures, showing the distinctive line shapes caused by different secondary structures. Reproduced from [7].

Concanavalin A). It can be seen that the FTIR spectrum for each protein is very similar but the 2D-IR spectra are all significantly different.⁷

β -sheet

Figure 1.15 shows the spectrum for an idealised β -sheet conformation. Two modes are visible, β^\perp (1620 cm^{-1}) and β^\parallel (1670 cm^{-1}) which are responsible for forming the ‘Z’ shape in the spectrum, with the coupling causing cross peaks in the off-diagonal region. The more intense β^\perp vibrational mode is due to inter-strand coupling i.e. through the H-bonding of strands. Whereas the less intense β^\parallel mode is due to the intra-strand coupling of oscillators i.e. bonded neighbours. The size of the β -sheet directly affects the β^\perp inter-strand coupling, which undergoes a redshift in frequency as well as an increase intensity as a result of a larger β -sheet. While the β^\parallel mode is largely unperturbed by the size of β -sheet. Therefore the $\beta^\perp/\beta^\parallel$ ratio and $\beta^\perp - \beta^\parallel$ splitting can be used to estimate changes in β -sheet size. Parallel

β -sheet structures are similar to anti-parallel ones except the β^{\parallel} mode has virtually no intensity.²⁷

α -helix

α -helical structures have two different vibrational modes, an **A** mode at 1640 cm^{-1} and a much less intense double degenerate mode **E**₁ at 1650 cm^{-1} . However, even though 2D-IR lineshapes are narrower than FTIR, the peak widths are generally still too broad to distinguish them. Therefore α -helix spectra result in two broad diagonal peaks as shown in Figure 1.15. The size of the α -helix is inversely proportional to the anharmonic splitting of, specifically, the **A** vibrational mode.²⁸ Random coil secondary structures exhibit a similar spectrum to α -helix, however, the transition dipole strength of vibrational modes can be calculated using 2D-IR spectroscopy and the two conformations can be distinguished using this.^{29,30}

1.5.1 Examples of 2D-IR spectroscopy

Proteins

Moran *et al.* applied 2D-IR to investigate the process of amyloid fibril formation (aggregation of disordered proteins), specifically the denaturation process of γ D-crystallin by either thermally inducing it or by subjecting it to acidic conditions.³¹ This involved labelling the N-terminal domain with ¹³C (shifts the amide I mode to lower wavenumber), then measuring the amide I region of the 2D-IR spectrum of γ D-crystallin in either a thermally induced

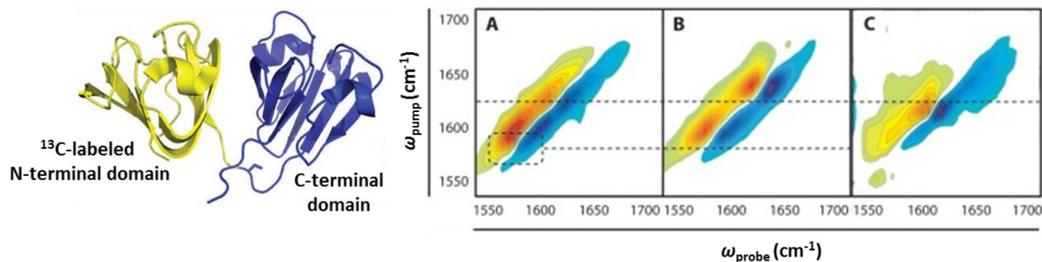


Fig. 1.16 Left, structural representations of the N- and C-terminal domains of γ D-crystallin. Right, 2D-IR spectra of γ D-crystallin for (A) thermally induced aggregates, (B) Native and (C) acid induced aggregates. This shows the different types of aggregates produced as a result of different denaturing processes. Reproduced from [18].

environment, or an acidic induced environment. Figure 1.16 shows the structures of the N- and C-terminal domains (yellow and blue respectively) with the 2D-IR spectra under (A) thermally induced, (B) native or (C) acid induced conditions. The native spectrum contains two sets of peaks, lower wavenumber region represents the doped N-terminal domain, the higher wavenumber region represents the C-terminal domain. When spectra A and B are compared, both regions shift to lower wavenumbers. However when spectra C and B are compared only the upper region (C-terminal domain) shifts to lower wavenumbers. It was therefore concluded that both the N- and C-terminal domains are denatured thermally, whereas only the C-terminal domain became denatured in acidic conditions.³¹

Deflores *et al.* investigated protein-solvent interactions using 2D-IR spectroscopy. The spectra were analysed in such a way that certain parts of the protein were observed to interact with D₂O while the other part of the protein interacted with itself. This was done by analysing the cross peaks between the amide I and II modes. The cross peaks of exposed parts of the protein appeared redshifted when compared with the solvent-shielded parts of the protein.³²

Water dynamics

2D-IR spectroscopy has also been used to investigate water dynamics which occur on a very fast timescale *ca.* 50-500 fs. Kraemer *et al.* investigated the temperature dependence (67 - 1 °C) of the water dynamics and found that while the structural correlations persist for longer at lower temperature, the intermolecular spectral diffusion is unaffected.³³ The dynamics of isotope doped ice has also been investigated; a very broad peak for the first excited state transition was observed in comparison with the ground state transition. This has been attributed to quantum beats between oxygen atoms of H-bonded water molecules.³⁴ The different dynamics of water molecules located in the bulk solution compared with molecules located at an interface (confined water) have also been investigated by collecting spectra of water located inside reverse micelles. It was observed that the dynamics of confined water was on the timescale of forming and breaking H-bonds, which was significantly slower than bulk water.³⁵

1.6 Transient 2D-IR

Transient 2D-IR spectroscopy (T-2D-IR) can be used to investigate how excited state reactions proceed because it provides femtosecond time resolution which is critical in probing photochemical dynamics. This technique was exploited in Chapter 2, where the photodissociation of Fe(CO)₅ was vibrationally controlled. T-2D-IR provides the same detailed information as ground state 2D-IR but probes the system in a non-equilibrium state.⁴ Essentially, the

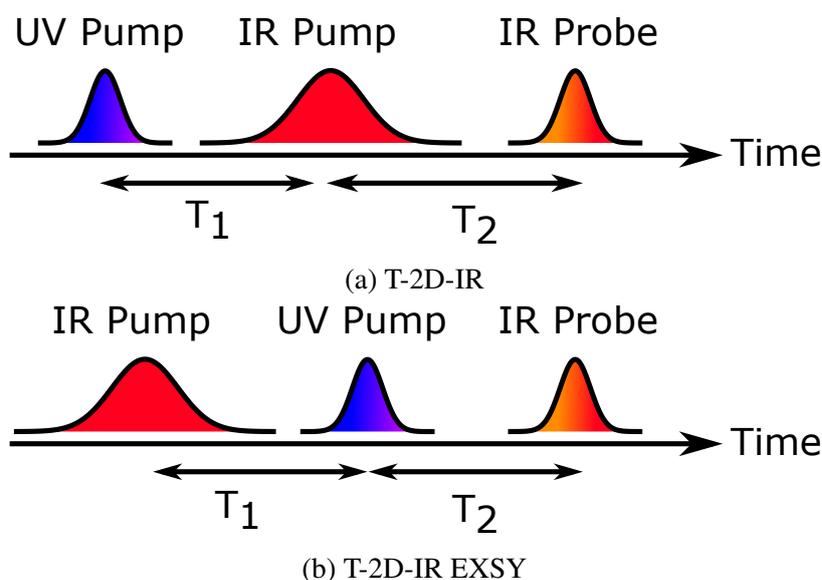


Fig. 1.17 The two different pulse sequences used to obtain T-2D-IR. (a) Conventional sequence of an initial UV_{pump} followed by an IR_{pump} after time T_1 , and finally the IR_{probe} after waiting time T_2 . (b) The pulse sequence used to obtain T-2D-IR-EXSY spectra, where the sample is initially excited by an IR_{pump} pulse followed by a UV_{pump} after T_1 and finally an IR_{probe} pulse after a waiting time T_2 .

process involves perturbing the system into a non-equilibrium state using either a UV/visible pulse or a sharp temperature change (T-jump), created using an IR pulse (typically excites the overtone of the OH stretch in water).⁴ The first example can be probed from a timescale of hundreds of femtoseconds up to nanoseconds or longer and examples include mapping ground vibrational states to electronically excited vibrational states. Whereas for the second example, it is typically on the timescale of microseconds and looks at processes such as protein folding.³⁶ The focus here will be on UV/visible excitation.

There are two T-2D-IR spectral techniques which contain three pulses (two pump and one probe pulse). The first is the most conventional and involves first exciting the sample with a UV_{pump} pulse followed by an IR_{pump} and finally an IR_{probe} . This is technically a fifth order process although if there is a long enough delay time between UV_{pump} and IR_{pump} then it can

be approximated to a third order process like normal 2D-IR. This technique allows reactions to be triggered and then watch how they evolve. The second method is called T-2D-IR-EXSY and involves the IR_{pump} being the first pulse followed by the UV_{pump} and finally the IR_{probe} pulse, this is shown in Figure 1.17. This sequence essentially ‘labels’ the vibrational modes and allows the observation of how the ground state vibrational modes translate into excited state vibrational modes.³⁷

Stewart *et al.* used T-2D-IR to elucidate the photoproduct formed from irradiation of (μ -propanedithiolate) $Fe_2(CO)_6$ – an inorganic analogue of a site in the hydrogenase enzyme. The T-2D-IR spectrum showed that there was only one photoproduct ((μ -propanedithiolate) $Fe_2(CO)_5$ (heptane)), this was because all of the ground state bleaches correlated to the new transient bands in the off-diagonal axis. The vibrational dynamics information was used to infer that the solvent coordinated to the new product. Finally, the likely structure of the photoproduct was determined by manipulating the polarisations of the various pulses.³⁸

T-2D-IR spectroscopy was used to characterise the spectral dynamics of the photoactivated catalyst $Re(CO)_3(bpy)Cl$ in tetrahydrofuran. This was the first example where the ground state 2D-IR spectrum was compared with the T-2D-IR spectrum. Changes in the spectral diffusion and vibrational relaxation between states were measured to give an insight into the microscopic effects the solvent has on charge transfer reactions.³⁹

Bredenbeck *et al.* demonstrated the first use of T-2D-IR-EXSY in 2004 with $[Re(CO)_3Cl(dmbpy)]$.⁴⁰ This molecule had already been studied extensively using TRIR, where the three carbonyl stretches were observed to blue shift when the molecule was excited and they

were thought to maintain their ordering. However, DFT calculations suggested that order of the modes switched. By employing T-2D-IR-EXSY, ground state vibrational bands were able to be labelled and tracked based on the off-diagonal cross peaks in the spectra when the molecule was excited. It was shown, conclusively, that the order of the carbonyl stretches was not preserved upon excitation of the molecule.⁴⁰

The previous sections above give an introduction to the technique of 2D-IR spectroscopy. Each of the following chapters and the two appendices contain specific applications of this and related techniques. Thus each chapter will contain its own specific introduction followed by detailed result and discussion sections. The two appendices report other research work I have been involved in in the form of a full paper which has either been submitted or is soon to be submitted.

Chapter 2

Vibrationally controlling the photodissociation of iron pentacarbonyl

2.1 Introduction

The overall aim of this work was to use vibrational excitations to control a photochemical process in solution. Many different phenomena can occur when a molecule is excited by light, for example intramolecular energy transfer, charge transfer, luminescence and dissociation.⁴¹ All of these different pathways are summarised in Figure 2.1. This chapter focusses on aiming to control the photodissociation process, for this $\text{Fe}(\text{CO})_5$ was selected as a model system to

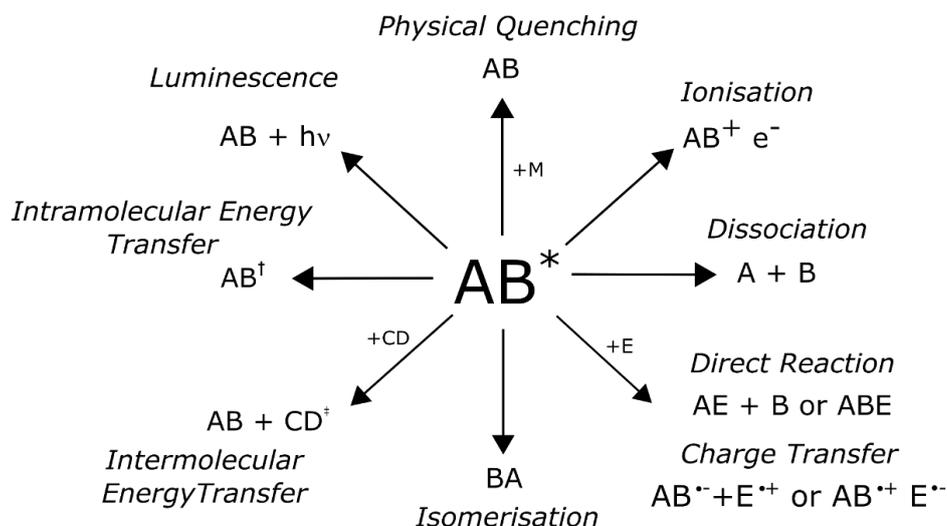


Fig. 2.1 A diagram showing all the different physical processes that can occur as a result of a photoexcitation. Reproduced from [41].

investigate. This is because it has been extensively studied meaning that the photochemistry and process of photodissociation (at least in the gas phase) is well understood.⁴²

Chemists have been trying to control chemical reactions for many centuries, various methods of control have been employed such as temperature, catalysts and light. Some of these provide very specific control of reactions such as light where the reaction can be switched on or off, whereas something like temperature provides far less control and mostly impacts the rate of reaction. A vibrational excitation along with a photoexcitation can provide extreme specificity to the chemical process where, not only can the reaction be switched on and off, but it can be steered down different pathways. This can eventually lead to increasingly complex systems to control.⁴³

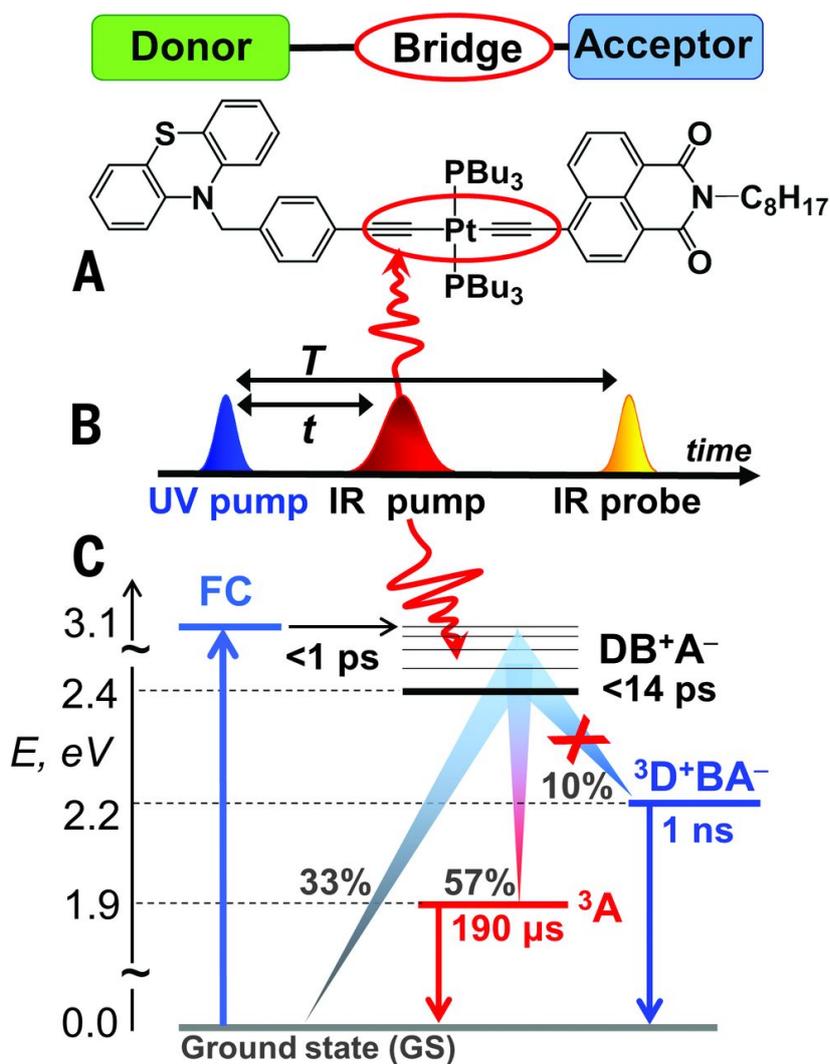


Fig. 2.2 A schematic representation of the experiment performed by Delor *et al.* (A) Molecular structure of donor-bridge-acceptor molecule. (B) The pulse sequence used in the experiment to influence the electron transfer. (C) A schematic showing the energy pathways the molecule takes during the experiment. If the bridge is vibrationally excited after UV excitation then this makes the charge separated state (³D⁺BA⁻) inaccessible. The molecule then has to relax via the other two pathways. Reproduced from [44].

2.1.1 Vibrational control of photochemical reactions

T-2D-IR spectroscopy can also be used to steer reactions by using the IR_{pump} pulse to perturb the energy of a photochemical reaction.⁴³ This can either increase or decrease the probability of a reaction occurring. An example of this is manipulating the process of charge transfer in donor–bridge–acceptor (DBA) assemblies.⁴⁴ Delor *et al.* investigated the photoexcitation and charge transfer (CT) of Pt(II) *trans*-acetylide. Here the excited state can decay via three pathways: straight back down to the ground state, via an acceptor-localised $\pi - \pi^*$ triplet state (³NAP) or via a charge separated state (CSS). When the certain stretching modes of the acetylene bridge are vibrationally excited, the CSS pathway suddenly becomes ‘blocked’ and the excited state is forced to decay via the other two pathways, this is shown in Figure 2.2.⁴⁵

The process of vibrationally controlling reactions takes advantage of vibronic coupling – in particular manipulating conical intersections. However, this process is very challenging to execute successfully as: vibrational lifetimes can be very small (e.g. H₂O has a lifetime of *ca.* 60 fs) and intramolecular vibrational redistribution (IVR) can result in a very fast transfer of energy into undesirable states.⁴⁵ In order to influence reactions there is a need for the correct frequency and timing of pulses so that the right vibrational mode is excited in the correct state. Ultrafast multidimensional spectroscopy technique provides the apparatus required for this fine control. If these factors are accounted for, the yield of certain photochemical reactions can be altered using IR_{pump} pulses.⁴³

Kern-Michler *et al.* employed a vibrational excitation to control the photochemistry of isotopomers.⁴⁶ The mixture of isotopomers in solution exhibit a near identical UV-Vis spectrum, however, this changes when the molecules are in a vibrationally excited state and since the molecules contain different isotopes their respective vibrational spectra are different. This allows the isotopomers to be selectively excited and therefore the photochemistry of near-identical molecules to be controlled. This work used the VIPER (Vibrationally Promoted Electronic Resonance) pulse sequence ($\text{IR}_{\text{pump}} \rightarrow \text{UV}_{\text{pump}} \rightarrow \text{IR}_{\text{probe}}$ which is effectively T-2D-IR) to selectively perform photochemistry.

2.1.2 Photochemistry of $\text{Fe}(\text{CO})_5$

The photodissociation of $\text{Fe}(\text{CO})_5$ has been extensively studied due to its interesting photochemistry.⁴⁷ It has been used as a model system to study spin forbidden reactions. Upon photolysis, the complex can lose one or more CO ligands, this has been studied at low temperatures as well as in gas and solution phases.⁴⁸⁻⁵⁰ $\text{Fe}(\text{CO})_5$ has been used in catalytic applications such as alkene hydrogenation and isomerisation.^{51,52} The main focus will be on the photodissociation of $\text{Fe}(\text{CO})_5$ in the solution phase, with a particular interest in the formation of $^3\text{Fe}(\text{CO})_3$. A brief summary of the literature follows.

Low temperature matrix isolation experiments

The photolysis of $\text{Fe}(\text{CO})_5$ was originally studied Stolz *et al.* using the matrix isolation technique which allows $\text{Fe}(\text{CO})_5$ to be studied at low temperatures, where molecules are

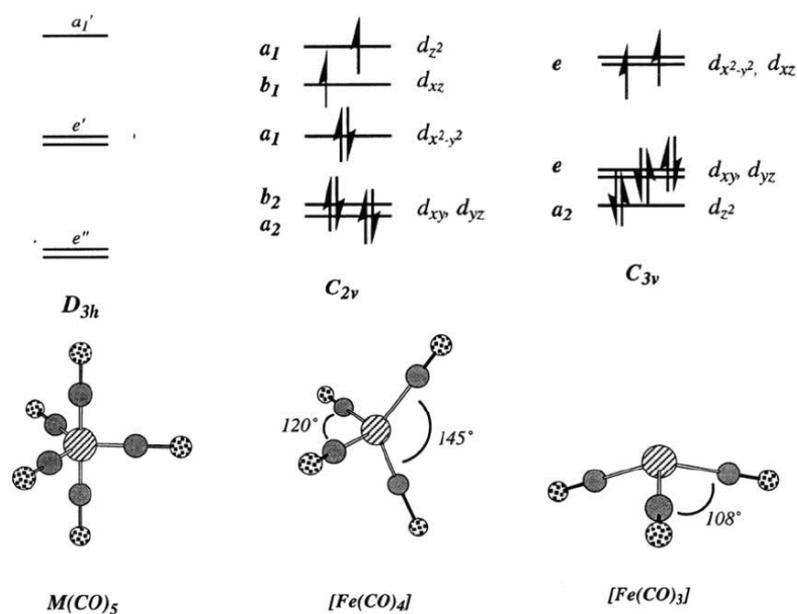


Fig. 2.3 The structures and electronic states are shown for $Fe(CO)_5$, ${}^3Fe(CO)_4$ and ${}^3Fe(CO)_3$. Reproduced from [47].

trapped in inert materials (polymers or frozen inert gases such as Xe or Ar). When photolysis occurs the products have long lifetimes due to the low temperature and inert nature of the matrix. It was shown that after UV photolysis of $Fe(CO)_5$ in various matrices at 4 K and 20 K that $Fe(CO)_4$ is generated, IR spectroscopy was used to measure this.⁵³ The structure was initially assumed to be C_{3v} , however this was later shown to be C_{2v} by Poliakoff and Turner who used ${}^{13}CO$ enriched $Fe(CO)_5$ to determine this (shown in Figure 2.3).⁴⁸ $Fe(CO)_4$ was predicted to be in the triplet ground state by Burdett using molecular orbital theory.⁵⁴ This was confirmed by Barton using magnetic dichroism studies and later with *ab initio* calculations.^{55,56}

Multiple other photoproducts were also observed in matrix isolation including $Fe(CO)_3$, polynuclear iron carbonyls and evidence for $Fe(CO)_4X$ (where $X = Xe, CH_4, N_2$).⁵⁷ The geometry of $Fe(CO)_3$ was determined to be C_{3v} (shown in Figure 2.3), again using isotope

substituted CO and was thought to be formed via sequential CO dissociation. Burdett also predicted that $\text{Fe}(\text{CO})_3$ would exist in a triplet ground state and this was confirmed by Barnes *et al.* using *ab initio* calculations.^{54,58} Bridged and unbridged polynuclear iron carbonyls were also observed in concentrated matrix isolation experiments where $\text{Fe}_2(\text{CO})_9$, $\text{Fe}_2(\text{CO})_8$ and $\text{Fe}_3(\text{CO})_{12}$ were all determined to be photoproducts.

Photochemical gas phase studies of $\text{Fe}(\text{CO})_5$

Low temperature matrix isolation techniques combined with IR spectroscopy was a very powerful combination because the photoproducts were stable long enough for steady state IR spectroscopy to measure them. However, in the gas phase the photoproducts of $\text{Fe}(\text{CO})_5$ were expected to be relatively short lived in comparison due to the nature of the unsaturated product $^3\text{Fe}(\text{CO})_4$. Time resolved IR spectroscopy (TRIR) was therefore required to observe these reaction intermediates.

Ouderkirk *et al.* observed the photodissociation of $\text{Fe}(\text{CO})_5$ in the gas phase with an argon buffer using a UV excitation wavelength of 248 nm.^{49,59} The first evidence of $^3\text{Fe}(\text{CO})_4$ and $^3\text{Fe}(\text{CO})_3$ was observed using TRIR with a time resolution of microseconds. $^3\text{Fe}(\text{CO})_3$ was observed to quickly convert into $\text{Fe}(\text{CO})_4$, and $\text{Fe}(\text{CO})_4$ is observed to slowly reform $\text{Fe}(\text{CO})_5$ (in a CO atmosphere), however not all of the photolyzed products reformed into $\text{Fe}(\text{CO})_5$.⁴⁹ Following this study, Seder *et al.* proposed that formation of $^3\text{Fe}(\text{CO})_3$ was due to a one photon and not a two photon process. This was determined based on the observation that increasing the laser power (flux) did not result in a change in the photoproduct yields.⁶⁰

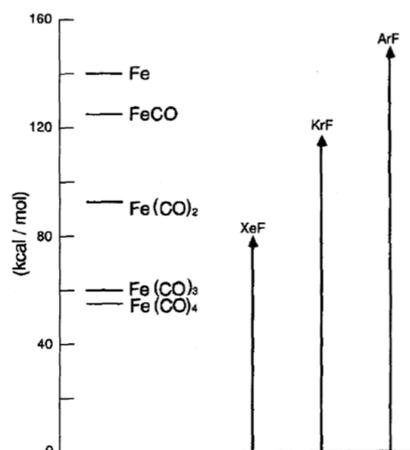


Fig. 2.4 A plot of the energy for different iron carbonyl complexes, showing the small amount of energy required to remove the second CO ligand. Reproduced from [63].

However, different photoproducts can be obtained by changing the excitation wavelength of the UV laser.⁶¹ The formation of $\text{Fe}(\text{CO})_3$ was thought to occur via a stepwise loss of CO ligands. After electronically exciting $\text{Fe}(\text{CO})_5$ – $^3\text{Fe}(\text{CO})_4$ is initially formed, while this is still vibrationally ‘hot’ another CO ligand is expelled forming $^3\text{Fe}(\text{CO})_3$.⁶² A laser photoelectron study of $\text{Fe}(\text{CO})_5$ determined the dissociation energy of losing successive CO ligands to be 55, 5, 32 and 23 kcal mol^{-1} for 1, 2, 3 and 4 CO ligands respectively (Figure 2.4).⁶³ This supported the possibility of the second CO ligand being expelled when the molecule is vibrationally ‘hot’.

Transient ionisation was performed by Trushin *et al.* to investigate the ultrafast photodissociation of $\text{Fe}(\text{CO})_5$. Five different processes were determined, ranging from 20 to 3300 fs, four of which were consecutive processes involved in the formation of $^1\text{Fe}(\text{CO})_4$. It was determined that $\text{Fe}(\text{CO})_5$ was initially excited into a MLCT state but then quickly crossed into a ligand field state, which is dissociative. This resulted in more than one of the equatorial

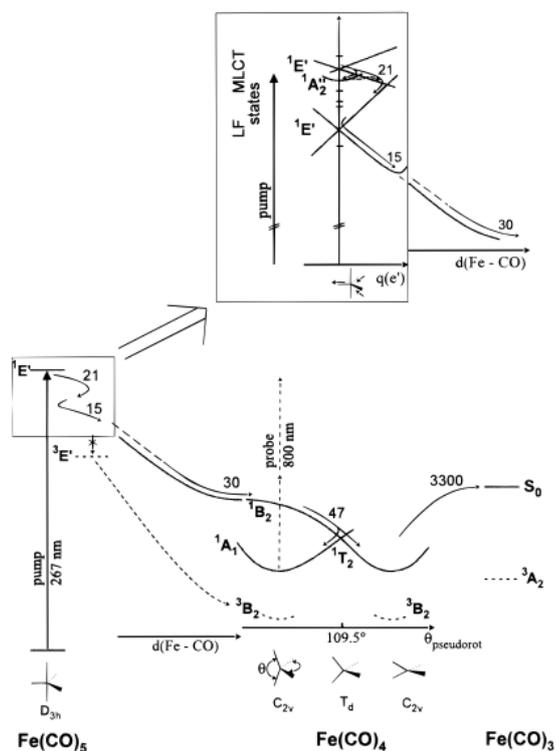


Fig. 2.5 A schematic of the proposed singlet pathway proposed by Trushin *et al.* for the photodissociation of $\text{Fe}(\text{CO})_5$, via a sequential pathway, in the gas phase. Reproduced from [64].

ligands accelerating with only one CO molecule ejected. The final step was assigned to the thermal loss of a further CO ligand to form $\text{Fe}(\text{CO})_3$. It was therefore concluded that sequential loss CO ligands occurred which is summarised by Figure 2.5.⁶⁴

Sequential loss of CO ligands for $\text{Fe}(\text{CO})_4$ was only recently observed directly by Wernet *et al.* using photoelectron spectroscopy.⁶⁵ The Fe metal centres and free CO molecules were probed and a kinetic model was fitted to the data which proved the sequential loss mechanism in the gas phase. A timescale of 3 ps was determined for the loss of the second CO ligand. These results also validated the mechanism photodissociation mechanism proposed by Trushin *et al.*

Bergt *et al.* controlled the photodissociation of $\text{Fe}(\text{CO})_5$ in the gas phase using coherent control techniques. This was done by attempting to maximise or minimise $\text{Fe}(\text{CO})_5/\text{Fe}$ ratio by controlling an 800 nm pulse using a pulse shaper. It was maximised when the pulse was very 'sharp' i.e. a short pulse. When the pulse was broadened by the pulse shaper the ratio is reduced by a factor of *ca.* 100. This suggested that it was likely a two photon process which was creating the Fe ions.⁶⁶

Photochemical solution phase studies of $\text{Fe}(\text{CO})_5$

The first solution phase TRIR study was performed by Church *et al.* in benzene- d_6 . $^1\text{Fe}(\text{CO})_4$ was observed to coordinate to a benzene- d_6 molecule. However, this experiment was performed on a microsecond timescale which was thought to be too long to observe triplet state photoproducts.⁶⁷ Snee *et al.* performed the first TRIR picosecond measurements in heptane where $^3\text{Fe}(\text{CO})_4$ was observed in solution phase for the first time (through the presence of two transient bands), however there was no evidence of $^3\text{Fe}(\text{CO})_3$.⁶⁸

Another, later study, again by Snee *et al.* but this time using an alcohol and PEt_3 as solvents, observed the presence of $^3\text{Fe}(\text{CO})_3$ (sol) and $^3\text{Fe}(\text{CO})_3$ (sol)₂ for the first time. $^3\text{Fe}(\text{CO})_3$ (PEt_3) was observed to form within 10 ps of a UV excitation of 295 nm, the $^3\text{Fe}(\text{CO})_3$ (PEt_3)₂ form was shown to be produced later via a further reaction with another solvent molecule. Another interesting result from the experiment was the formation of $^1\text{Fe}(\text{CO})_4$ (PEt_3) only in PEt_3 whereas $^3\text{Fe}(\text{CO})_4$ was formed followed by solvent coordination when alcohol was used.⁶⁸

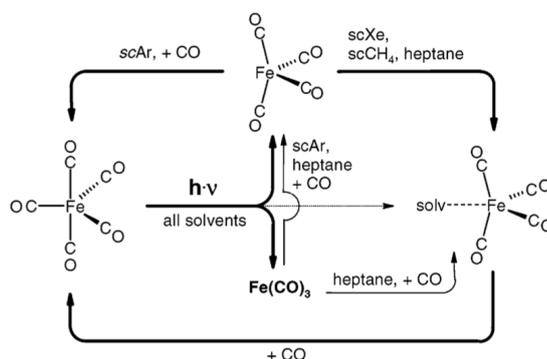


Fig. 2.6 A schematic of the different pathways of $\text{Fe}(\text{CO})_5$ after photoexcitation at 266 nm. Reproduced from [50].

$\text{Fe}(\text{CO})_5$ was also investigated in cyclohexane using TRIR, it was shown that $^3\text{Fe}(\text{CO})_3$ (sol) is formed via a one photon process as in the gas phase.⁶⁹ This was further supported by Portius *et al.*, where the photochemistry of $\text{Fe}(\text{CO})_5$ was studied in a variety of solvents: heptane, supercritical argon, xenon and methane with a UV excitation of 266 nm.⁵⁰ Figure 2.7 shows TRIR spectra obtained of $\text{Fe}(\text{CO})_5$ in heptane. Early time delays (<20 ps) show broad peaks due to vibrationally hot products, whereas at later times these peaks become more resolved as the excess energy is dissipated by solvent molecules. The two negative parent bands at 2000 and 2020 cm^{-1} are the ground state vibrational modes of $\text{Fe}(\text{CO})_5$, the two positive transient peaks at 1988 and 1967 cm^{-1} are assigned to $^3\text{Fe}(\text{CO})_4$ and the one small transient peak at 1926 cm^{-1} is assigned to $^3\text{Fe}(\text{CO})_3$. These products were observed to be stable up to the ns timescale. Interestingly, the yield of $\text{Fe}(\text{CO})_3$ relative to $\text{Fe}(\text{CO})_4$ increases from heptane \rightarrow sc CH_4 \rightarrow scXe \rightarrow scAr. This was thought to be explained by a sequential loss of CO ligands, e.g. $\text{Fe}(\text{CO})_5 \rightarrow \text{Fe}(\text{CO})_4 \rightarrow \text{Fe}(\text{CO})_3$, as well as a contribution from the solvent cage. Figure 2.6 summarises the different processes that occur after $\text{Fe}(\text{CO})_5$ has been photoexcited. A more recent study showed that $^3\text{Fe}(\text{CO})_3$ remains uncoordinated

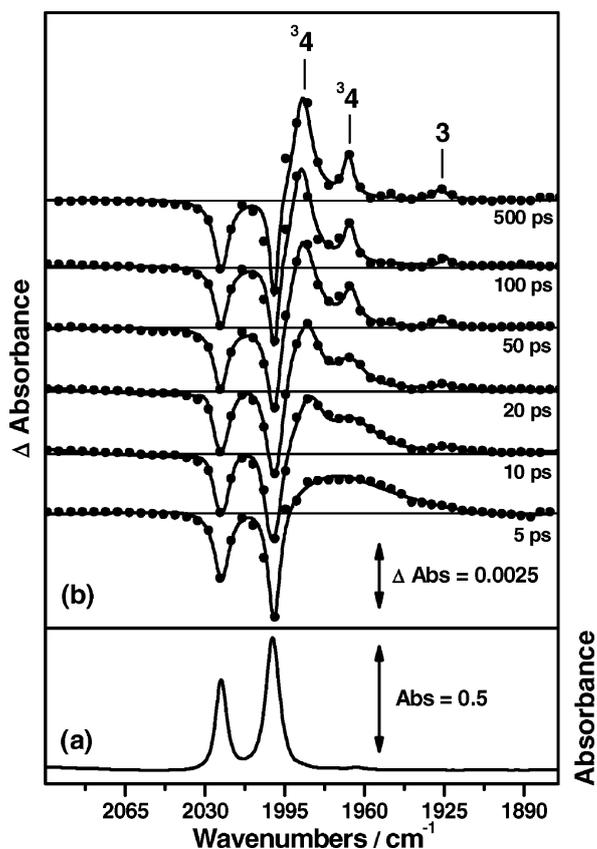


Fig. 2.7 TRIR spectra of $\text{Fe}(\text{CO})_5$ in heptane excited at 266 nm. Two negative bands are observed for $\text{Fe}(\text{CO})_5$, two large positive transient bands at 1988 and 1967 cm^{-1} are assigned to ${}^3\text{Fe}(\text{CO})_4$ and the small transient band at 1926 cm^{-1} is assigned to ${}^3\text{Fe}(\text{CO})_3$. Reproduced from [50].

in alkanes but coordinates to the solvent if it is an alcohol on the ps timescale. It was also shown that the triplet spin state is preserved upon coordination of the solvent.⁷⁰

A recent study used resonant inelastic X-ray scattering (RIXS) to investigate the photodissociation of $\text{Fe}(\text{CO})_5$ in ethanol at 266 nm. Only $\text{Fe}(\text{CO})_4$ was observed as a photoproduct, no $\text{Fe}(\text{CO})_3$ was observed in the spectra up to a time limit of a few picoseconds. The authors attributed this to not enough $\text{Fe}(\text{CO})_3$ having formed in solution at that stage. This, however, is unlikely considering TRIR experiments with $\text{Fe}(\text{CO})_5$ in alcohol have shown broad peaks indicating $\text{Fe}(\text{CO})_3$ has already formed at these short time timescales.^{70,71}

2.1.3 2D-IR study of Fe(CO)₅

The fluxional rearrangement of Fe(CO)₅ in dodecane was investigated using 2D-IR spectroscopy.⁷² The exchange of axial and equatorial CO ligands was probed by collecting spectra with a range of waiting times between IR_{pump} and IR_{probe} pulses and at a variety of temperatures. This allowed the exchange of vibrational energy between modes to be probed by viewing the off-diagonal peaks in particular. A small energy barrier was found between the exchange of vibrational energy, furthermore the rate of this exchange was found to increase by 78% over the temperature range. If this vibrational energy exchange was due to IVR then only a smaller increase of 26% would be expected. Therefore the small energy barrier is consistent with a pseudo rotation mechanism (and not IVR) where the axial and equatorial ligands are exchanged. This occurs on the timescale of 8 ps at room temperature.⁷²

2.1.4 Aims

The aim of the work in this chapter was to control the photodissociation of Fe(CO)₅ in solution using a vibrational excitation. More specifically, can we vibrationally excite Fe(CO)₅ in solution and would that result in a change in the relative amounts of Fe(CO)₄ and Fe(CO)₃ formed as a result of the UV excitation at 266 nm? According to previous literature,^{50,64} it was hypothesised that Fe(CO)₃ is formed via a vibrationally excited Fe(CO)₄ state, i.e. via sequential CO loss. It was thought that the yield of Fe(CO)₃ could be enhanced by vibrationally exciting the Fe(CO)₄ molecules immediately after the UV photoexcitation.

Photodissociation of ligands occurs on the ultrafast timescale (*ca.* fs-ps)⁷³ and therefore the excitation and measurement of these experiments must be performed on a similar timescale ($\text{Fe}(\text{CO})_3$ is stable up to the nanosecond timescale).⁵⁰ Therefore, T-2D-IR spectroscopy was the ideal technique to perform this experiment with. It involves an UV and IR (vibrational) excitation pulse, as well as an IR probe pulse, which was sufficient to detect and distinguish all the photoproducts on the femtosecond timescale.

2.2 Results

This section presents TRIR, 2D-IR and T-2D-IR spectra obtained for the photodissociation of $\text{Fe}(\text{CO})_5$ in heptane and CH_2Cl_2 . 266 nm UV and IR pump pulses (at a range of frequencies) were used to excite the sample, and an additional IR probe pulse was used to measure the photoproducts. The IR probe pulse was typically *ca.* 100 ps after the UV excitation, which was a time delay that ensured the spectral peaks were well resolved and not vibrationally 'hot'.⁵⁰ The T-2D-IR spectra were obtained using the double resonance technique as opposed to the pulse-shaping method.⁴ Two different types of plots are displayed below: contour maps which display a T-2D-IR spectrum across probe and pumps axes and spectral slices at a set IR_{pump} frequency, both for a given time delay.

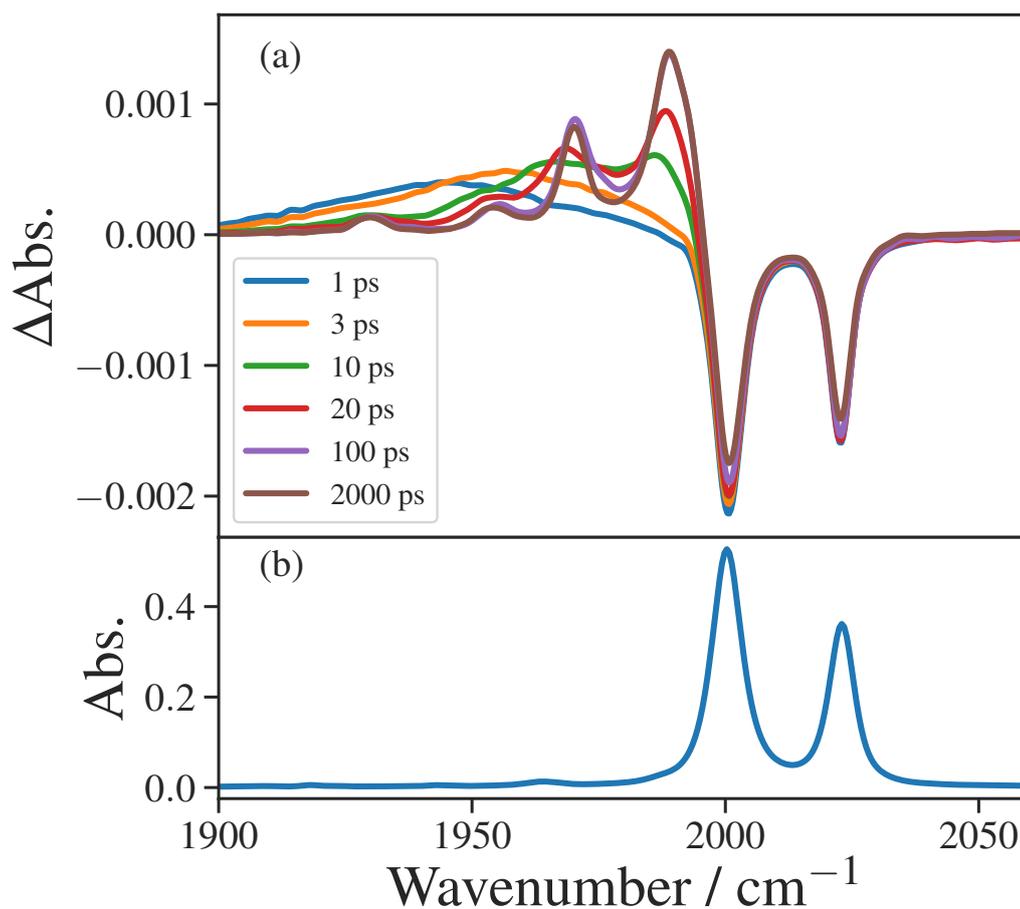


Fig. 2.8 (a) TRIR spectra of $\text{Fe}(\text{CO})_5$ in heptane excited at 266 nm at various delay times showing the vibrational cooling. There are two parent bleaches at 2000 and 2020 cm^{-1} representing $\text{Fe}(\text{CO})_5$, two large transient peaks at 1985 and 1965 cm^{-1} for the formation of $\text{Fe}(\text{CO})_4$ and a small transient peak at 1925 cm^{-1} indicating $\text{Fe}(\text{CO})_3$. (b) The FTIR spectrum for the same sample.

2.2.1 TRIR

Figure 2.8 shows TRIR spectra of $\text{Fe}(\text{CO})_5$ in heptane, excited at 266 nm, at a range of delay times. The spectrum at 100 ps shows two large negative peaks (parent bleaches) at 2000 and 2020 cm^{-1} , these indicate the excited $\text{Fe}(\text{CO})_5$ molecules and match the peaks in the FTIR spectrum. The two large positive transient peaks at slightly lower wavenumber (*ca.* 1985 and 1965 cm^{-1}) indicate the formation of $\text{Fe}(\text{CO})_4$. There is a single peak at *ca.* 1925 cm^{-1} which also shows that $\text{Fe}(\text{CO})_3$ is formed. At early time delays these well resolved transient peaks are instead very broad and this is due to the photoproducts being vibrationally ‘hot’. As the time delays increase from 1 ps to 100 ps the excess vibrational energy in the molecules are dissipated by solution and result in well-resolved peaks. This process of vibrational cooling occurs on a timescale of *ca.* 10 ps. At 2 ns all three peaks are still present and have not reduced dramatically in intensity, meaning that both these photoproducts are stable on the nanosecond timescale. These results are consistent with those of Portius *et al.*⁵⁰

The TRIR spectra in Figure 2.8 were fitted using pseudo-Voigt peaks. The full width at half maximum (FWHM) was plotted against delay time for one of the fitted peaks and a single exponential function was fitted to it, as shown in Figure 2.9. The fit resulted in a decay time of 8.2 ± 1.5 ps which is a similar timescale to that stated by Portius *et al.*⁵⁰

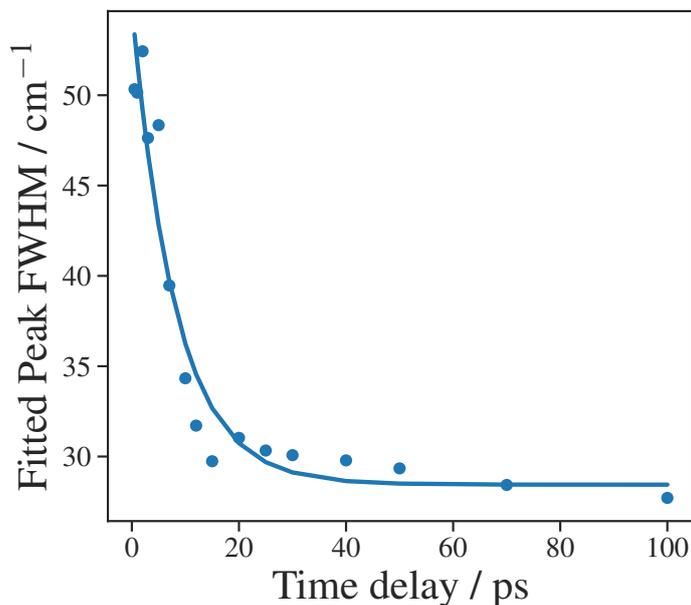


Fig. 2.9 A plot of the FWHM of a fitted peak to TRIR spectra versus time delay, showing vibrational cooling. A single exponential function was fitted (line) with a lifetime of 8.2 ± 1.5 ps.

2.2.2 Ground state 2D-IR

Frequency-domain 2D-IR spectroscopy

The ground state 2D-IR spectrum of $\text{Fe}(\text{CO})_5$ in heptane is shown in Figure 2.10. It was collected using the double resonance technique using a waiting time (T_W) of 100 ps. Two parent bleaches can be observed (blue) from the fundamental transition and the first two hot bands can be observed at a slightly lower wavenumber (red). This 2D-IR spectrum is consistent with those obtained of $\text{Fe}(\text{CO})_5$ in dodecane at long waiting times.⁷⁴ Each peak can be seen to extend along the ω_{pump} axis, this is partially due to the lineshape ‘leakage’ from the pump pulse ($\text{FWHM} = ca. 28 \text{ cm}^{-1}$) but is mostly due to fluxional rearrangement of the ligands, arising from and assigned to be due to Berry psuedo-rotation, which results in

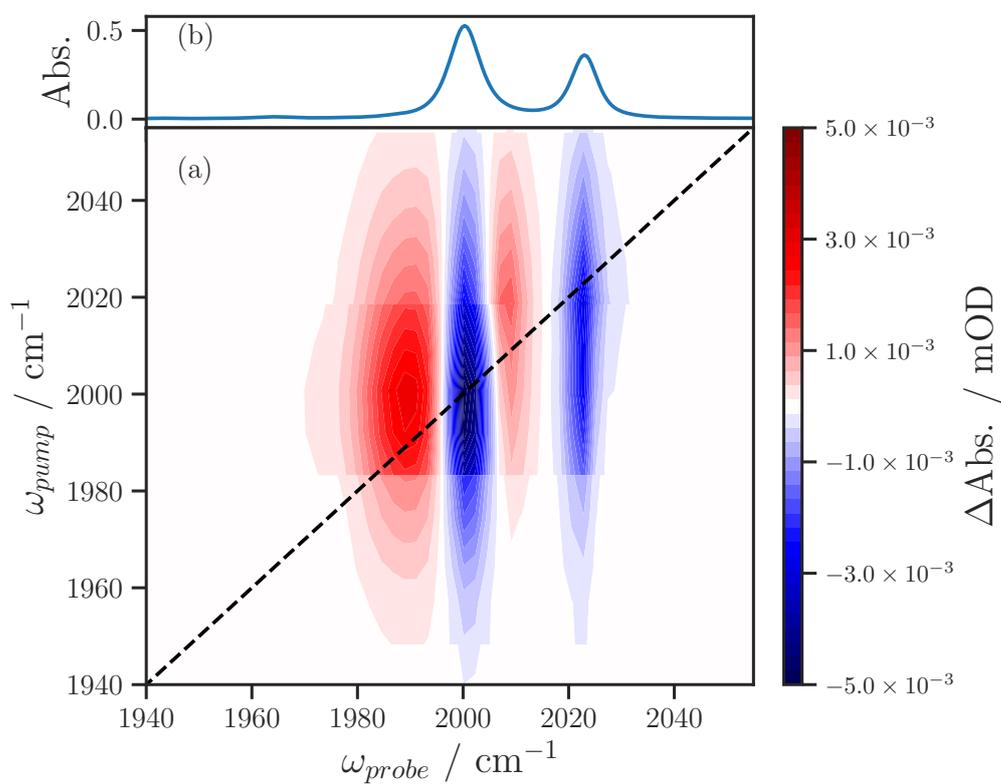


Fig. 2.10 (a) The ground state 2D-IR spectrum of Fe(CO)₅ at $T_W = 100$ ps, blue and red contours represent positive and negative changes in absorbances respectively. (b) The FTIR spectrum for the same sample.

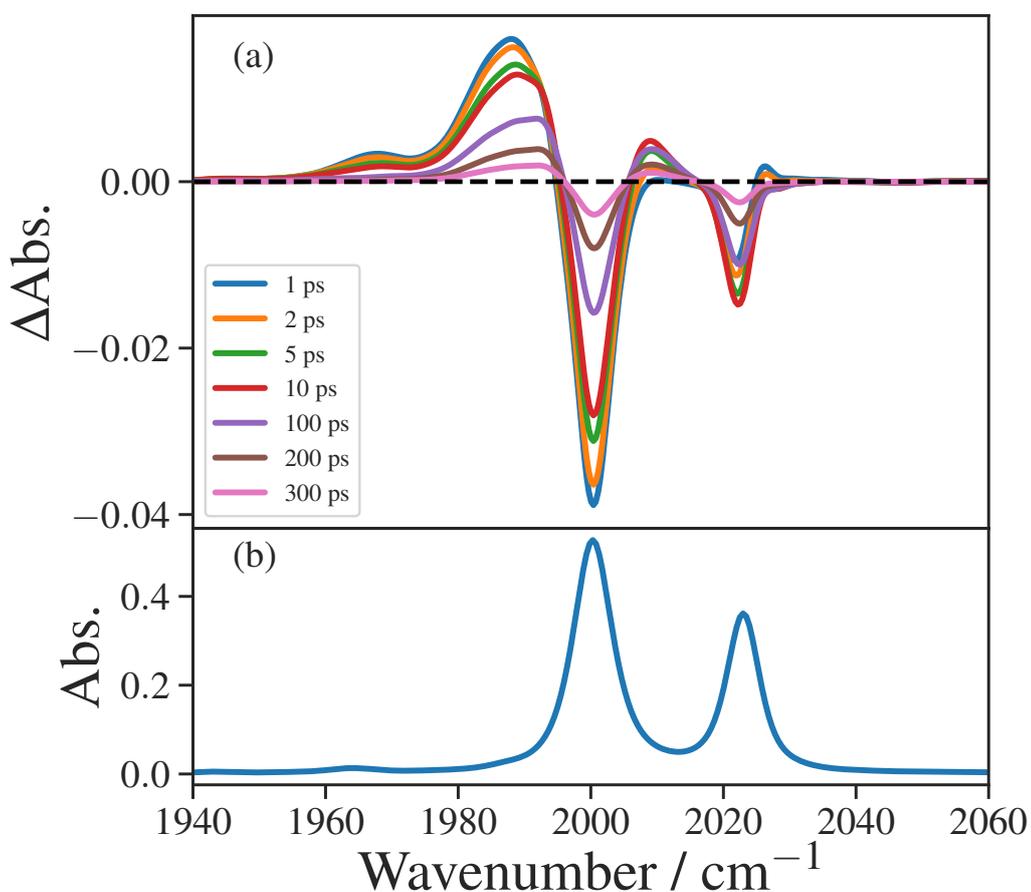


Fig. 2.11 (a) 2D-IR pump slice spectra of $\text{Fe}(\text{CO})_5$ in heptane vibrationaly excited at 2000 cm^{-1} and at various waiting times showing the vibrational energy relaxation. There are two parent bleaches at 2000 and 2020 cm^{-1} representing the fundamental modes of $\text{Fe}(\text{CO})_5$, one large transient peak at 185 cm^{-1} indicates the hot band of $\text{Fe}(\text{CO})_5$. (b) The FTIR spectrum for the same sample.

cross peaks. Here the vibrational modes become mixed, i.e. one mode is pumped, there is then a pseudo rotation of the ligands and the vibrational energy is subsequently observed in both states.

2D-IR pump-slice spectra are shown in Figure 2.11 at various waiting times for a vibrational excitation at 2000 cm^{-1} . There are two negative parent bleaches for the fundamental transitions of $\text{Fe}(\text{CO})_5$ (matching the FTIR spectrum). The peak at 2000 cm^{-1} (\mathbf{e}') is a

doubly degenerate mode consisting of out of phase stretching of the equatorial ligands. The peak at 2023 cm^{-1} (a_2'') is due to the out of phase stretch of the two axial ligands.⁷² There would normally be two positive transient peaks at lower frequency, however, only one is present. The main reason being, that the hot band for the higher frequency mode overlaps with the negative parent bleach of the lower energy fundamental mode, this results in the intensities almost cancelling out. There is another minor reason, the IR excitation mainly excites the lower energy mode (this also explains the intensity difference between the two parent bleaches) and higher energy state is only populated by Berry pseudo-rotation over time.

There is a small transient peak at *ca.* 1965 cm^{-1} which is unexpected. It could be due to the ^{13}C isotope, however, the intensity ratio compared to the main peak is too small for this to be the case. Alternatively, the $\nu_{2\rightarrow 3}$ transition is another option but the frequency is too high for that to make sense. Another option is the formation of bridged complexes but the frequencies of the peaks do not match up again.⁵⁰ The kinetics of the peak are reflected well by the global fit and are similar to the main transient peak at 1987 cm^{-1} . This spectrum was collected towards the end of a series of experiments. It is therefore assumed to be a small impurity present in the sample which is also being excited by the vibrational pump pulse.

The kinetics of the decay and growth of peaks in Figure 2.11 are described well by two timescales and were globally fitted (Figure 2.12). The first timescale, $3.0 \pm 0.1\text{ ps}$, represents population transfer between vibrational modes (pseudo-rotation process), the second timescale of $185 \pm 9\text{ ps}$ is due to vibrational population relaxation. The peaks at 1987 and

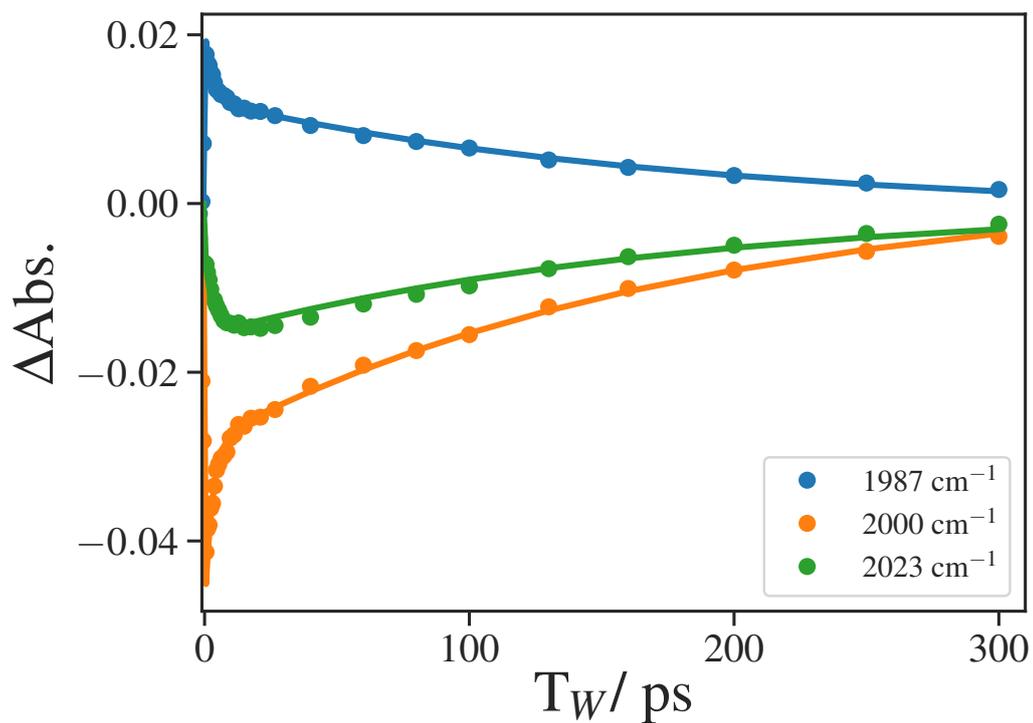


Fig. 2.12 A plot of the maximum peak intensities at 1987 , 2000 and 2023 cm^{-1} versus waiting time showing the vibrational energy relaxation occurring. Globally fitted exponential decay curves are displayed and were fitted with two lifetimes of $3.0 \pm 0.1 \text{ ps}$ and $185 \pm 9 \text{ ps}$, respectively.

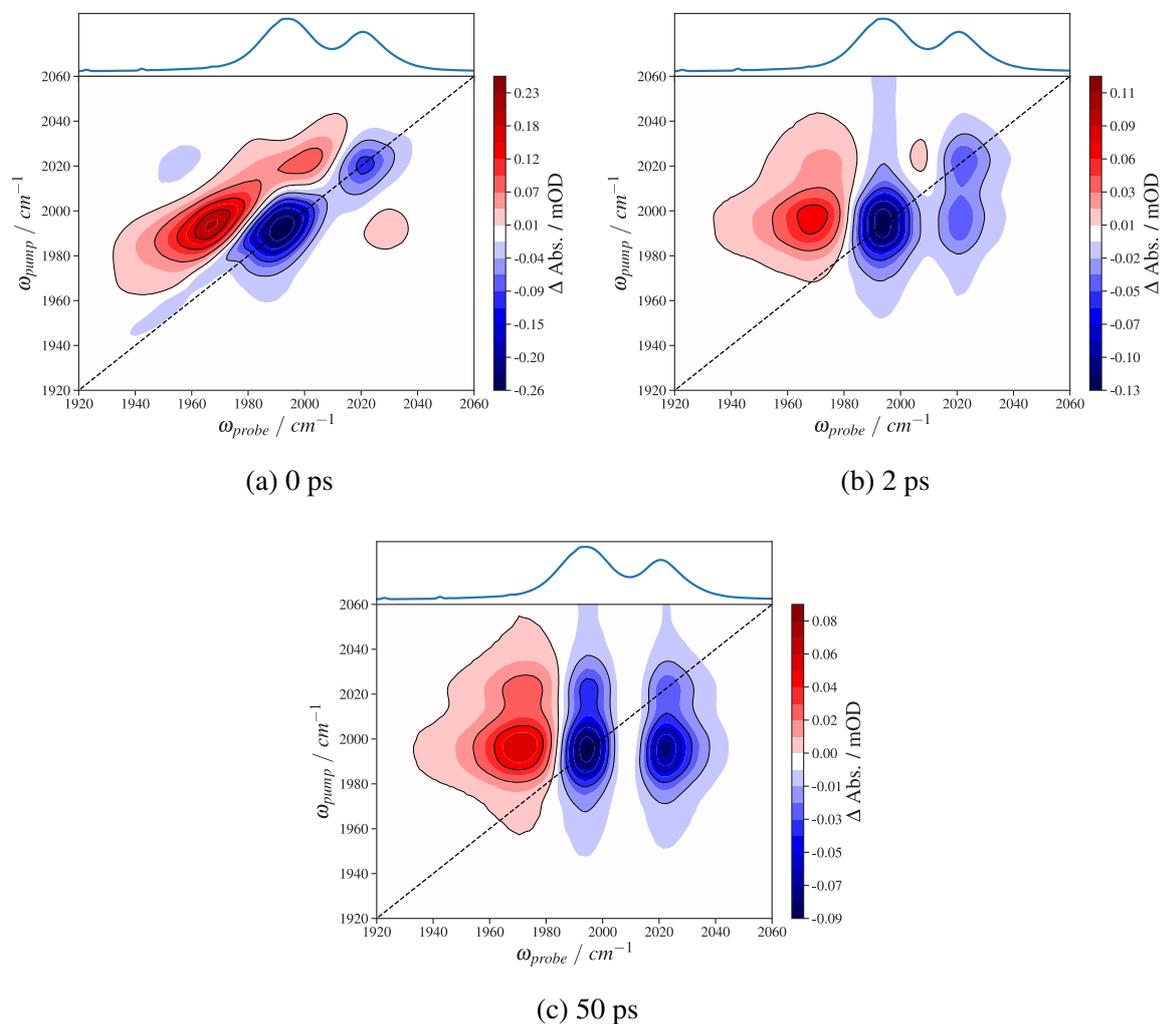


Fig. 2.13 2D-IR spectra of $\text{Fe}(\text{CO})_5$ in CH_2Cl_2 at different waiting times, showing cross peaks occurring as the waiting time is increased.

2000 cm^{-1} show decay characteristics for both timescales whereas the peak at 2023 cm^{-1} shows an initial growth due to population transfer from the excited \mathbf{e}' mode.

Time-domain 2D-IR spectroscopy

Unfortunately ground state 2D-IR spectra of $\text{Fe}(\text{CO})_5$ in CH_2Cl_2 were not collected at the time. Instead, spectra were obtained in Nottingham using the time domain technique. Ground

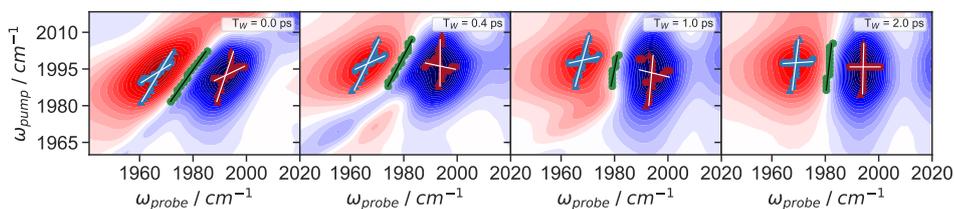


Fig. 2.14 Spectral diffusion of 2D-IR peaks at various waiting times. The central line slope of each peak is measured and the NLS is calculated from this (plotted in green). This allows the rate of spectral diffusion to be quantified.

state 2D-IR spectra of $\text{Fe}(\text{CO})_5$ in CH_2Cl_2 at 0, 2 and 50 ps which show the cross peaks forming as the waiting time is increased are shown in Figure 2.13. The line shapes of the peaks are noticeably different to the previous spectra obtained in the frequency domain and the time resolution is much better (*ca.* 300 fs). Because of this, the spectral diffusion was then calculated for the spectra at early waiting times which is shown in Figure 2.14. The nodal line shape (NLS) method was found to be the best method for this set of data (CLS (triangles) and ivCLS (circles) were also calculated for each peak and are also shown).⁷⁵ The decay of the NLS gradient was then fitted with a single exponential function, resulting in a lifetime of 0.8 ± 0.2 ps (Figure 2.15). This is equivalent to the frequency-frequency correlation function (FFCF) and is a measure of the inhomogeneous spectral broadening as a result of solvent fluctuations.

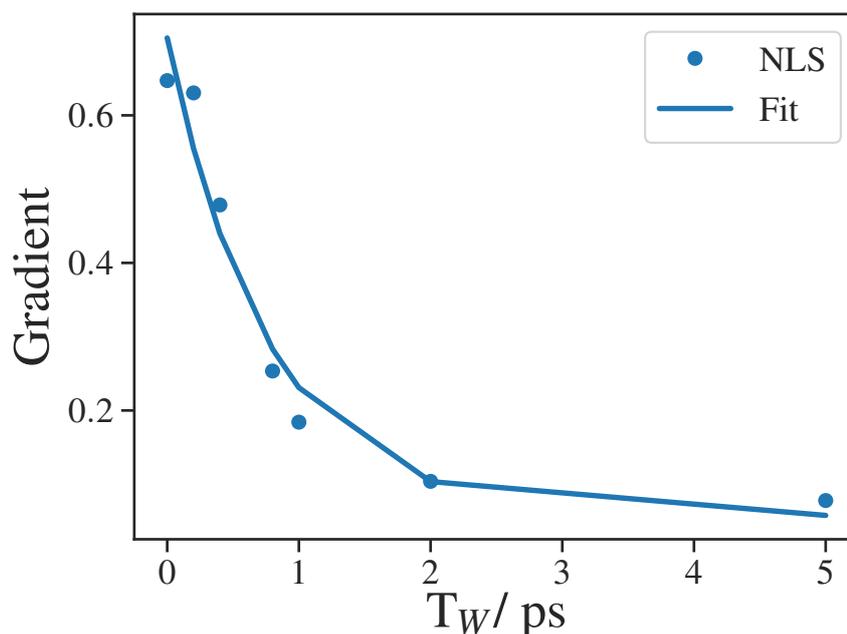


Fig. 2.15 Gradient of the NLS versus waiting time showing the spectral diffusion of the peak at *ca.* 1990, 1980 cm^{-1} for $\text{Fe}(\text{CO})_5$ in CH_2Cl_2 . An exponential decay curve has been fitted giving a lifetime of 0.8 ± 0.2 ps.

2.2.3 T-2D-IR spectroscopy

Heptane

Figure 2.16 shows three different types of difference spectra ((b) TRIR, (c) 2D-IR, (d,e) T-2D-IR) obtained for $\text{Fe}(\text{CO})_5$ in heptane. The TRIR spectrum ($T_{UV} = 95$ ps) shows two parent bleaches where the ground state carbonyl stretches were, together with three transient bands at lower wavenumber. The first two transient bands at 1985 cm^{-1} and 1965 cm^{-1} are from the formation of $\text{Fe}(\text{CO})_4$ after UV excitation and the single peak at 1925 cm^{-1} represents the formation of $\text{Fe}(\text{CO})_3$. A second peak would be expected for $\text{Fe}(\text{CO})_3$, however, it is thought to overlap with a large parent band and the intensity is also assumed to be weak. This is consistent with the TRIR spectra from Portius *et al.*⁵⁰ The 2D-IR spectral slice ($\text{IR}_{\text{pump}} =$

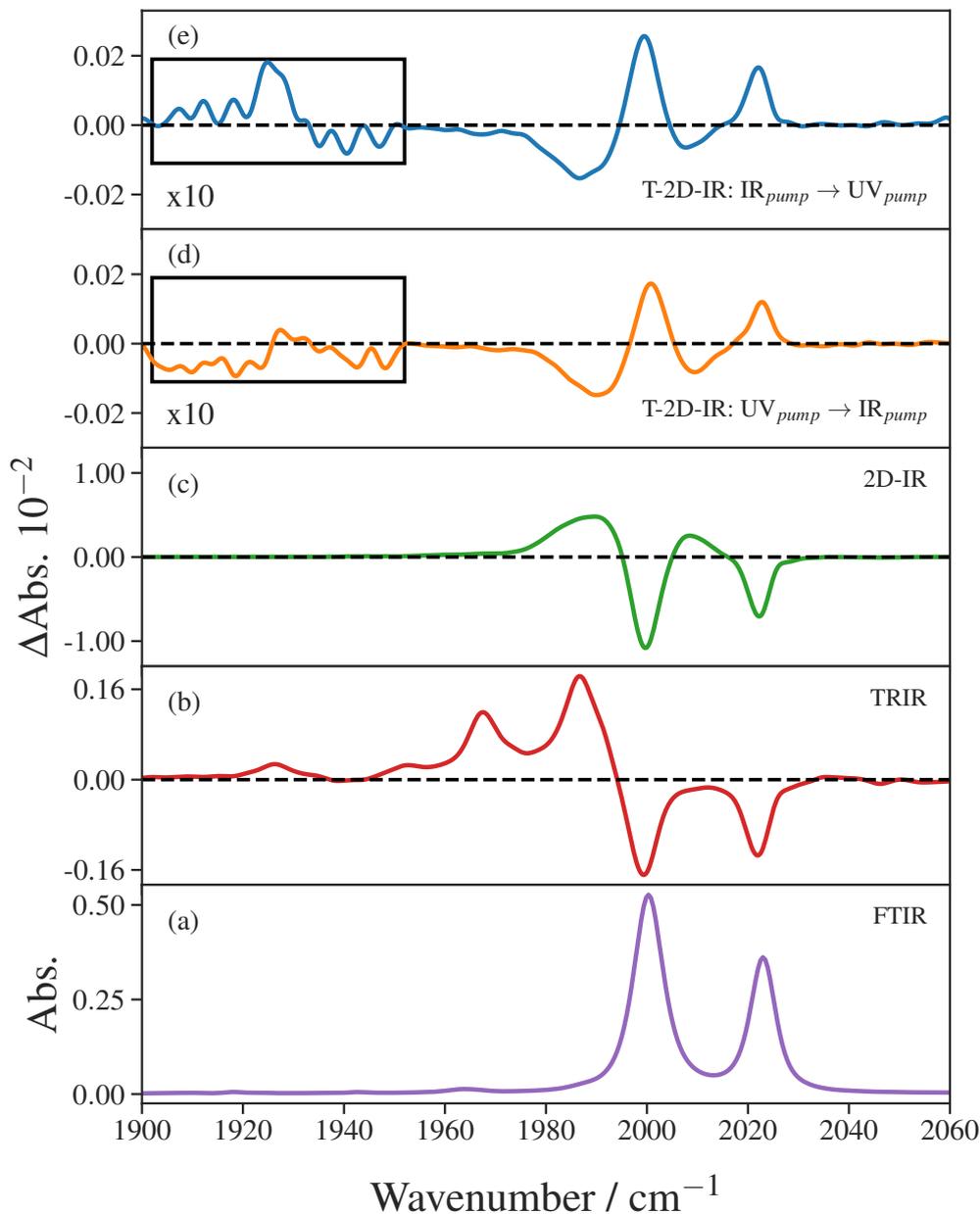


Fig. 2.16 (a) FTIR spectrum of $\text{Fe}(\text{CO})_5$ in heptane. (b) TRIR difference spectrum of $\text{Fe}(\text{CO})_5$ after it has been excited by a UV pulse, showing transient peaks at 1985 cm^{-1} and 1965 cm^{-1} assigned to $\text{Fe}(\text{CO})_4$ and a single peak at 1925 cm^{-1} from $\text{Fe}(\text{CO})_3$. (c) 2D-IR difference spectral slice showing negative bands for ω_{01} and positive bands for ω_{12} for an $\text{IR}_{\text{pump}} = 2000 \text{ cm}^{-1}$. (d) and (e) T-2D-IR spectral slice where any peaks present in the spectrum are directly due to the IR pump pulse when compared with the TRIR spectrum. For (e), a small positive peak is observed at 1925 cm^{-1} , showing an increase in the amount of $\text{Fe}(\text{CO})_3$ formed when the IR pump (2000 cm^{-1}) is first incident on the sample followed by the UV pump 5 ps later and finally the IR probe 100 ps later. This is not observed in (d) where IR_{pump} is directly after the UV_{pump} .

2000 cm^{-1}) again shows two bleaches where the ground state vibrational modes are located; there are also two positive peaks shifted to lower wavenumber (probed at $T_W = 100$ ps) – these represent the first hot transitions (ω_{12}). Finally for the double difference T-2D-IR spectral slices, any peaks observed in this spectrum are as a direct result of the IR pump pulse and are compared with the TRIR spectrum. A small positive peak is evident at 1925cm^{-1} , demonstrating that the IR excitation, at 2000 cm^{-1} , 5 ps before the UV pump pulse (and an IR probe after 100 ps) results in an increase in the amount of $\text{Fe}(\text{CO})_3$. There is also a small negative peak observed at 1965 cm^{-1} , which indicates less $\text{Fe}(\text{CO})_4$ is being formed as a result of the IR pump pulse at this time delay. However, most of the $\text{Fe}(\text{CO})_4$ region in the spectrum is obscured by the inverted ground state 2D-IR signal generated from $\text{Fe}(\text{CO})_5$ and prevents the observation of clearer evidence. No peak at 1925 cm^{-1} was observed when the order of the pump pulses was reversed ($\text{UV}_{pump} \rightarrow \text{IR}_{pump}$) indicating that the vibrational excitation after the UV excitation had no effect on the resultant photoproducts.

The T-2D-IR spectrum of $\text{Fe}(\text{CO})_5$ in heptane is shown in Figure 2.17, this was obtained using the following pulse sequence: $\text{IR}_{pump} \rightarrow 2\text{ ps} \rightarrow \text{UV}_{pump} \rightarrow 98\text{ ps} \rightarrow \text{IR}_{probe}$, where the IR_{pump} frequency was scanned across the two parent bands. The important feature of this spectrum is the small positive band located at 1925 cm^{-1} in the probe axis. This represents the increase in the formation of $\text{Fe}(\text{CO})_3$ when the IR_{pump} frequency is between *ca.* 1990 and 2025 cm^{-1} . The positive peak occurs across the IR_{pump} frequency range, suggesting that the increase in $\text{Fe}(\text{CO})_3$ is not dependent on which ground state vibrational mode of $\text{Fe}(\text{CO})_5$ is excited first.

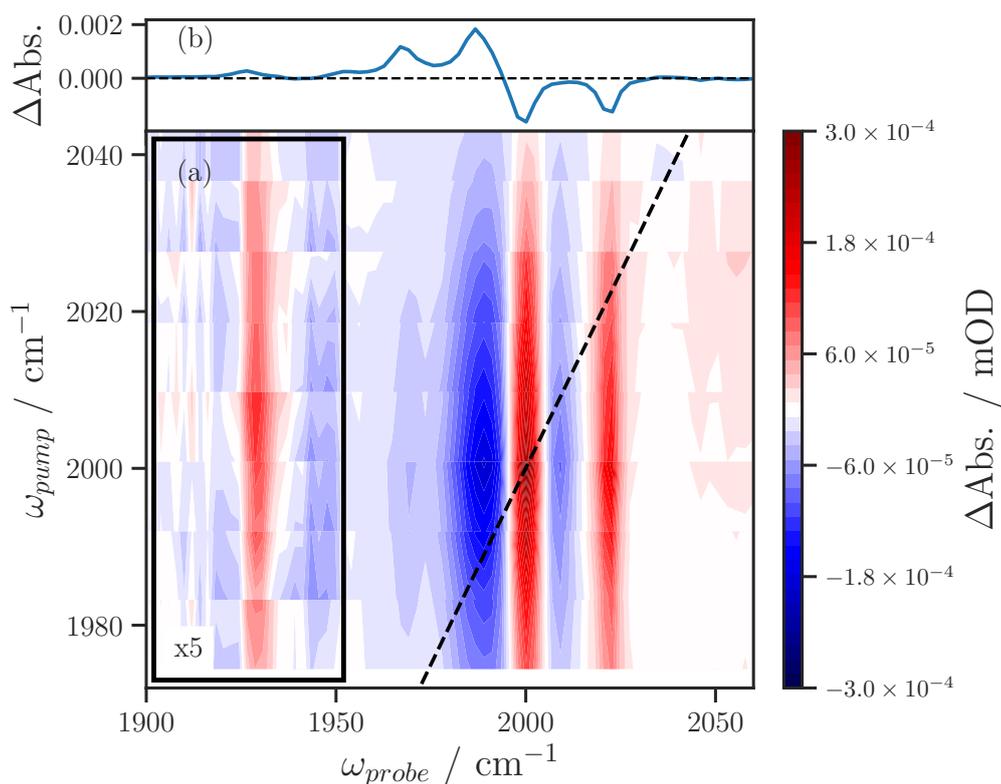


Fig. 2.17 (a) T-2D-IR spectrum of $\text{Fe}(\text{CO})_5$ in heptane obtained with a 2 ps delay time between IR and UV pumps and a waiting time of 100 ps between the initial pump and probe pulses. The $\text{Fe}(\text{CO})_3$ region of the spectrum ($1915\text{-}1945 \text{ cm}^{-1}$) has been enhanced by a factor of 10 to highlight the peaks observed. A small increase (red) can be observed at 1925 cm^{-1} which shows that there is an increase in $\text{Fe}(\text{CO})_3$ formed when the sample is excited using this pulse sequence. (b) TRIR spectrum of $\text{Fe}(\text{CO})_5$ at a 98 ps time delay.

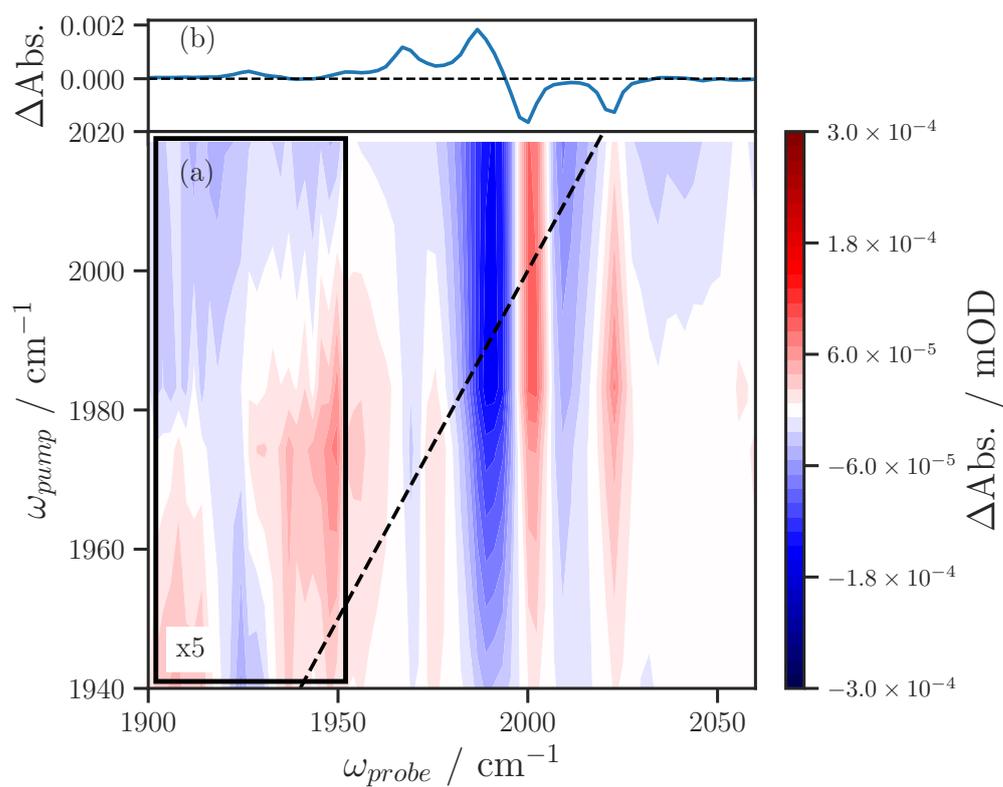


Fig. 2.18 (a) T-2D-IR spectrum of $\text{Fe}(\text{CO})_5$ in heptane with the IR_{pump} pulse incident immediately after the UV pump pulse (0 ps) and IR_{probe} pulse incident after 100 ps. No peak is observed at 1925 cm^{-1} , indicating there is no change in the amount of $\text{Fe}(\text{CO})_3$ formed. (b) TRIR spectrum of $\text{Fe}(\text{CO})_5$ at a 100 ps time delay.

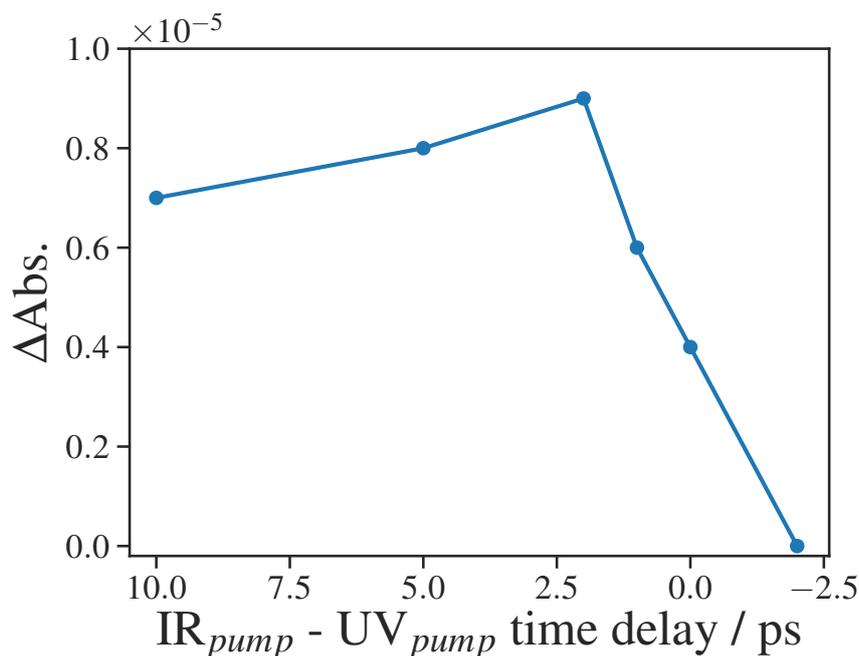


Fig. 2.19 A plot of the intensity of the $\text{Fe}(\text{CO})_3$ peak at 1925 cm^{-1} from T-2D-IR spectral slices for an IR excitation at 2000 cm^{-1} versus the delay time between the two excitation pulses. A positive delay time indicates an IR before UV excitation and the order swaps when the delay time becomes negative. The large decrease in peak intensity from 2 to -2 ps shows that the amount of $\text{Fe}(\text{CO})_3$ formed only changes when the IR pump pulse is incident before the UV pump pulse.

Figure 2.18 shows the T-2D-IR spectrum for exciting the $\text{Fe}(\text{CO})_5$ in heptane with an IR_{pump} pulse immediately after the UV photo-excitation and probing the sample 100 ps after the initial excitations. From the zoomed in $\text{Fe}(\text{CO})_3$ region (1925 cm^{-1}), there is no obvious peak observed at any IR_{pump} frequency, whether it be $\text{Fe}(\text{CO})_5$ ground state vibrational modes or vibrationally hot $\text{Fe}(\text{CO})_4$ modes. This indicates that an IR_{pump} pulse directly after the UV excitation results in no change in the amount of $\text{Fe}(\text{CO})_3$ formed.

Figure 2.19 shows how the intensity of the $\text{Fe}(\text{CO})_3$ peak (for T-2D-IR spectral slices at 2000 cm^{-1}) varies against the delay time between the IR and UV excitations. The intensity of the peak increases as the delay time between IR and UV excitations decreases up to 2

ps. Then between 2 ps and -2 ps (when the order of the excitations swap i.e. UV then IR excitation) the intensity decreases substantially to 0, indicating no change in the amount of $\text{Fe}(\text{CO})_3$ formed. This is consistent with the T-2D-IR spectrum shown in Figure 2.18 and clearly shows that IR_{pump} before the UV_{pump} pulse results in more $\text{Fe}(\text{CO})_3$ being formed. Whereas when order is switched and the UV_{pump} pulse is immediately before the IR_{pump} pulse, then no change in the amount of $\text{Fe}(\text{CO})_3$ is observed.

Dichloromethane

Figure 2.20 shows the FTIR and various difference spectra of $\text{Fe}(\text{CO})_5$ in CH_2Cl_2 . The first observation is that the peaks are broader and have been shifted to a slightly lower wavenumber. These broader spectral features make it harder to resolve, especially transient ones for example in the TRIR spectra. The TRIR and 2D-IR spectra were obtained with a delay time of 98 and 100 ps, respectively. The TRIR spectrum displays two negative $\text{Fe}(\text{CO})_5$ parent bands at 1995 and 2020 cm^{-1} , two positive transient peaks for $\text{Fe}(\text{CO})_4$ at 1975 and 1955 cm^{-1} and a single positive transient peak at 1920 cm^{-1} for the formation of $\text{Fe}(\text{CO})_3$. The T-2D-IR spectral slice was obtained with the following pulse sequence: IR_{pump} (1990 cm^{-1}) \rightarrow 2 ps \rightarrow UV_{pump} \rightarrow 98 ps \rightarrow IR_{probe} . A small negative band can be observed at 1920 cm^{-1} , representing a decrease in the amount of $\text{Fe}(\text{CO})_3$ formed because of the IR_{pump} pulse. Unfortunately, no information can be determined about the amount of $\text{Fe}(\text{CO})_4$ formed as the broad bands from the inverted $\text{Fe}(\text{CO})_5$ 2D-IR signal obscures this.

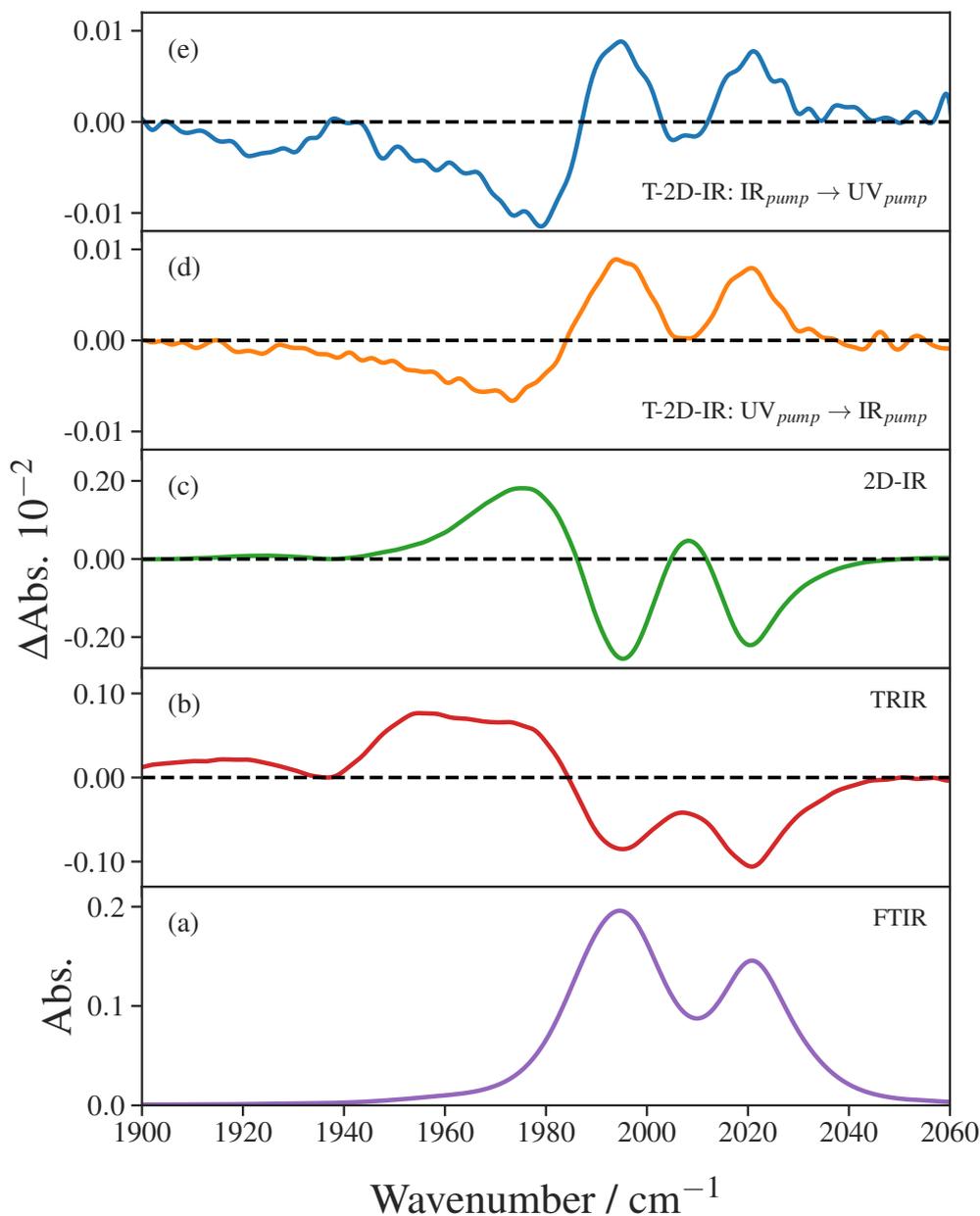


Fig. 2.20 (a) FTIR spectrum of $\text{Fe}(\text{CO})_5$ in CH_2Cl_2 . (b) TRIR difference spectrum of $\text{Fe}(\text{CO})_5$ after it has been excited by a UV pulse (266 nm), showing transient peaks at 1970 cm^{-1} and 1955 cm^{-1} assigned to $\text{Fe}(\text{CO})_4$ and a single peak at 1920 cm^{-1} from $\text{Fe}(\text{CO})_3$. (c) 2D-IR spectral slice ($\text{IR}_{\text{pump}} = 1990\text{ cm}^{-1}$) showing negative bands for ω_{01} and positive bands for ω_{12} transitions. (d) and (e) T-2D-IR spectral slice which shows the changes that occur because of the IR pump pulse when compared with the TRIR spectrum. For (e), a small negative peak is observed at 1920 cm^{-1} , indicating a decrease in the amount of $\text{Fe}(\text{CO})_3$ formed when the IR pump pulse (1990 cm^{-1}) is first incident on the sample followed by the UV pump 2 ps later and finally the IR probe 98 ps after that. For (d), no negative peak is observed when the order of the pump pulses are reversed to $\text{UV}_{\text{pump}} \rightarrow \text{IR}_{\text{pump}}$.

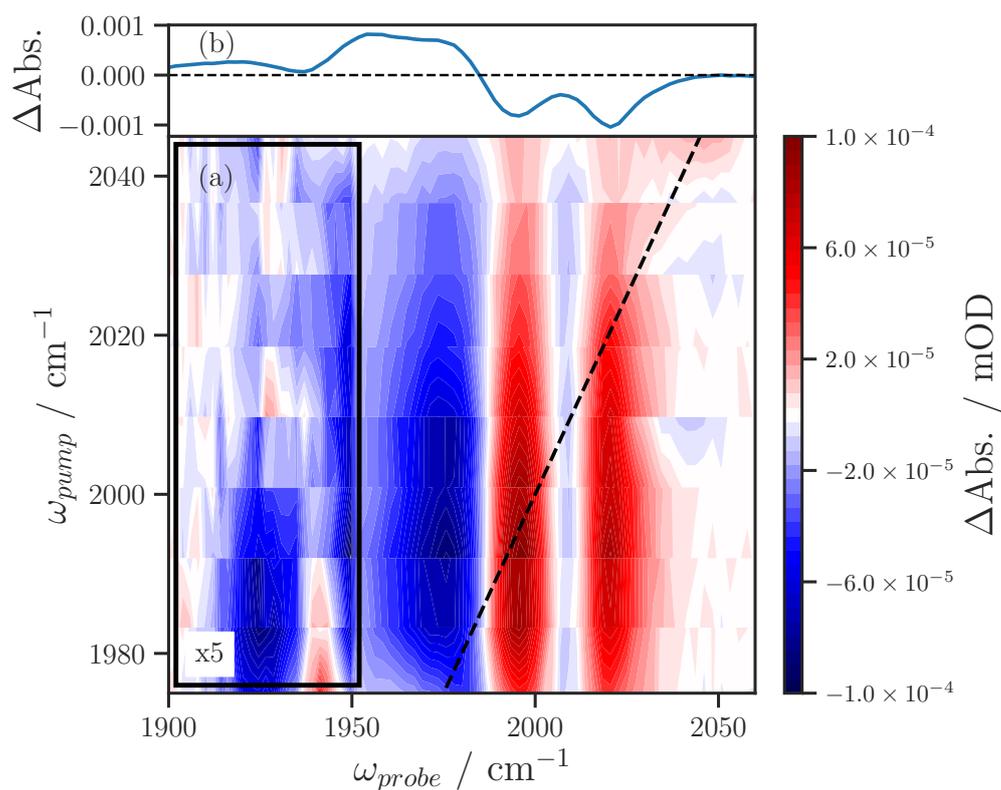


Fig. 2.21 T-2D-IR spectrum of $\text{Fe}(\text{CO})_5$ in CH_2Cl_2 obtained with a 2 ps delay time between IR and UV pumps and waiting time of 100 ps between the initial pump and probe pulses. A small decrease (blue) can be observed at 1920 cm^{-1} on the ω_{probe} axis and between 1975 and 2000 cm^{-1} on the ω_{pump} axis. This shows that there is a decrease in $\text{Fe}(\text{CO})_3$ formed when the sample is excited using this sequence.

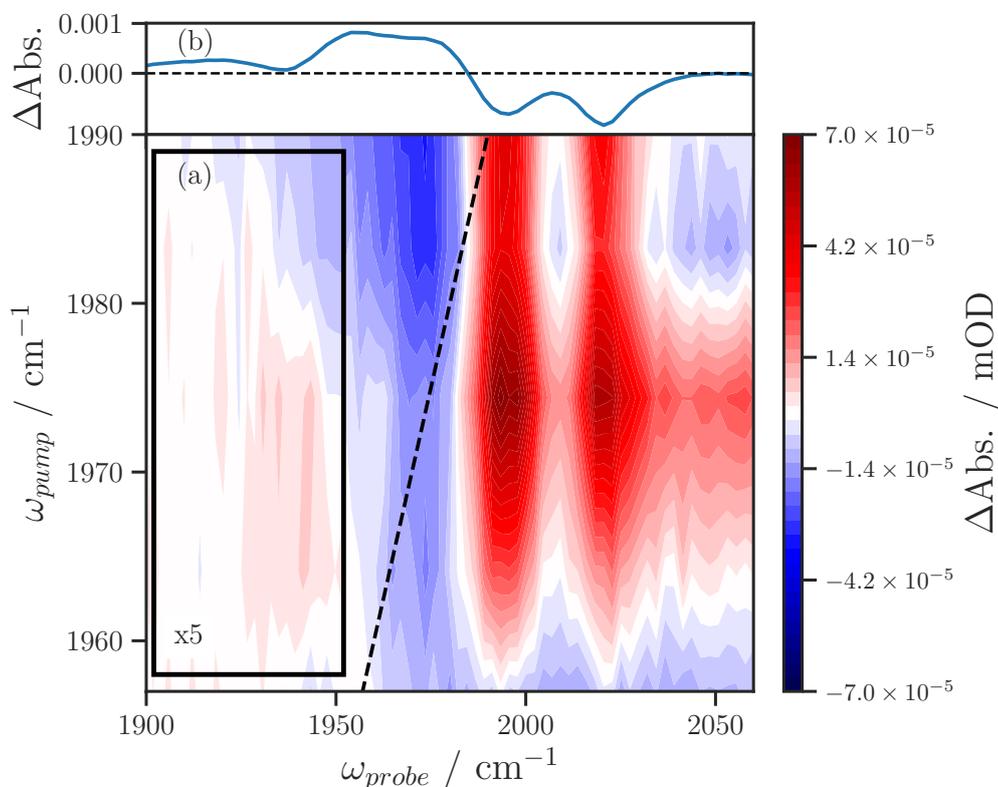


Fig. 2.22 T-2D-IR spectrum of $\text{Fe}(\text{CO})_5$ in CH_2Cl_2 with the IR_{pump} pulse incident immediately after the UV pump pulse (0 ps) and IR_{probe} pulse incident after 100 ps. No peak is observed at 1920 cm^{-1} , indicating there is no change in the amount of $\text{Fe}(\text{CO})_3$ formed.

The T-2D-IR spectrum for $\text{Fe}(\text{CO})_5$ in CH_2Cl_2 is shown in Figure 2.21, this was obtained with an initial IR_{pump} pulse followed 2 ps later by a UV excitation pulse and probing the sample 98 ps after that. The IR_{pump} frequency was scanned across the range of the ground state vibrational modes of $\text{Fe}(\text{CO})_5$. A negative peak can be observed at 1920 cm^{-1} on the probe axis, this represents the decrease in $\text{Fe}(\text{CO})_3$ being formed, it is only present though in the pump axis range of $1975 - 2000 \text{ cm}^{-1}$. This suggests that the change in $\text{Fe}(\text{CO})_3$ is IR_{pump} frequency dependent (vibrational mode specific), i.e. exciting only the lower wavenumber CO stretch (e') results in a change in the amount of $\text{Fe}(\text{CO})_3$ formed.

Figure 2.22 shows the T-2D-IR spectrum for exciting the $\text{Fe}(\text{CO})_5$ in CH_2Cl_2 with an IR_{pump} pulse immediately after the UV excitation and probing the sample 100 ps after the initial excitation. From the $\text{Fe}(\text{CO})_3$ region ($\omega_{\text{probe}} = 1920 \text{ cm}^{-1}$), there is no obvious peak observed at any IR_{pump} frequency, whether it be the low frequency $\text{Fe}(\text{CO})_5$ ground state vibrational mode or vibrationally hot $\text{Fe}(\text{CO})_4$ modes. This means that an IR_{pump} pulse directly after the UV excitation results in no change in the amount of $\text{Fe}(\text{CO})_3$ formed.

2.3 Discussion

2.3.1 Pulse sequence

It has been observed that there are only changes in the amount of $\text{Fe}(\text{CO})_3$ formed when an IR_{pump} pulse is incident before a UV_{pump} pulse. It is thought that there is no change when the IR_{pump} is incident after the UV excitation, because the loss of CO ligands occurs on too fast a timescale for an IR_{pump} pulse with a temporal resolution of *ca.* 1 ps to perturb the process. Portius *et al.* previously hypothesised that $\text{Fe}(\text{CO})_3$ was formed from a vibrationally excited state of $\text{Fe}(\text{CO})_4$ and within 5ps of the UV excitation.⁷⁶ There are two possibilities for the formation of $\text{Fe}(\text{CO})_3$: it is formed directly from $\text{Fe}(\text{CO})_5$ with the simultaneous loss of two CO ligands or it is formed via vibrationally hot $\text{Fe}(\text{CO})_4$ with the loss of a CO ligand, i.e. a sequential loss of two ligands. While the sequential loss of the ligands cannot be ruled out entirely, it does not appear likely given vibrationally exciting $\text{Fe}(\text{CO})_4$ when it is first formed does not result in a change in $\text{Fe}(\text{CO})_3$ being formed – it is however possible that this

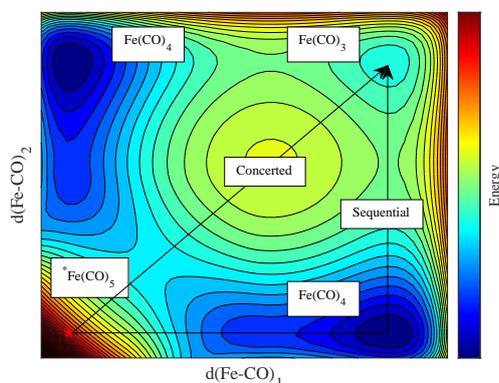


Fig. 2.23 A diagram showing a potential energy landscape for the dissociation of two CO ligands from an excited $\text{Fe}(\text{CO})_5$ dissociative state. There are two possible routes to $\text{Fe}(\text{CO})_3$, concerted and sequential, both represented by arrows.

process is too fast to perturb. A more likely scenario is the simultaneous loss of the ligands, and the evidence for the IR excitation after the UV one resulting in no change would support this. Furthermore, if there was sequential dissociation of the CO ligands, it is unlikely that a vibrational pre-excitation would result in a change in the amount of $\text{Fe}(\text{CO})_3$ formed. This would be more consistent with a simultaneous or concerted loss of ligands from $\text{Fe}(\text{CO})_5$ as the electronic excited state will be directly perturbed by the vibrational excitation. Both these processes are summarised by Figure 2.23.

The vibrational pre-excitation could result in a change in the molar absorptivity coefficient (i.e. a red-shift or blue-shift in the UV/Vis spectrum), however, this would not explain the results in heptane.⁴⁶ A change would only result in an increase in both $\text{Fe}(\text{CO})_3$ and $\text{Fe}(\text{CO})_4$, however, in heptane an increase in $\text{Fe}(\text{CO})_3$ and decrease in $\text{Fe}(\text{CO})_4$ was observed. In CH_2Cl_2 , a decrease in $\text{Fe}(\text{CO})_3$ was observed but no conclusion about $\text{Fe}(\text{CO})_4$ can be made

as the broad, inverted, 2D-IR signal obscures this. It was therefore assumed to be analogous to $\text{Fe}(\text{CO})_5$ in heptane.

2.3.2 Solvent dependence

A very interesting result that has arisen is the increase in the formation of $\text{Fe}(\text{CO})_3$ in heptane, but in CH_2Cl_2 a decrease of $\text{Fe}(\text{CO})_3$ is observed. At this stage it is unclear why this occurs, but it could relate to the polarity of solvents and solvatochromic effect that that has. It is also possible that the solvent cage is partially responsible which was thought to have an effect by Portius *et al.*⁷⁶ Both of these factors will affect and perturb the potential energy surface (PES) significantly. Experiments were attempted in perfluoro(methylcyclohexane) (PFMCH), unfortunately the small signals made it difficult to resolve any small peaks in the T-2D-IR spectrum. A higher concentration solution could allow the smaller peaks to be resolved and this would allow a trend in polarity to be assessed. It would also be interesting to investigate the effect much more polar solvents such as an alcohol have on the formation of $\text{Fe}(\text{CO})_3$.

2.3.3 Pump frequency dependence

The increase in formation of $\text{Fe}(\text{CO})_3$ when $\text{Fe}(\text{CO})_5$ is first vibrationally excited in heptane has been shown to not depend on the vibrational mode being excited. However, the formation of $\text{Fe}(\text{CO})_3$ appears to be vibrationally mode specific when $\text{Fe}(\text{CO})_5$ is solvated in CH_2Cl_2 . Only exciting the lower energy vibrational mode results in a change (e'), this suggests that

it is the equatorial CO ligands that dissociate. This could relate to solvent cage around the molecule and more specifically the effect the polarity of solvent has on dipole moment of the molecule.

For heptane, the T-2D-IR spectrum in Figure 2.16, shows that either vibrational mode can be excited and this results in a change of $\text{Fe}(\text{CO})_3$. As Cahoon *et al.* have shown, Berry-pseudorotation occurs for $\text{Fe}(\text{CO})_5$ and this is responsible for the vibrational energy transfer between modes.⁷⁴ It is possible that the \mathbf{a}_2'' mode is excited, the vibrational energy is transferred to the \mathbf{e}' mode of $\text{Fe}(\text{CO})_5$ and then the UV excitation and photodissociation occurs. This is, however, unlikely as the intensity of the $\text{Fe}(\text{CO})_3$ peak in the T-2D-IR spectrum does not significantly vary between the two modes. Figure 2.19 shows a gradual increase in the amount of $\text{Fe}(\text{CO})_3$ formed as the delay time between IR and UV pump pulses decreases to 2 ps. This gradual increase is related to the solution dissipating less vibrational energy from $\text{Fe}(\text{CO})_5$ molecules as the delay time decreases. If the vibrational pre-excitation was mode specific, this trend would follow a more substantial increase in intensity versus delay time and would be on a similar timescale to the Berry-pseudorotation process, which it is not. Interestingly, in CH_2Cl_2 , it might be expected to see a very small change in $\text{Fe}(\text{CO})_3$ formed when the higher energy \mathbf{a}_2'' mode was excited since the lower energy vibrational mode would be populated via the Berry-pseudorotation process. This is not observed in Figure 2.21, most likely due to a combination of two factors: firstly, the signals involved in T-2D-IR are very small and so any change that can be observed must be significant to be visible above the noise, and secondly it is only 2 ps between IR and UV excitations and therefore the population of the \mathbf{a}_2'' mode is too small to observe a change. Unfortunately data

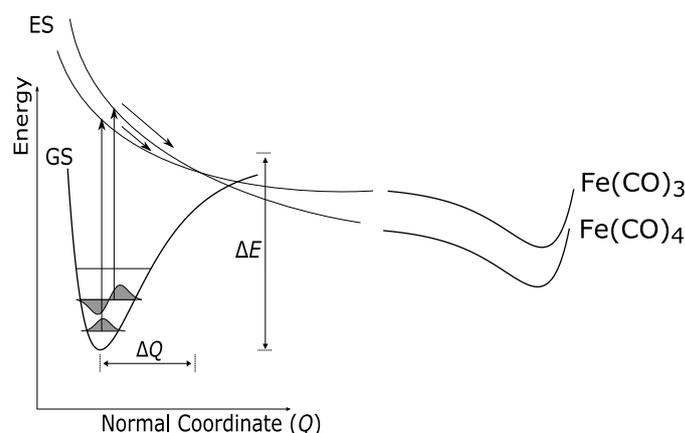


Fig. 2.24 A simplified diagram showing the ground (GS) and dissociative excited states (ES) potentials of $\text{Fe}(\text{CO})_5$ as well as potentials for the two photoproducts $\text{Fe}(\text{CO})_4$ and $\text{Fe}(\text{CO})_3$. Two different electronic transitions: one from the vibrational ground state (left) and one from the 1st vibrationally excited state (right) are displayed.

at longer time delays between excitations were not collected. It is of course assumed that Berry-pseudorotation for $\text{Fe}(\text{CO})_5$ occurs in CH_2Cl_2 , but solvent assisted IVR would also follow the same arguments. There is also the issue of lineshape ‘leakage’ in the pump axis, where the lineshape of the IR_{pump} is reflected in the spectrum. This makes it more difficult to conclusively say whether the IR_{pump} is vibrationally mode specific or not because it has a FWHM of *ca.* 28 cm^{-1} . However, there is substantial contrast between $\text{Fe}(\text{CO})_5$ in heptane and CH_2Cl_2 and so it is not thought to be significant. The spatial resolution in the pump axis could also be improved by using the pulse-shaping technique.

The vibronic coupling of $\text{Fe}(\text{CO})_5$ in solution is clearly a major factor in these results. The vibrational pre-excitation will perturb the vibrational wavepacket relaxation pathway within the excited state potential. Figure 2.24 shows a cartoon representation of this process, with $\text{Fe}(\text{CO})_5$ excited into a dissociative state before eventually forming either $\text{Fe}(\text{CO})_4$ or $\text{Fe}(\text{CO})_3$. The slight change in wavepacket relaxation pathway could result in less chance of

lower energy states being accessed and therefore a greater or lesser probability of $\text{Fe}(\text{CO})_3$ forming in heptane or CH_2Cl_2 , respectively. This is likely responsible for the change in formation of $\text{Fe}(\text{CO})_3$. It is also possible that with $\text{Fe}(\text{CO})_5$ in a vibrationally excited state the ‘extended’ CO ligands would result in a greater probability of dissociation, however, this is not evidenced by a decrease in $\text{Fe}(\text{CO})_3$ with $\text{Fe}(\text{CO})_5$ in CH_2Cl_2 . By vibrationally exciting the molecule first, the ‘extra’ energy is provided before the photodissociation process occurs and so it is logical that more energy in results in a change in an increase in the amount of $\text{Fe}(\text{CO})_3$ produced, however, this again is far too simplistic and not correct, as a decrease is observed for CH_2Cl_2 . Therefore the process involved in vibrationally controlling the photodissociation $\text{Fe}(\text{CO})_5$ is a very complex problem where more experiments are required to elucidate this.

2.4 Conclusion and future work

2.4.1 Conclusion

It has been demonstrated that IR_{pump} pulses can be used to steer the photodissociation of $\text{Fe}(\text{CO})_5$. It was determined that the UV_{pump} pulse followed by the IR_{pump} pulse resulted in no change in the formation of $\text{Fe}(\text{CO})_4$ and $\text{Fe}(\text{CO})_3$. Whereas the IR_{pump} followed by the UV_{pump} resulted in an increase in the formation of $\text{Fe}(\text{CO})_3$ in heptane, but a decrease in $\text{Fe}(\text{CO})_3$ was observed in CH_2Cl_2 . The IR_{pump} was found to be mode specific for CH_2Cl_2

but not heptane as a solvent. It is clear that further work needs to be carried out to understand these observations.

2.4.2 Future work

These experiments could be repeated using the pulse shaping method, this provides better control over the time and frequency resolution which would limit lineshape 'leakage'. It would also be interesting to carry out this experiment again in more polar solvents such as acetonitrile or an alcohol – leading to a better understanding of the solvent effects on the formation of $\text{Fe}(\text{CO})_3$. $\text{IR}_{\text{pump}}\text{-UV}_{\text{probe}}$ experiments could be used to observe if there is a major shift in the electronic transition at 266 nm of $\text{Fe}(\text{CO})_5$ as a result of the vibrational excitation. Another very interesting experiment would be to try exciting the CO ligand overtone band and observing the effect this has on the formation of $\text{Fe}(\text{CO})_3$, however, this may be difficult as signals could be very weak. This could inform us whether the changes observed are caused simply by the molecule being excited to a higher energy, or it is the perturbation of the vibrational wavepacket relaxation in the excited state that causes this.

DFT calculations for the photodissociation of $\text{Fe}(\text{CO})_5$ could be used to gain an understanding into these observations. PES could be calculated for the dissociation of CO ligands for a ground state molecule as well as for a vibrationally and electronically excited molecule. This could be further expanded from gas state to solvated molecule calculations, the impact of different solvents on the PES could be determined this way. However, since these calculations

would involve modelling conical intersections and vibronic coupling, these calculations would be extremely challenging and would require bespoke methods not available in conventional *ab initio* calculation software packages which were not available during my studies, and would require collaboration with more specialist theory groups.

Chapter 3

Gelation of Carrageenan

3.1 Introduction

3.1.1 Structure and Function

Carrageenan is a sulfated polysaccharide that is an extract of red seaweed. It is used throughout the food and cosmetic industries as a thickening, stabilising or gelling agent. This is because it is a hydrocolloid, which can change the viscosity, solubility and water binding affinity of component mixtures all of which contribute to the critical property, texture. Carrageenan is useful compared to other gelation agents because it can form gels at very low concentrations (*ca.* 1% w/w).⁷⁷ It has also been shown that it limits ice crystals growth in frozen foods which is crucial to the stability and lifespan of the product.⁷⁸

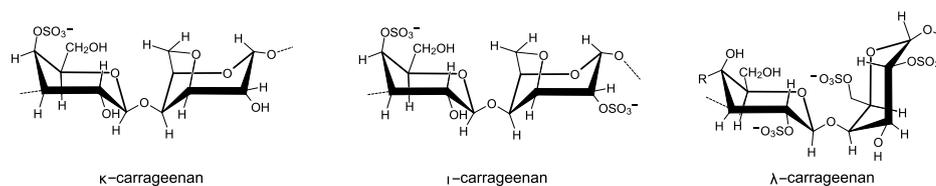


Fig. 3.1 The chemical structures for κ -, ι - and λ -carrageenan.

Carrageenan polymer chains are formed from repeating galactose and 3,6 anhydrogalactose units which can be substituted with sulfur groups at various positions, glycosidic bonds link the various units together. There are up to 15 different forms of carrageenan, however, the three main forms are: κ -, ι - and λ -carrageenan,⁷⁹ differing from each other in the number and position of the organosulfate groups. The average molecular weight of carrageenan chains is *ca.* 500 kDa.⁸⁰ The chemical structures of these forms are shown in Figure 3.1. κ - and ι -carrageenan are both gel-forming, whereas λ -carrageenan does not form a gel.⁸¹ The interactions with cations with and around these functional groups are thought to be critical to the gel's properties and the gelation mechanism.

3.1.2 Gelation Process

Carrageenan exists in disordered random coils in solution at high temperatures.⁸² As the temperature is lowered past the gel-sol transition, double helix structures start to form. These double helices can then aggregate as the temperature is lowered further (Figure 3.2).⁸³ It has been shown that cations contained within solution stabilise the helical structures both intramolecularly and intermolecularly.⁸⁴ Many different factors affect the gelation of carrageenan, these include: salt type (more specifically, size and valency of the cation), salt concentration, pH of solution and carrageenan concentration.⁸⁵

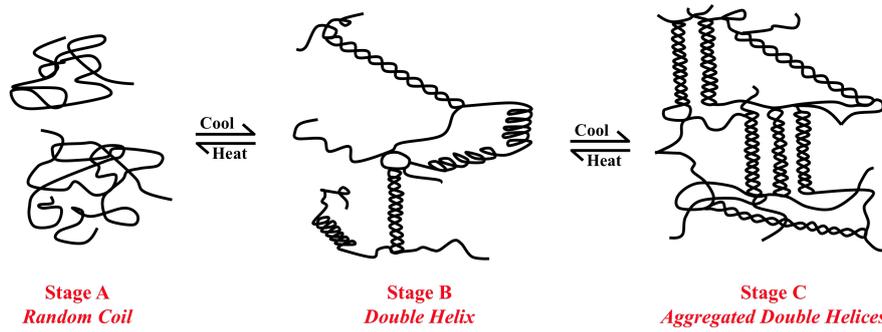


Fig. 3.2 Cartoon showing how carrageenan forms helical structures and aggregates from a random coil state as the temperature of solution is lowered. Reproduced from [86].

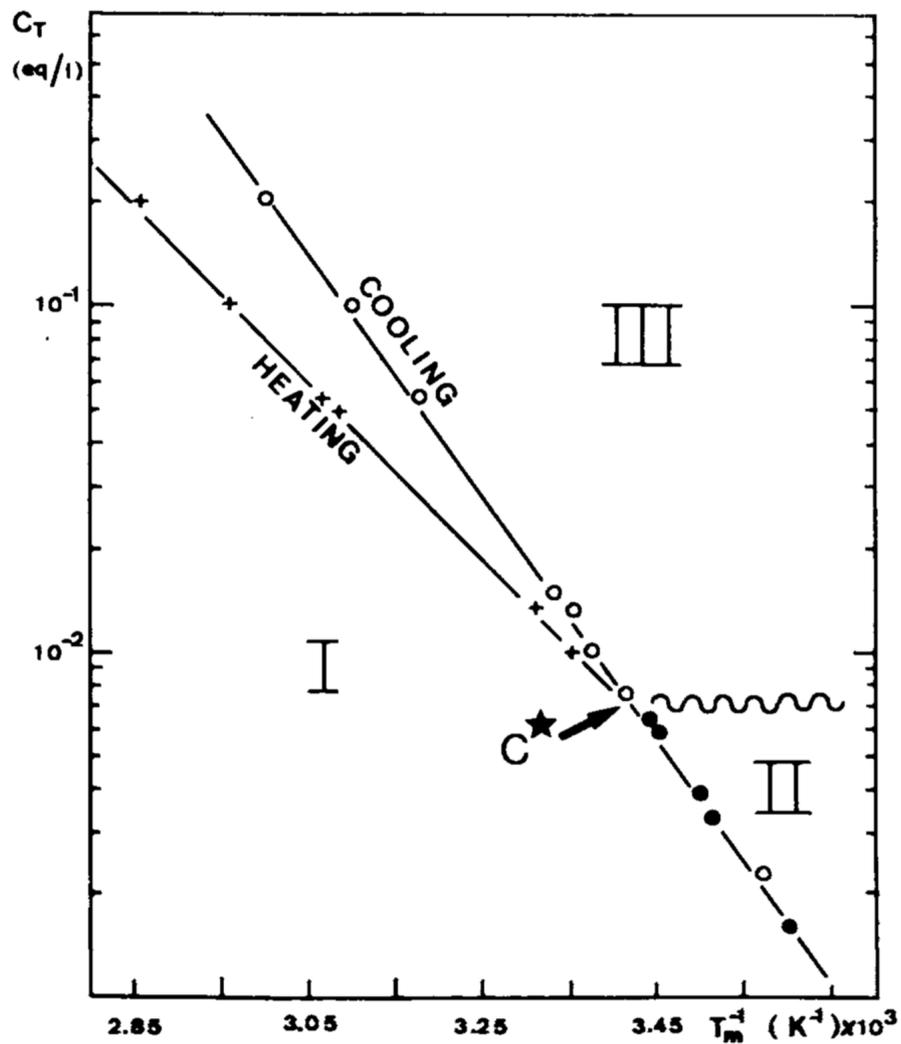


Fig. 3.3 Phase diagram showing gel-sol transition of κ -carrageenan and the three different phases. The plot displays heating and cooling plotted with the free concentration of K^+ (C) against melting temperature (T_m). Reproduced from [87].

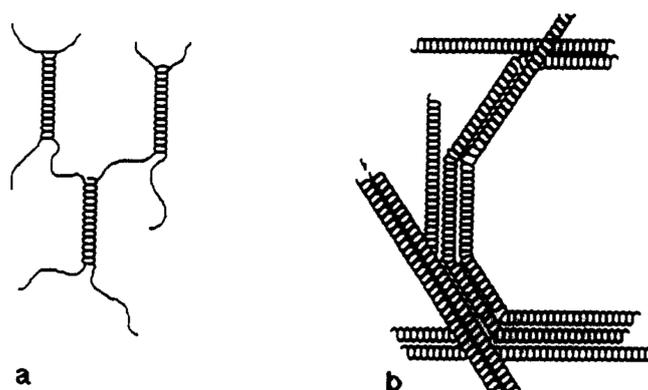


Fig. 3.4 Cartoon representation showing the different types of gelation that can occur with carrageenan via (a) aggregation of helices (superhelix formation) (b) cross-linking junctions. Reproduced from [83].

The cation concentration and the temperature are the two main factors that affect the gelation of carrageenan. Figure 3.3 shows the phase diagram of κ -carrageenan which has been plotted with the critical K^+ concentration (C_T) against melting temperature (T_m) with heating and cooling curves. The carrageenan firstly exists in phase I which is the solution phase, here the carrageenan chains are in a disordered conformation. Phase II exists when the temperature is lowered at small concentrations of K^+ . No gelation occurs because it is believed that there is not a high enough concentration of cations to stabilise the organosulfate groups that have to be in close proximity to each other.⁸⁸ Phase III is the gelation phase where the κ -carrageenan double helices aggregate together as the temperature is lowered at higher K^+ concentrations. It is thought that these aggregates are stabilised by the cations.⁸³

κ -carrageenan

κ -carrageenan forms brittle gels. This is thought to be because the double helices aggregate into super helical conformations (as shown in Figure 3.4).⁸⁹ It gels well with large univalent

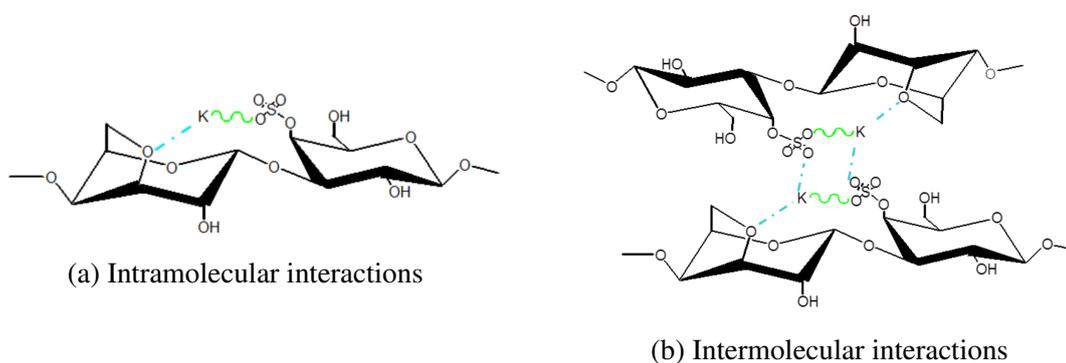


Fig. 3.5 The chemical structure of κ -carrageenan showing the ionic and electrostatic cation interactions respectively represented by green and blue lines. Reproduced from [85].

cations present such as K^+ , Rb^+ and Cs^+ . However, smaller univalent cations such as Na^+ and Li^+ have a detrimental effect on the gelation of κ -carrageenan, even preventing gelation altogether providing the concentration is high enough.^{88,90,91} Interestingly, large anions can also prevent gelation. For example iodide ions have been shown to reduce gelation, this is thought to be caused by the screening effect the large anions play with respect to the cations present, preventing the helices from aggregating.⁹² It is thought that in the case of K^+ , it interacts mainly with the oxygen of the organosulfate group as well as with the oxygen of the ether group, this is shown in Figure 3.5b. In the case of intermolecular interactions, the cation again has two interactions with the first molecule whilst having a third electrostatic interaction with the sulfate group of the other molecule (shown in Figure 3.5b).⁸⁵

***ι*-carrageenan**

Conversely, ι -carrageenan forms more flexible gels, this is because large helical aggregates are not thought to form. Instead, the gelation occurs because various junctions form, connecting the double helices together (shown in Figure 3.4). Evidence based on the lack hysteresis

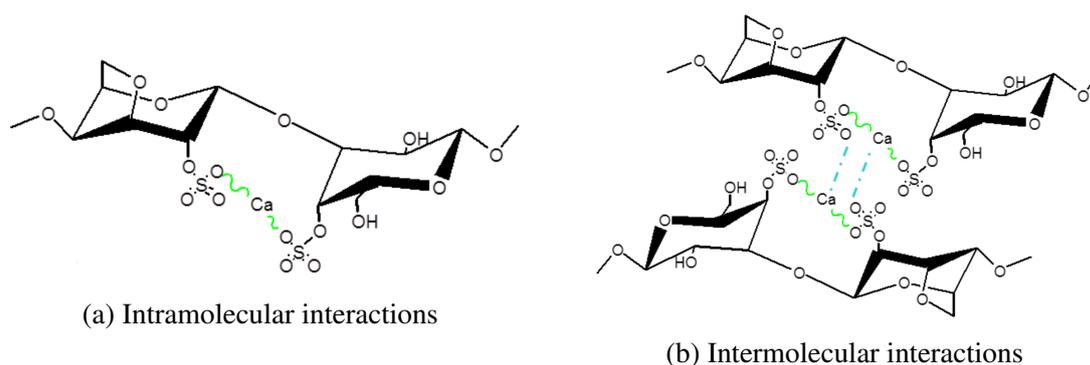


Fig. 3.6 The chemical structure of ι -carrageenan showing the ionic and electrostatic cation interactions respectively represented by green and blue lines. The charge density is clearly higher when compared with κ -carrageenan. Reproduced from [85].

compared with the κ -carrageenan supports this.⁸³ ι -carrageenan undergoes gelation with divalent cations such as Ca^{2+} but not with smaller univalent cations like Na^+ .⁹⁰ This is because the ι form contains another organosulfate group and hence is more negatively charged, therefore a larger charge is required to stabilise both organosulfate groups. Intramolecularly, the cation experiences ionic interactions with oxygens from both organosulfate groups as shown in Figure 3.6a. The cation can then experience a third electrostatic interaction in the intermolecular case. This is from one of the sulfate groups on the other molecule and is a weaker electrostatic interaction, this is shown in Figure 3.6b.⁸⁵

λ -carrageenan

λ -carrageenan is known as the non-gelling carrageenan because of the three organosulfate groups located on the disaccharide unit which are thought to repel each other. However, it has been suggested that it does gel at very high concentrations *ca.* 5 % w/w or in the presence of trivalent cations such as Fe^{3+} .⁹³

3.1.3 Previous research

The gelation and structure of carrageenan has been investigated for over 60 years using a whole variety of methods. For example, x-ray scattering experiments by Bayley provided the first evidence for the helical structure of κ -carrageenan.⁹⁴ Since then, other x-ray scattering experiments have provided evidence for the double helix structure found in both κ and ι -carrageenan, as well as the aggregation of helical structures.^{84,95,96} Atomic force microscopy has been used to elucidate how the polymer strands coil and aggregate.⁹⁷ A host of other experimental techniques have been used to analyse carrageenan, for example NMR, dynamic light scattering and differential scanning calorimetry (DSC).^{89,98,99} The section below highlights a variety of experimental techniques used to investigate carrageenan.

Rheology has been the main technique used to study the gelation of carrageenan. The sol-transition is defined in rheology as the crossover point between the bulk and storage modulus. For 2 % κ -carrageenan solutions the transition temperature was determined to be *ca.* 30 °C and for 2 % ι -carrageenan solutions it was determined to be *ca.* 50 °C.^{87,100} There have been many rheological studies investigating both low and high temperature regimes, as well as different solution concentrations, cation types and concentrations of carrageenan.^{90,101–103}

A light scattering study by Meunier *et al.* looked at how the rate of aggregation of κ -carrageenan changed with temperature and salt concentration.⁹⁹ They concluded that the aggregation rate increases with decreasing temperature and increasing salt concentration. It

was also stated, based on optical rotation measurements, that helical conformation is not affected by aggregation.⁹⁹

DSC was used by Viebke *et al.* to measure the transition enthalpies of κ -carrageenan related to the formation of aggregates and investigate how these change with salt concentration and type. There is evidence that the stability of the aggregates was only dependent on the concentration of caesium cations and not the type of anion present. This suggested that the structure of aggregates is highly dependent on the cations present.⁸⁹ However, the structure and formation of the aggregates is still relatively unknown and it is not clear if these aggregates form smaller structures before combining into larger ones.

NMR is a very powerful spectroscopic technique which has been applied to a wide variety of research.¹ Carrageenan studies using NMR have even been undertaken. For example Bosco *et al.* used ^1H and ^{13}C NMR to investigate the structure of κ -carrageenan in the disordered phase (*ca.* 60 °C).⁸² It was found that the chains explored rather restricted conformations and this was further supported by molecular dynamics simulations. It was concluded that in the disordered state the carrageenan chains form loose helices or coils and that there was no evidence for the interaction between chains to form double helices in this phase. Although molecular dynamic models support these results, DMSO was used as a solvent for the carrageenan, therefore it is unknown exactly how carrageenan interacts with the solvent and caution must be exercised when extrapolating these findings to using water as a solvent.⁸²

A very interesting and detailed study again utilised ^1H and ^{13}C NMR to investigate the effect that small impurities (such as μ -carrobiose) have on the structure of carrageenan gels. Evidence was found to suggest that these impure units disrupted the helical structure and were therefore likely to play a role in the interactions between various chains at junctions (shown in Figure 3.4). These main interactions at junctions are thought to be hydrogen bonds forming between various μ -carrobiose units.¹⁰⁴

Naumann *et al.* investigated both ι - and κ -carrageenan using ^2H and ^{23}Na NMR, specifically looking at how the anisotropy of the gel changed as it was stretched using specialist apparatus. Both ι and κ forms were shown to be isotropic but became anisotropic when stretched. A larger anisotropy was observed for the ι form, this is most likely due to the Na^+ cations electrostatically interacting with the organosulfate groups. As gelation is being investigated, the solvent is an essential component of the structure, however, most NMR solvents are deuterated and these can cause slightly different gel structures.¹⁰⁵ Therefore care must be taken when interpreting results from experiments that use D_2O as a solvent.¹⁰⁶

3.1.4 Vibrational Spectroscopy

Vibrational spectroscopy is a very powerful technique which can provide vast amounts of information about the system being studied. It is very flexible, requires very little sample preparation and often allows the sample to be studied *in situ*. There have been many studies of carrageenan using FTIR and to a lesser extent Raman spectroscopy. The first recorded

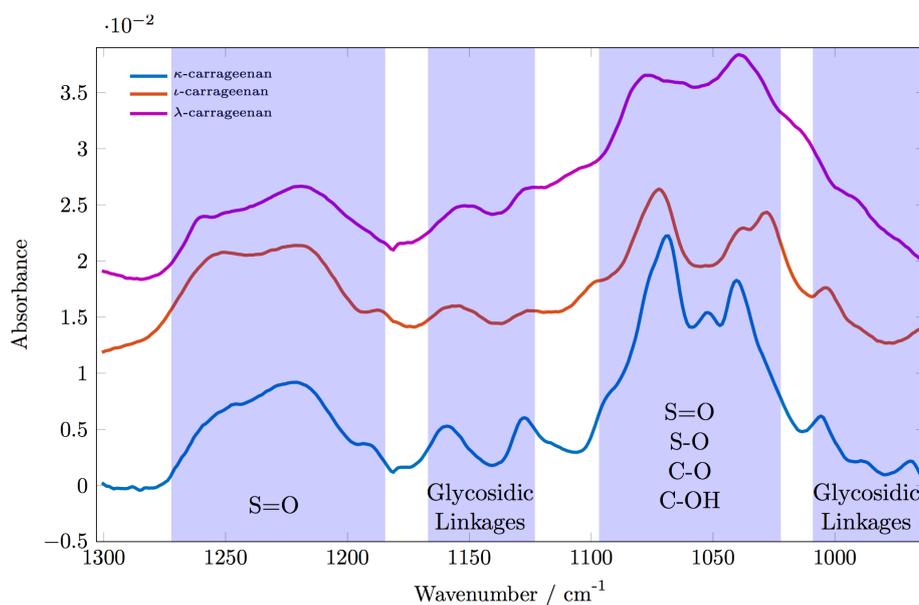


Fig. 3.7 ATR-FTIR spectra of κ -, ι and λ -carrageenan at 20 °C. The main spectral features have been highlighted. Reproduced from [113].

FTIR spectra of carrageenan was by Bayley in 1955, since then many studies have been undertaken.^{107–112}

Many early studies concentrated on band assignments by comparing differences between the various types of carrageenan.^{107–111,114,115} These have been summarised in Table 3.1 and Figure 3.7 which shows ATR-FTIR spectra of the three main types of carrageenan at room temperature. Belton *et al.* first investigated the effect of different cations on the FTIR spectra of κ and ι -carrageenan.¹⁰⁷ It was concluded that the spectral variations observed for different cations were due to differences in the carrageenan structure as opposed to specific interactions of the cations with the carrageenan chains. Following this, a variable temperature study was conducted over a temperature range of 80 - 30 °C, with various cations.¹⁰⁸ The spectral changes observed were again attributed with conformational changes in the carrageenan structures and gel-sol transition temperatures were identified.¹⁰⁸ The conclusion from the

Table 3.1 Spectral band assignments for κ -, ι and λ -carrageenan, +/- represents present/not present in spectrum. Reproduced from [114].

Wavenumber / cm^{-1}	Functional Group	κ	ι	λ
3400-3000	O-H	+	+	+
2920	C-H (stretching)	+	+	+
1380-1355	Sulfates	+	+	+
1250-1230	O=S=O (antisymmetric stretching)	+	+	+
1190	S=O (antisymmetric stretching)	+	+	-
1160-1155	C-O-C (antisymmetric stretching)	+	+	+
1125	Glycosidic bonds (antisymmetric stretching)	+	+	+
1090	S-O (symmetric stretching)	+	+	+
1080-1040	C-O and C-OH	+	+	+
1045	C-OH and S=O	+	+	+
1026	S=O in C2 (pseudo-symmetric stretching)	-	+	+
1012	S=O in C6 (pseudo-symmetric stretching)	-	-	+
1002	Glycosidic bonds	+	+	+
970-965	Glycosidic bonds	+	+	+
930	C-O-C (3,6-anhydrogalactose)	+	+	+
900-890	C6 group in β -D-galactose	+	+	+
850-840	C4-O-S group in galactose (stretching)	+	+	-
830-825	C2-O-S galactose	-	-	+
820-810	C6-O-S (stretching)	-	-	+
805-800	C2-O-S in 3,6-anhydrogalactose	-	+	-
740-725	C-O-C α (1,3) (stretching)	+	+	-
615-608	O=S=O (bending)	+	+	+
580	O=S=O (bending)	+	+	+

earlier study, that the spectral changes were not directly due to the different cations bound to the chains but a result of inherent structural changes, was again reinforced.¹¹⁰

More recently, FTIR spectroscopy has been used to help identify the different types of carrageenan, as well as the respective purity of the samples. This has been achievable by combining the ease of ATR-FTIR spectroscopy with partial least squares (PLS) modelling; which utilises a set of calibration samples to form a model which can then be used to predict the purity of samples.^{109,114}

Calum Welsh most recently used ATR-FTIR spectroscopy in conjunction with multivariate analysis and 2D correlation spectroscopy (2DCOS) to investigate the gelation of carrageenan.¹¹³ Figures 3.8 and 3.9 show the multivariate curve resolution (MCR) components (akin to pure component spectra) and scores (concentration profile) of ι - and κ -carrageenan solutions as they cooled from 80 to 10 °C, respectively. Sharp changes in gradient are observed with the scores, which are thought to correspond to the start of molecular rearrangements as part of the sol-gel transition. For ι -carrageenan, this starts at *ca.* 60 °C and the crossover point occurs at *ca.* 53 °C which are both significantly above the rheologically determined transition temperature of *ca.* 35 °C. It is thought that the ATR-FTIR spectra are detecting the molecular rearrangements of the carrageenan, which must occur before the macroscopic sol-gel transition is detected by rheology. These molecular rearrangements continue to occur (although slowing) throughout the remainder of the cooling process. For κ -carrageenan, the sharp change in the score gradients occur at a much lower temperature of *ca.* 33 °C (30 °C at the crossover point) which is much closer to the rheologically determined

sol-gel temperature of 35 °C. The scores also appear to stop changing after 20 °C and suggest that κ -carrageenan has a much smaller temperature window for its sol-gel transition than ι -carrageenan.

A study recently undertaken by Makshakova, utilised FTIR spectroscopy in combination with DFT calculations to investigate cation binding and structural changes associated with the gelation of κ -carrageenan. It was found there were three different conformations for the organosulfate groups assigned to: one cation free state and two different cation bound states. Upon cooling, specific cation binding of the organosulfate groups preceded the formation of the helical structure. The DFT calculations indicated that the specific cation binding helped stabilise the helical conformation of κ -carrageenan. This was by interacting with two oxygen atoms of the organosulfate groups and one oxygen from the ether group on the galactose monomer unit.¹¹⁶

3.1.5 Aims

2D-IR spectroscopy allows us to directly probe the organosulfate functional groups and their environment as gelation occurs. The aim of this chapter was to further the understanding of the process of gelation carrageenan by investigating the molecular rearrangements that occur during this process. The solvent environment within carrageenan solutions is also important to investigate because, as stated previously, carrageenan has been shown to slow the growth of ice crystals in frozen solutions and the presence of carrageenan must therefore directly affect the Oswald ripening process.¹¹⁷ In the next sections the gelation of carrageenan is

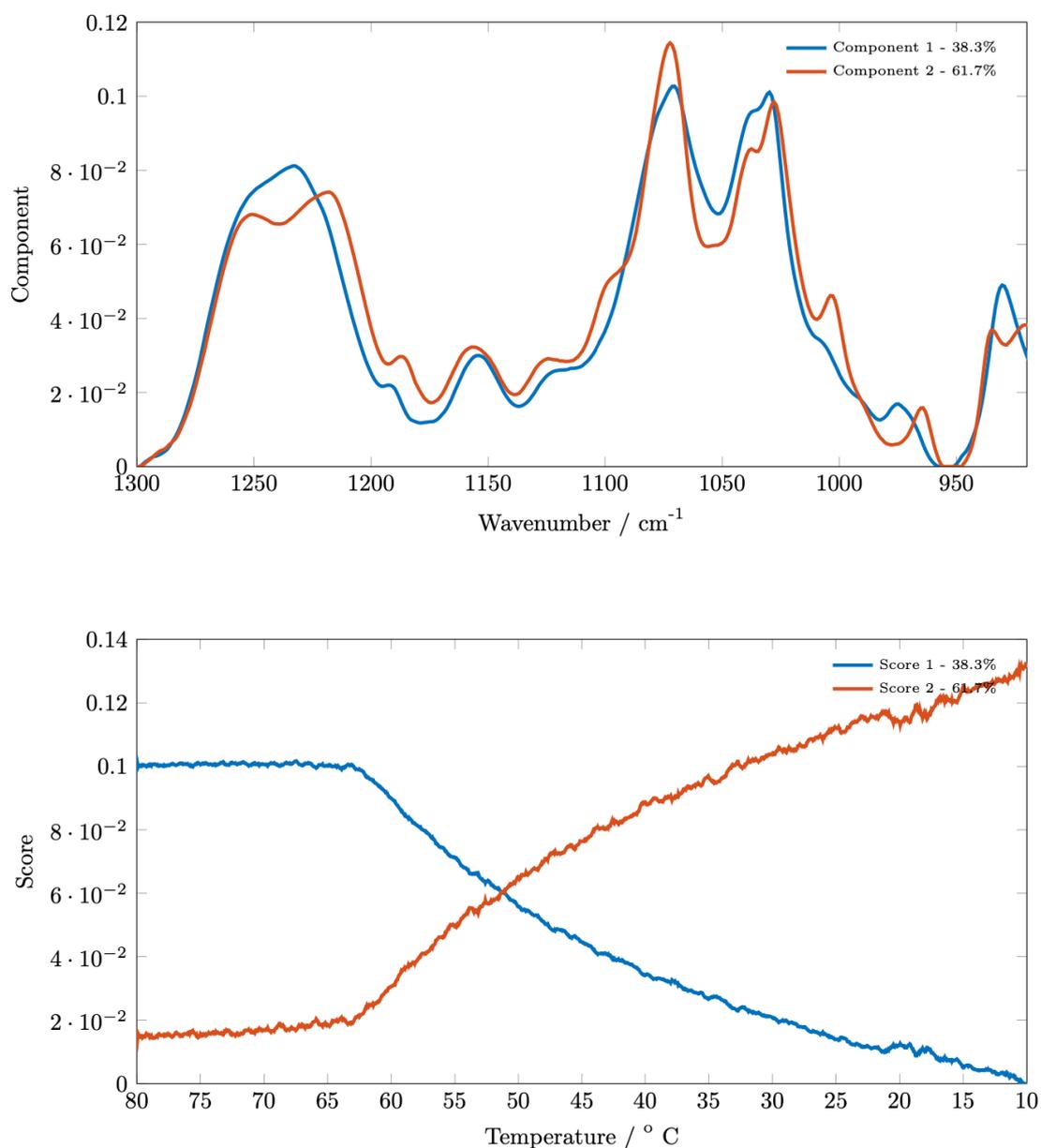


Fig. 3.8 MCR components (above) and scores (below) of 2% ι -carrageenan solution ATR-FTIR spectra as it is cooled from 80 to 10 °C. The process of gelation only appears to start *ca.* 60 °C and continue past the lower temperature limit of 10 °C, with the midpoint of gelation occurring at *ca.* 50 °C. Reproduced from [113].

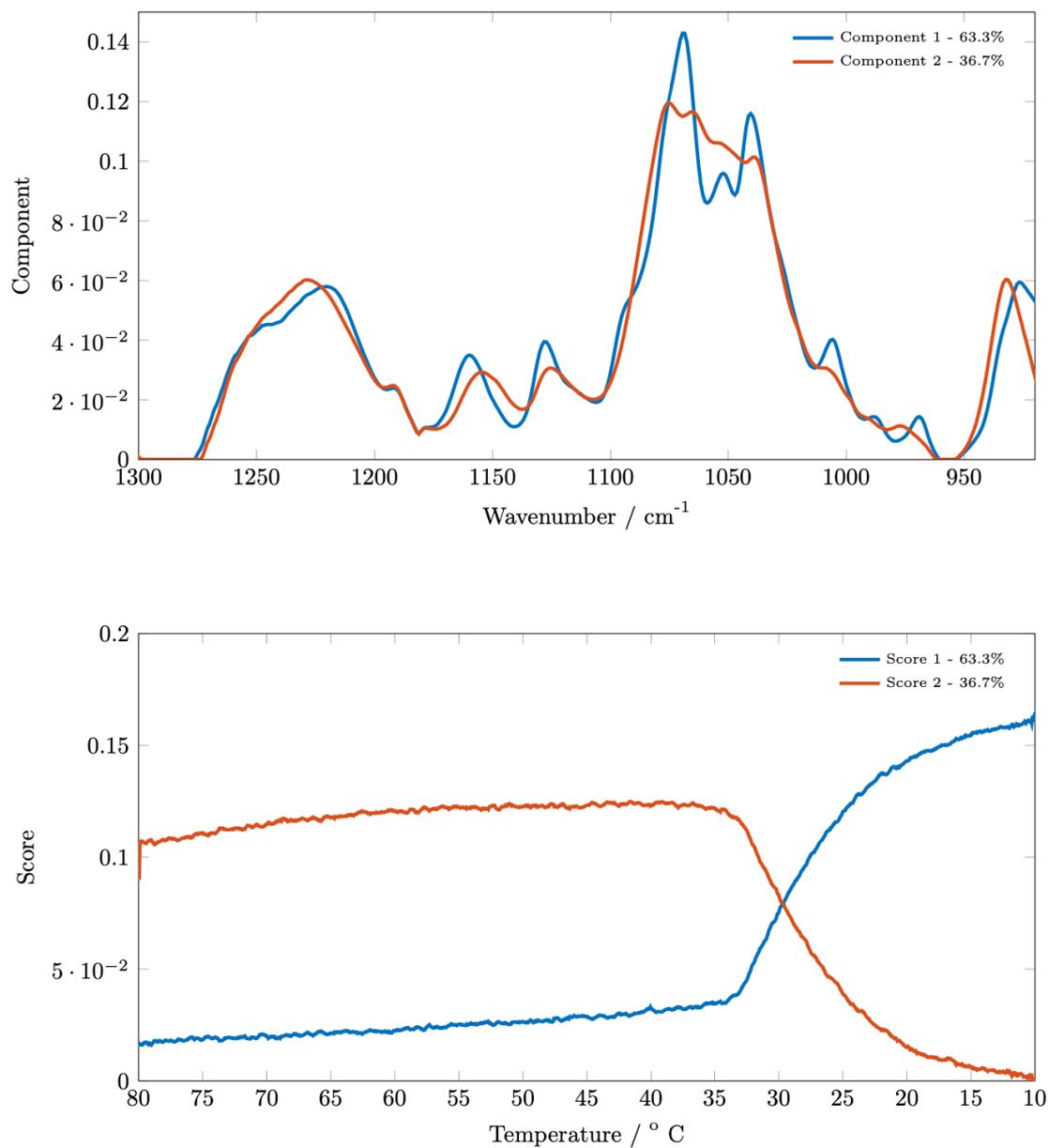


Fig. 3.9 MCR components (above) and scores (below) of 2% κ -carrageenan solution ATR-FTIR spectra as it is cooled from 80 to 10 °C. The process of gelation only starts much later *ca.* 35 °C and appears to stop just past the lower temperature limit of 10 °C, with the midpoint of gelation occurring at *ca.* 30 °C. Reproduced from [113].

investigated using 2D-IR spectroscopy. Firstly, by probing how the organosulfate functional groups change throughout the gelation process. Following this, the water dynamics of carrageenan solutions were investigated via the use of a probe molecule (ferrocyanide) and looking specifically at how spectral diffusion varies with temperature. This will serve as a basic foundation from which sub-zero temperature solutions can be investigated in the future.

3.2 Results

This section presents 2D-IR spectra obtained for κ -carrageenan and ι -carrageenan solutions at a range of temperatures. The first set of spectra are of 2 % κ - and ι -carrageenan solutions, focussing on the organosulfate band at *ca.* 1250 cm^{-1} , at temperatures of 70, 45 and 22 °C as gelation occurs. These spectra were collected using the frequency domain technique. The second set of spectra are of 4 % ι - and κ -carrageenan solutions containing ferrocyanide and investigate the solvent dynamics. The spectra were collected using the pulse-shaping method in the time domain and looked at the CN stretch of potassium ferrocyanide in solution. Figure 3.10 shows the structure and FTIR spectrum of potassium ferrocyanide in water, only one peak is observed at *ca.* 2040 cm^{-1} in a clear region of the spectrum. Spectra were collected at a range of temperatures from 20-80 °C and analysed to assess the effect of gelation on the solvent dynamics.

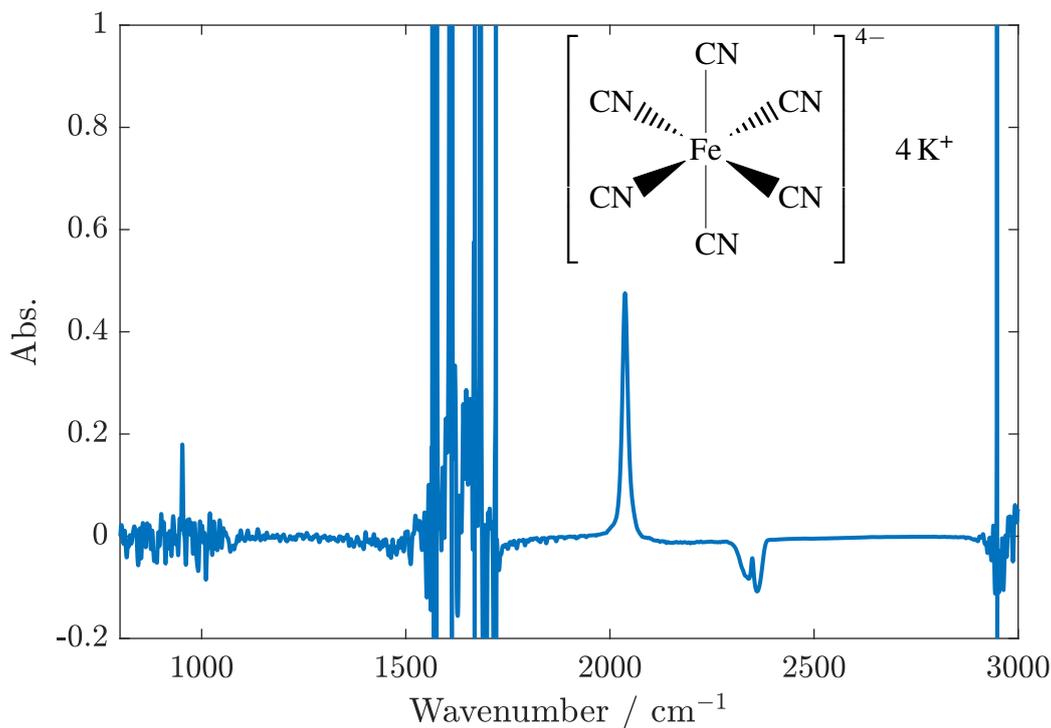


Fig. 3.10 FTIR spectrum of potassium ferrocyanide showing only one peak at 2040 cm^{-1} in the spectral region of interest. The chemical structure of potassium ferrocyanide is displayed above the spectrum.

3.2.1 *t*-Carrageenan

Looking first at *t*-carrageenan, the 2D-IR spectrum of a 2% w/w *t*-carrageenan solution at $70\text{ }^{\circ}\text{C}$ (solution state) is shown in Figure 3.11. Two large broad peaks are the main features of this spectrum. Firstly, the blue peak ($\omega_{probe} = ca. 1260\text{ cm}^{-1}$) represents a depletion in the population of the ground vibrational state (bleach $\nu_{0\rightarrow 1}$). Secondly, the red peak ($\omega_{probe} = ca. 1210\text{ cm}^{-1}$) represents an increase in the population of the first vibrationally excited state ($\nu_{1\rightarrow 2}$).

As the temperature of the *t*-carrageenan solution is reduced to $45\text{ }^{\circ}\text{C}$, changes are observed in the 2D-IR spectrum shown in Figure 3.12. The most significant change is the appearance

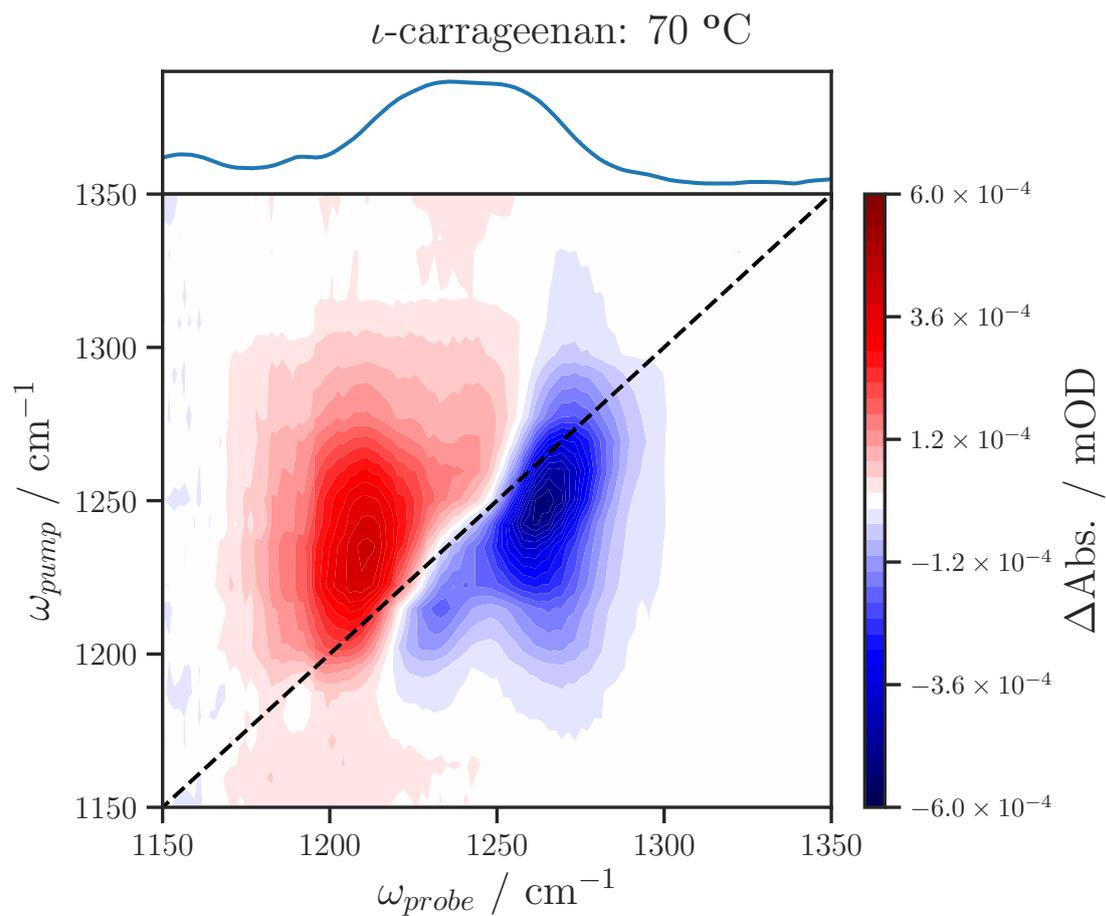


Fig. 3.11 2D-IR spectrum of 2% w/w solution of *t*-carrageenan at 70 °C, red and blue contours represent positive and negative changes in absorbances respectively. Above is the FTIR spectrum for the same sample.

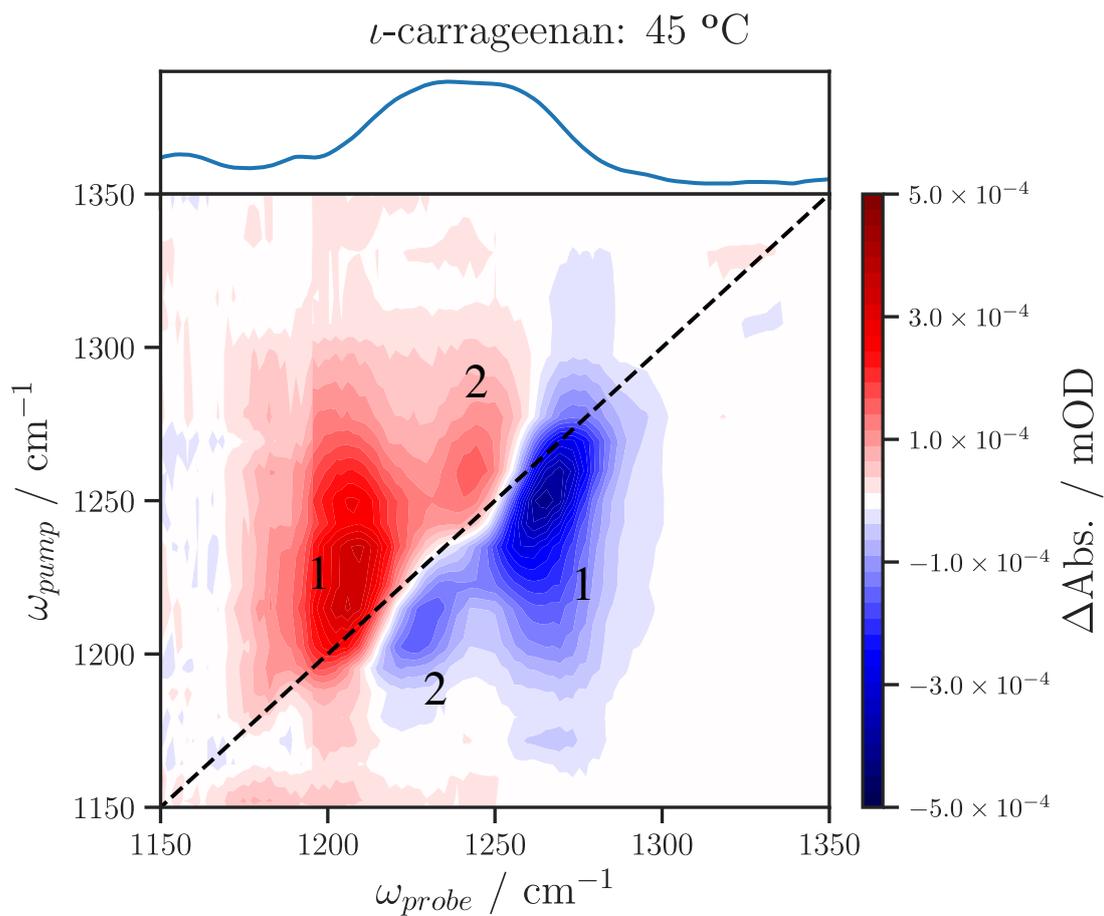


Fig. 3.12 2D-IR spectrum of 2% w/w solution of *ι*-carrageenan at 45 °C, red and blue contours represent positive and negative changes in absorbances respectively. Numbers 1 and 2 represent different spectral features. Above is the FTIR spectrum for the same sample.

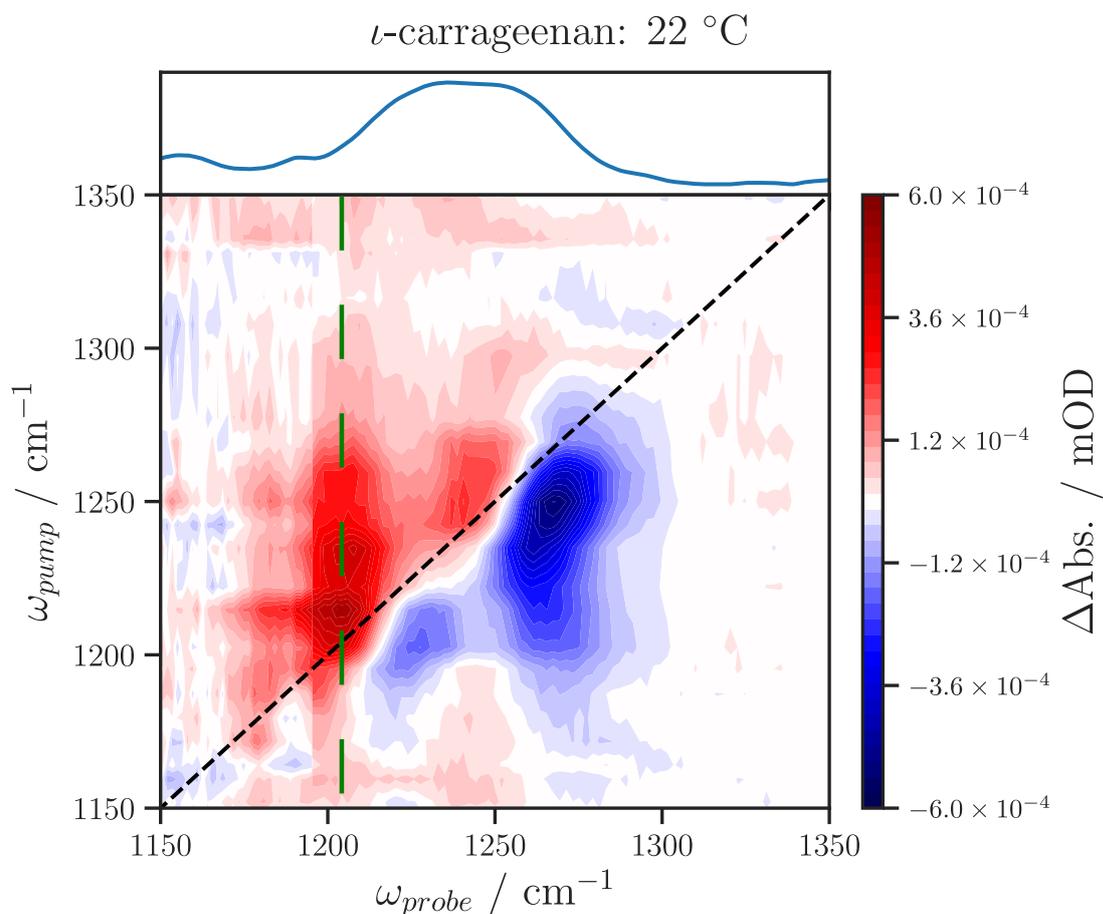


Fig. 3.13 2D-IR spectrum of 2% w/w solution of *t*-carrageenan at 22 °C, red and blue contours represent positive and negative changes in absorbances respectively. The dashed green line shows where spectral slices were extracted from to highlight cross peaks at lower temperature. Above is the FTIR spectrum for the same sample.

of this small hot transition peak centred at $\omega_{probe} = ca. 1245 \text{ cm}^{-1}$ and $\omega_{pump} = ca. 1255 \text{ cm}^{-1}$ (labelled red feature **2**). This is thought to be related to the gelation of *t*-carrageenan as the sol-gel transition should be occurring at this temperature. There is also a small ground state bleach which also becomes more prominent as the temperature is decreased ($\omega_{probe} = ca. 1225 \text{ cm}^{-1}$ and $\omega_{pump} = ca. 1210 \text{ cm}^{-1}$, blue feature **2**) and is again thought to be related to the gelation.

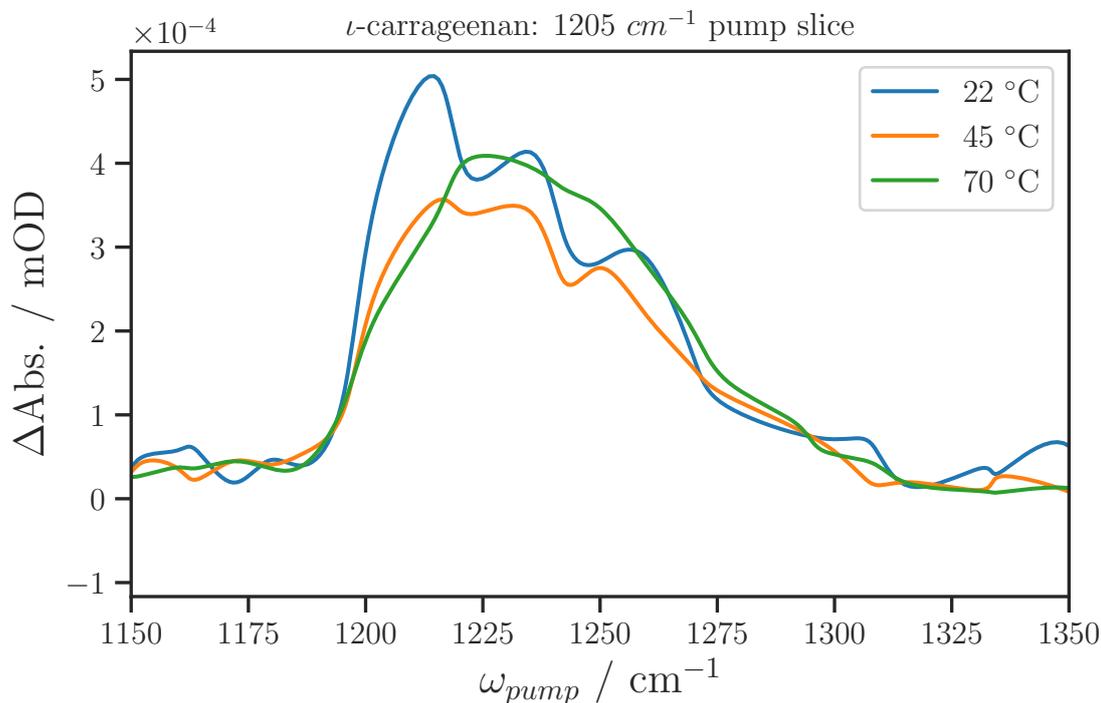


Fig. 3.14 2D-IR interpolated spectral slice of ι -carrageenan along the ω_{pump} axis at 1205 cm^{-1} at different temperatures. Cross peaks become more resolved as the temperature is decreased.

Figure 3.13 shows the 2D-IR spectrum when the temperature ι -carrageenan is further decreased to $22 \text{ }^\circ\text{C}$. The small hot transition peak becomes more prominent, as well as the ground state bleach. It appears to become even more ‘detached’ from the main peaks, the hot transition also appears to blueshift in frequency. Cross peaks can now be observed along the pump axis in the main hot transition. There appears to be three separate peaks in the band – which could indicate separate conformations relating to the organosulfate functional groups.

In order to analyse the cross peaks observed, a probe slice from the 2D-IR spectrum is extracted along the pump axis (represented by the green dashed line in Figure 3.13). These

Table 3.2 The peak positions and relative areas of peaks fitted to ι -carrageenan spectral slices at different temperatures. These show the redistribution in spectral intensity to cross peaks at lower temperature.

Temperature / °C	Peak	Peak position / cm^{-1}	Normalised area
22	1	1212	0.47
	2	1234	0.26
	3	1258	0.28
45	1	1213	0.43
	2	1233	0.25
	3	1257	0.32
70	1	1213	0.33
	2	1233	0.30
	3	1257	0.37

probe slices are shown in Figure 3.14 for each temperature and are obtained at 1205 cm^{-1} on the probe axis. Again, starting at high temperature there is one broad peak, then as the temperature decreases, separate peaks start to become resolved. Each probe slice was fitted with three peaks and the relative areas of the peaks are shown in Table 3.2.* As the temperature decreases, the relative area of the three peaks becomes more distributed towards to the lower frequencies, indicating an overall redshift, as well as the range of environments becoming more limited. At higher temperatures, the relative area of each peak is distributed evenly, suggesting a large range of environments for the organosulfate groups.

*Three peaks were selected based on the spectrum at 22 °C. At 70 °C the overall band is likely formed of many different peaks.

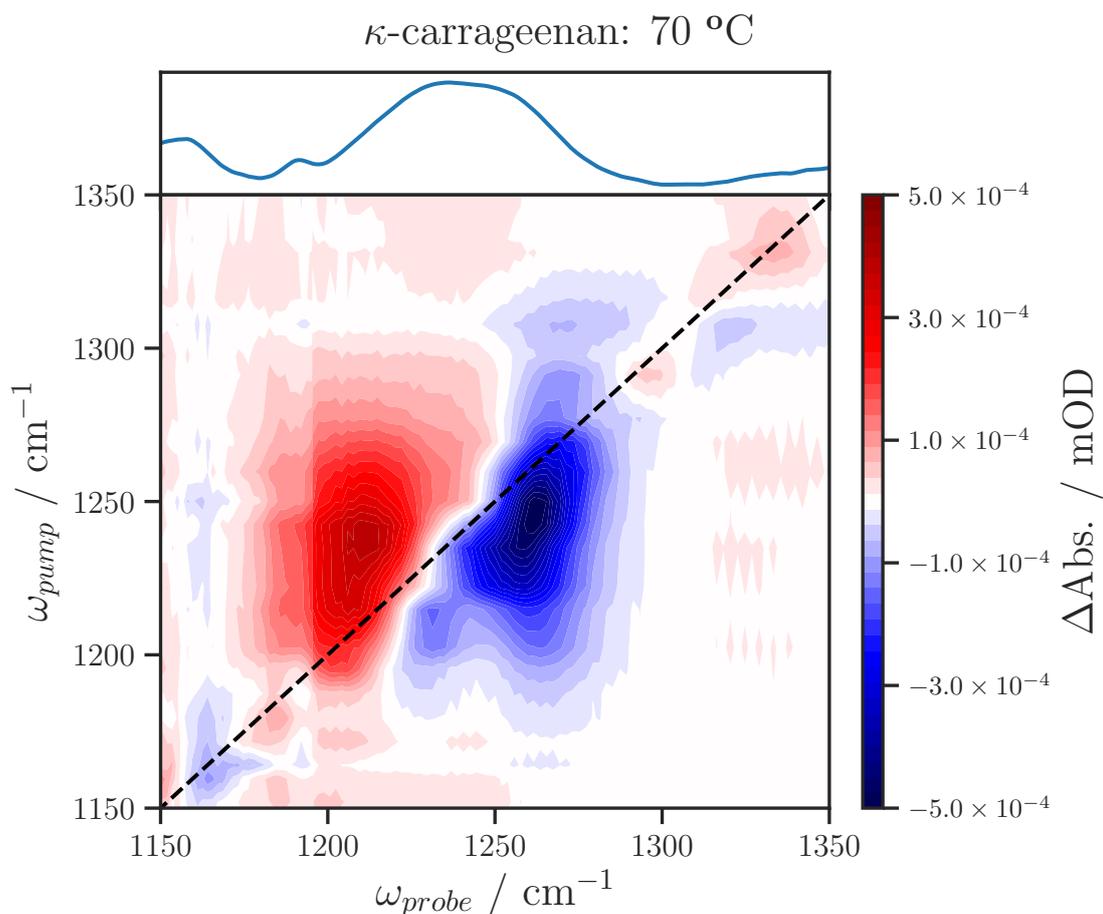


Fig. 3.15 2D-IR spectrum of 2% w/w solution of κ -carrageenan at 70 °C, red and blue contours represent positive and negative changes in absorbances respectively. Above is the FTIR spectrum for the same sample.

3.2.2 κ -Carrageenan

Moving now to look at κ -carrageenan, Figure 3.17 shows the 2D-IR spectrum at 70 °C. Once again the main features are the blue ground state bleach and the red hot transition – these are similar for ι -carrageenan.

As the sample is cooled to 45 °C (past the gel-sol transition) the 2D-IR spectrum changes slightly (Figure 3.16). A small hot transition starts to appear at $\omega_{probe} = ca. 1240 \text{ cm}^{-1}$ and

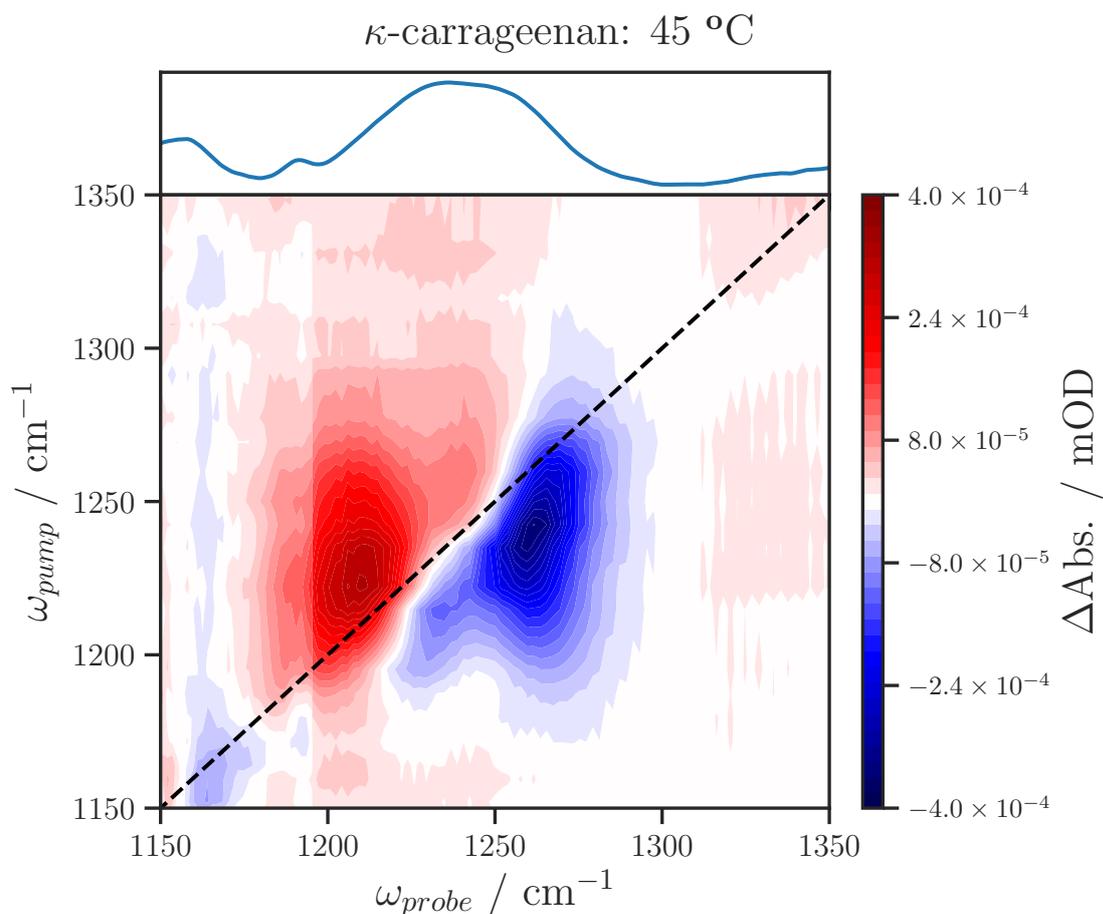


Fig. 3.16 2D-IR spectrum of 2% w/w solution of κ -carrageenan at 45 °C, red and blue contours represent positive and negative changes in absorbances respectively. Above is the FTIR spectrum for the same sample.

$\omega_{pump} = ca. 1250 \text{ cm}^{-1}$ which is a shoulder of the main band. The small ground state bleach ($\omega_{probe} = ca. 1225 \text{ cm}^{-1}$ and $\omega_{pump} = ca. 1205 \text{ cm}^{-1}$) appears to be unchanged with this decrease in temperature. However, the main fundamental transition peak becomes slightly narrower.

Upon a further decrease in temperature to *ca.* 20 °C, these small features in the spectrum become more exaggerated. Figure 3.17 shows the 2D-IR spectrum where the small hot transition appears separate from the main band, as does the small ground state bleach. Both

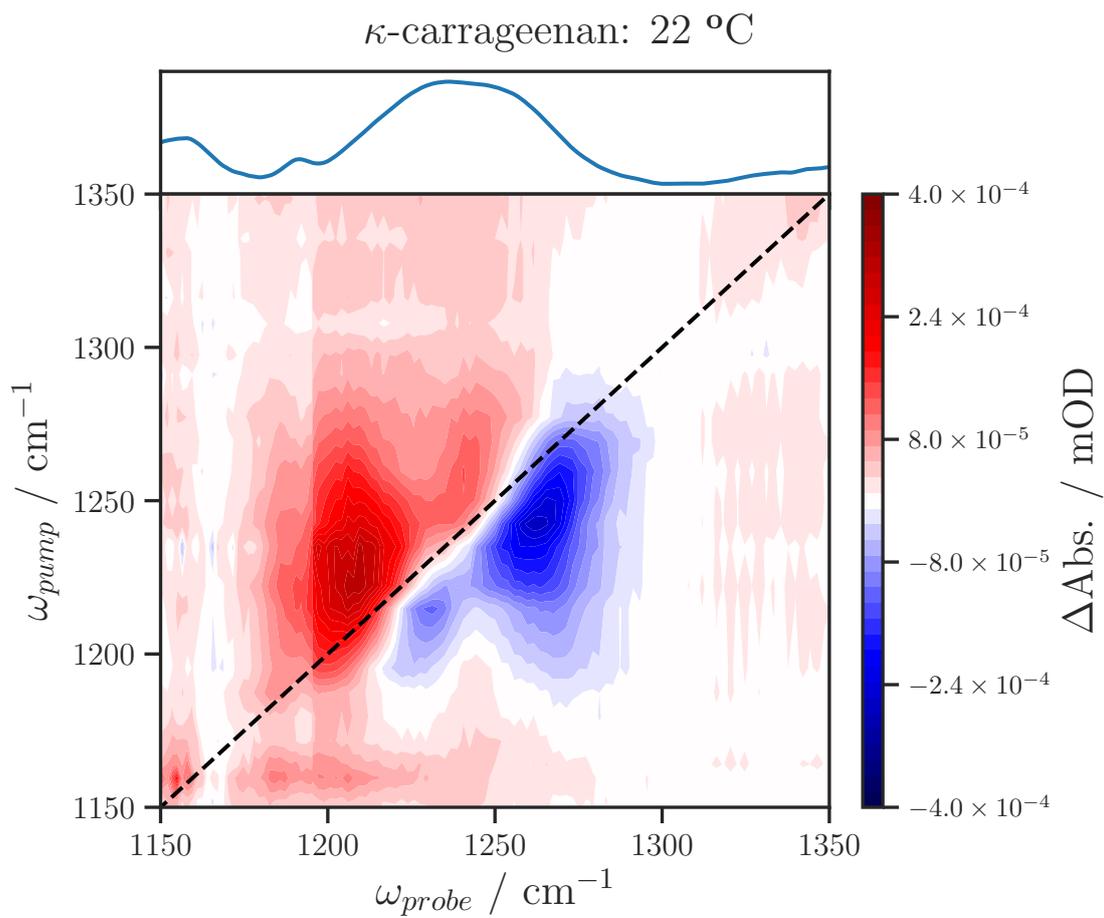


Fig. 3.17 2D-IR spectrum of 2% w/w solution of κ -carrageenan at 22 °C, red and blue contours represent positive and negative changes in absorbances respectively. Above is the FTIR spectrum for the same sample.

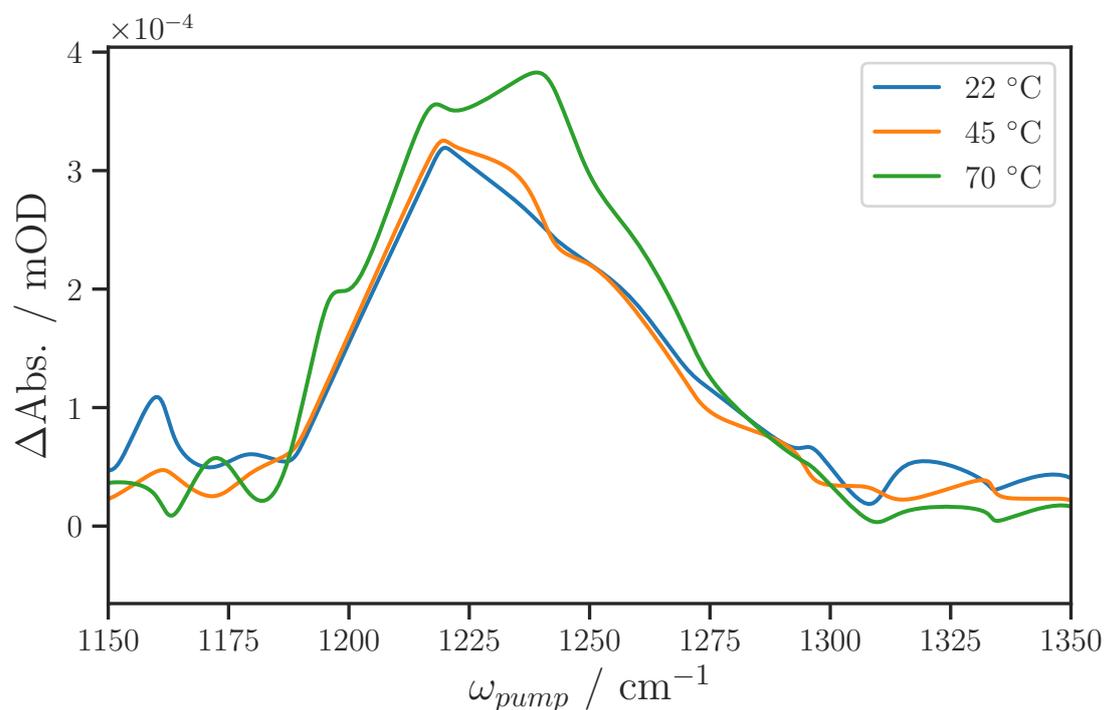


Fig. 3.18 2D-IR interpolated spectral slice of κ -carrageenan along the ω_{pump} axis at 1205 cm^{-1} at different temperatures. A significant change in the spectral slice is only seen at higher temperature.

main peaks appear to narrow with a decrease in temperature suggesting a reduction in the range environments that exist as gelation occurs. These small features could suggest that a separate environment of the organosulfate functional groups exists as the temperature is decreased further.

Figure 3.18 shows the spectral slices along the ω_{pump} axis ($\omega_{probe} = 1205 \text{ cm}^{-1}$) of the κ -carrageenan sample at different temperatures. As the temperature is decreased from 70 to $45 \text{ }^\circ\text{C}$, there is a redshift of the maximum of the spectral slice, as well as an overall narrowing of the band. Upon a further decrease in temperature to $22 \text{ }^\circ\text{C}$, there is no significant change in

the spectral slice.[†] This suggests that there is very little change occurring in this temperature range.

Unfortunately, the time resolution of the frequency domain technique (*ca.* 1-2 ps) is approximately the same as the vibrational decay lifetime of the sample. The kinetic decay of the peaks are therefore not particularly useful. An attempt was made to use the time domain technique using the 2D-IR spectrometer in Nottingham, however, there was little to no mid-IR output in the *ca.* 1250 cm⁻¹ region. Figure 3.19 shows the mid-IR laser output set at a range of wavelengths. A sharp decrease in intensity at *ca.* 1250 cm⁻¹ is clearly evident. This prevented repeating these measurements with a better time resolution.

[†]The overall band shapes of the spectral slices do appear slightly abnormal.

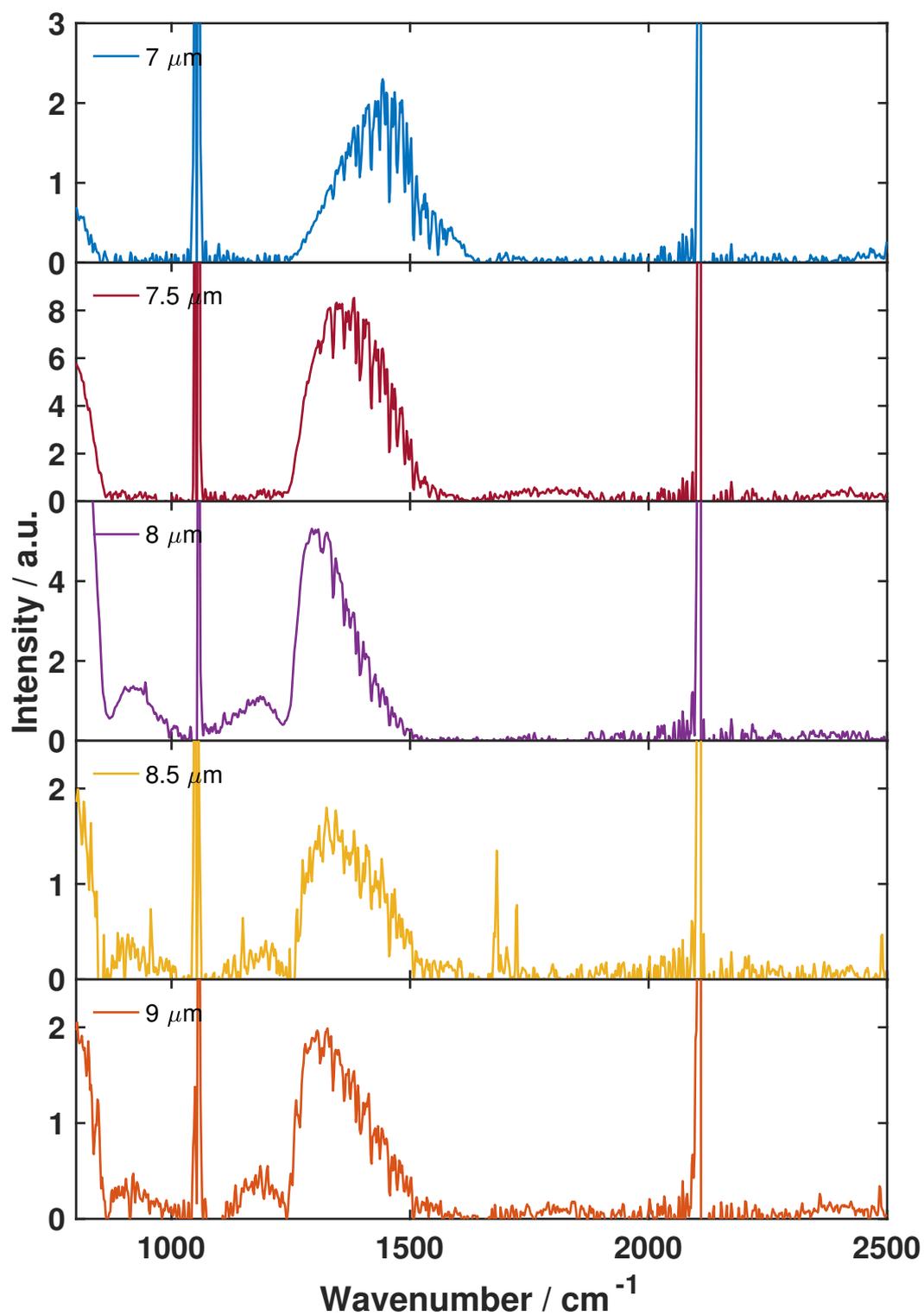


Fig. 3.19 Mid-IR spectral output at different tuning wavelengths measured using an FTIR spectrometer. Each spectrum represents a different target wavelength. At *ca.* 1250 cm^{-1} , a large intensity drop can be observed indicating the OPA and DFG output is affected in this region.

3.2.3 Solvent dynamics

In order to assess what effect gelation had on the solvent environment for carrageenan, an anionic probe molecule, ferrocyanide, was used. Carrageenan is an anionic polymer, therefore in theory, this results in the probe molecule being immersed in the solvent environment and not near carrageenan chains. 4 % w/w ι and κ -carrageenan solutions were made with a potassium ferrocyanide concentration of *ca.* 0.18 M. The high concentration of carrageenan was selected as it was thought this was the best opportunity to observe changes in the solvent environment with gelation.¹¹⁸ 2D-IR spectra of the two carrageenan solutions and water (as a control) were measured in the CN stretching region using the time domain technique at a range of temperatures (20 - 80 °C). Unfortunately, due to limited experimental time, experiments had to be performed over multiple days and could not be performed in one day. As a result a decision was made to collect spectra with increasing, instead of decreasing, temperature as it had been done previously.

Hysteresis does occur with carrageenan samples, for example the gel-sol and sol-gel transition temperatures for 4 % κ -carrageenan samples have been measured at 45 and 61 °C, respectively.¹¹⁹ The impact of different cations and anions has been extensively studied with carrageenan and so adding ferrocyanide does have the potential to affect the gelation.¹²⁰

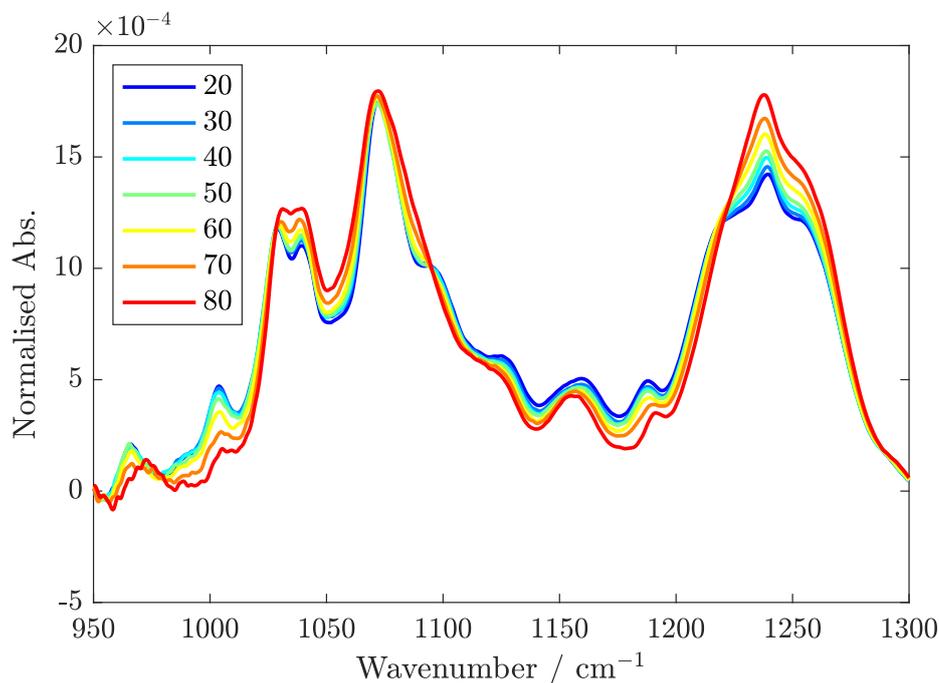


Fig. 3.20 FTIR spectra of 4 % w/w ι -carrageenan solutions with 0.18 M ferrocyanide at a range of temperatures (20 - 80 °C). Significant changes with temperature are mainly observed at *ca.* 1250 cm^{-1} which represent the organosulfate groups.

FTIR Spectroscopy

The gelation of carrageenan is dependent on the salt type as well as the concentration of the sample. For example, for κ -carrageenan the strength of gel formed is reduced in the presence of iodide ions.⁹² 4 % κ -carrageenan has a different sol-gel transition temperature to a concentration of 1 %.¹¹⁹ For this reason, it was important to assess the gelation of carrageenan at the higher 4 % concentration as well as with the added ferrocyanide in the solution. FTIR spectroscopy in combination with multivariate analysis was used to assess the gelation of carrageenan. Calum Welsh developed this technique which used MCR to assess the sol-gel transition temperature as well as the molecular rearrangements that occur during this process.¹¹³

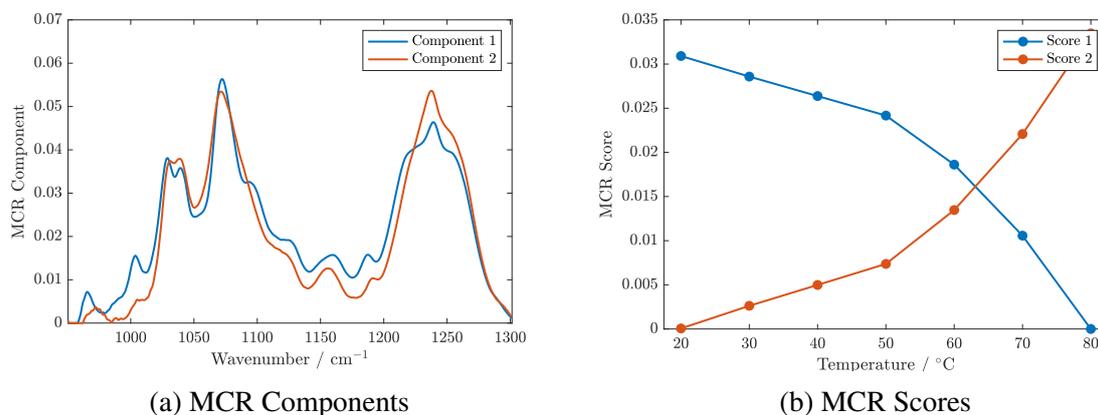


Fig. 3.21 MCR components and scores from the MCR analysis of *t*-carrageenan FTIR spectra. Scores show gelation occurring at *ca.* 65 °C.

The FTIR spectra are shown in Figure 3.20 and 3.22 for *t*- and κ -carrageenan, respectively. They both show similar changes to those observed in [113] for 2 % concentration solutions without ferrocyanide present. There are however a few major differences: firstly, the temperature is increased here instead of decreased (assessing the gel-sol transition instead of the sol-gel transition), secondly, these spectra were measured using a variable temperature Harrick cell in transmission mode as opposed to ATR-FTIR spectroscopy. The two MCR factors (gel and sol states) extracted for both types of carrageenan are consistent with those in [113], with changes observed in the organosulfate region and glycosidic linkage regions (Figures 3.21 and 3.23). The sol-gel transition was defined as the temperature when the two MCR scores cross, it was assumed to be the same for the gel-sol transition. For *t*-carrageenan this was 62 °C and for κ -carrageenan it was 50 °C. This is consistent with the literature of 4 % solutions and gel-sol transition temperatures (for κ -carrageenan it is 61 °C and for *t*-carrageenan a higher temperature is expected).¹¹⁹ Based on the FTIR spectra and MCR analysis, it was concluded that ferrocyanide does not have a significant impact on the gelation

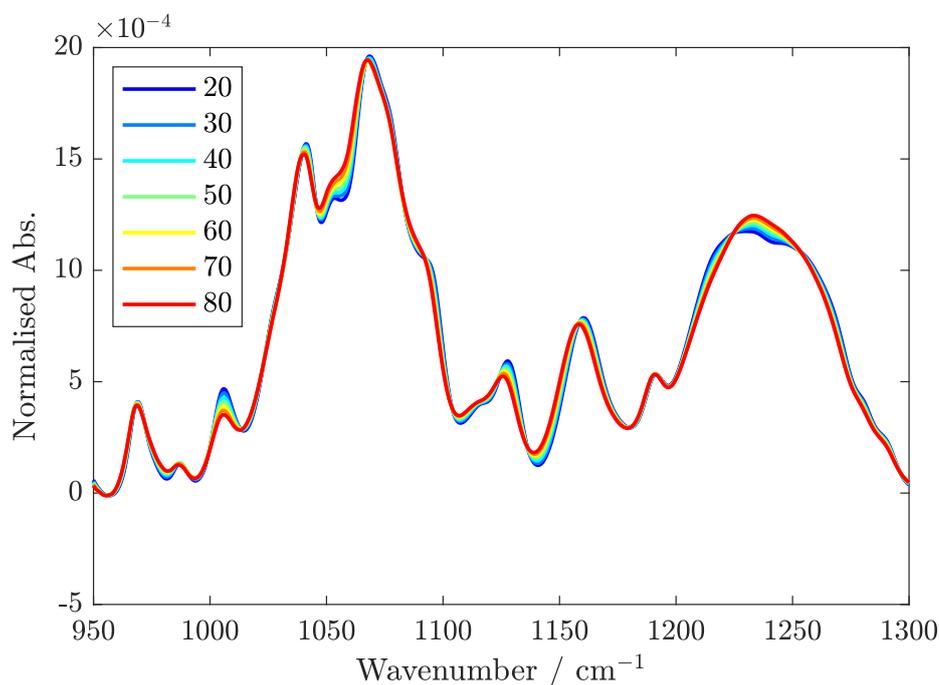


Fig. 3.22 FTIR spectra of 4 % w/w ι -carrageenan solutions with 0.18 M ferrocyanide at a range of temperatures (20 - 80 °C). Again, the most significant changes with temperature are mainly observed at *ca.* 1250 cm^{-1} which represent the organosulfate groups.

process of carrageenan. This is thought to be because there is already a significant KCl concentration present within the carrageenan sample which is integral to the gelation process. It is also concluded that transmittance FTIR spectroscopy can also be used with MCR analysis instead of ATR-FTIR spectroscopy, although the experiments are more time consuming and the temperature resolution is also significantly reduced (from 0.1 °C to > 1 °C).¹¹³

Looking at the FTIR spectra in the CN stretching region for the carrageenan samples (Figure 3.24), there is very little change observed in the peak widths or maximum peak position as the temperature is changed. This suggests that the solvent environment is not changing significantly over the temperature range based on FTIR spectroscopy.

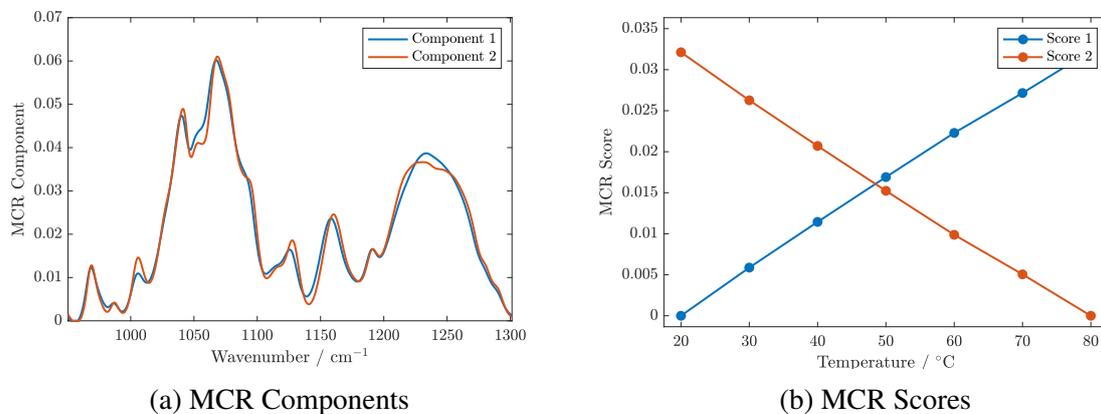


Fig. 3.23 MCR components and scores from the MCR analysis of κ -carrageenan FTIR spectra. The scores indicate gelation is occurring at *ca.* 55 °C.

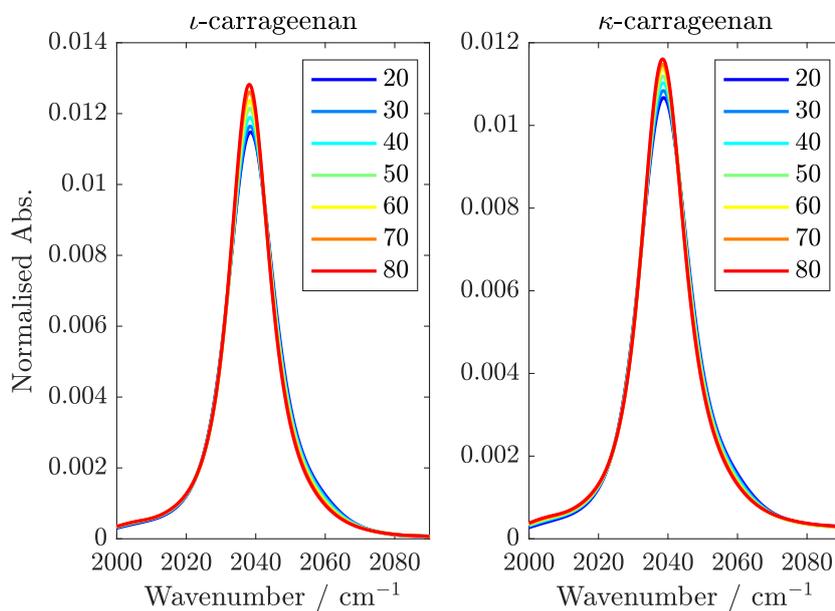


Fig. 3.24 FTIR spectra of 4 % w/w ι and κ -carrageenan solutions with 0.18 M ferrocyanide at a range of temperatures (20 - 80 °C) in the CN stretching region. No significant variation is observed with temperature for the FTIR spectra.

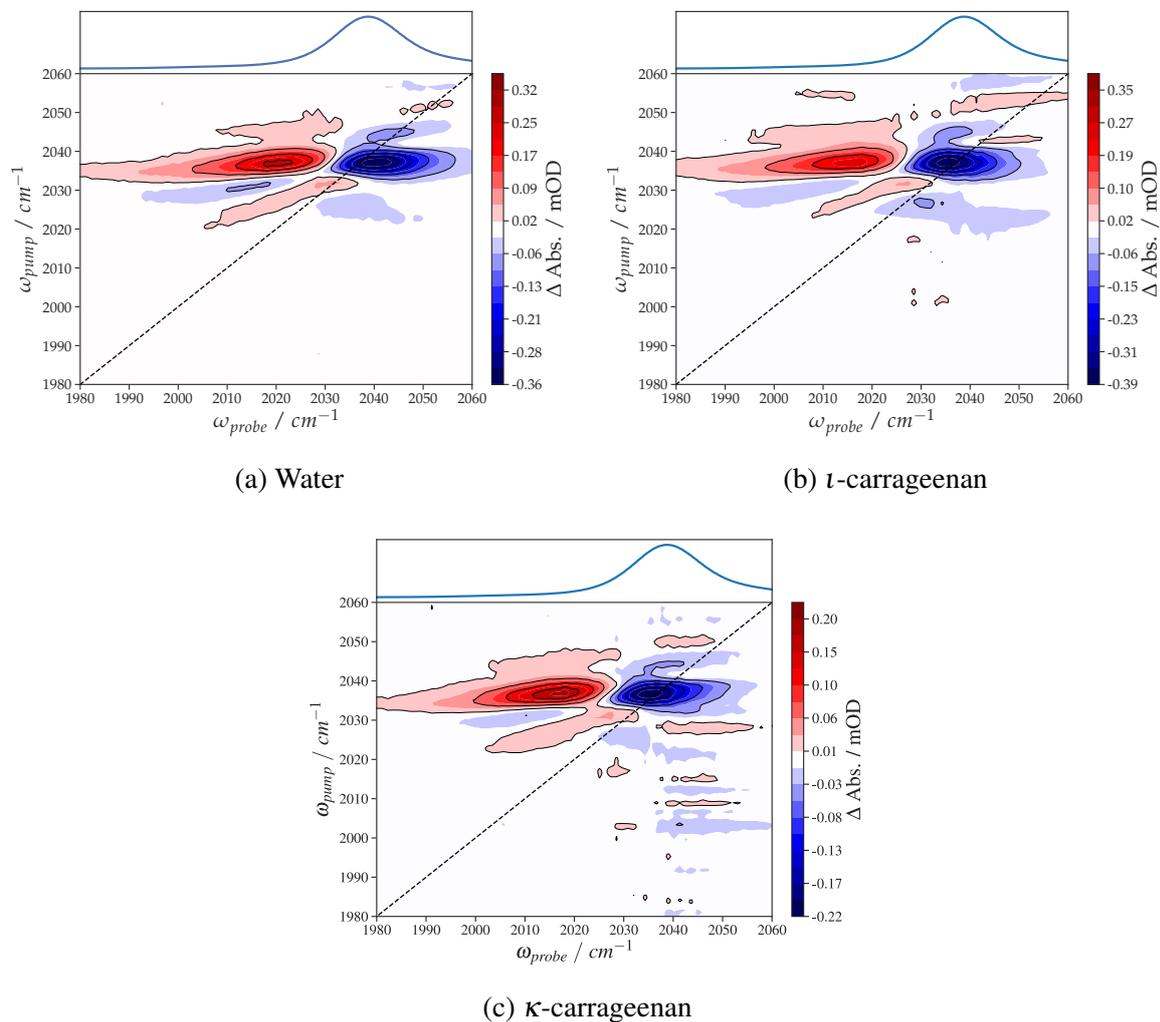


Fig. 3.25 2D-IR spectra of ferrocyanide in different solutions at a waiting time of 0.5 ps. No significant differences are observed between the different solutions.

2D-IR

2D-IR is a much more sensitive technique than linear FTIR spectroscopy, it also provides a lot more information on the dynamics and structure of a system. The dynamics of the CN stretch has been analysed in order to assess the solvent environment in carrageenan solutions when compared with water. Ferrocyanide has been used as the probe molecule here and has been studied with 2D-IR spectroscopy before.

Figure 3.25 shows 2D-IR spectra of the CN stretch of ferrocyanide in (a) water, (b) 4 % ι -carrageenan and (c) 4 % κ -carrageenan. They all contain the blue $\nu_{0 \rightarrow 1}$ and red $\nu_{1 \rightarrow 2}$ transitions and appear very similar except for scatter artefacts like the FTIR spectra.

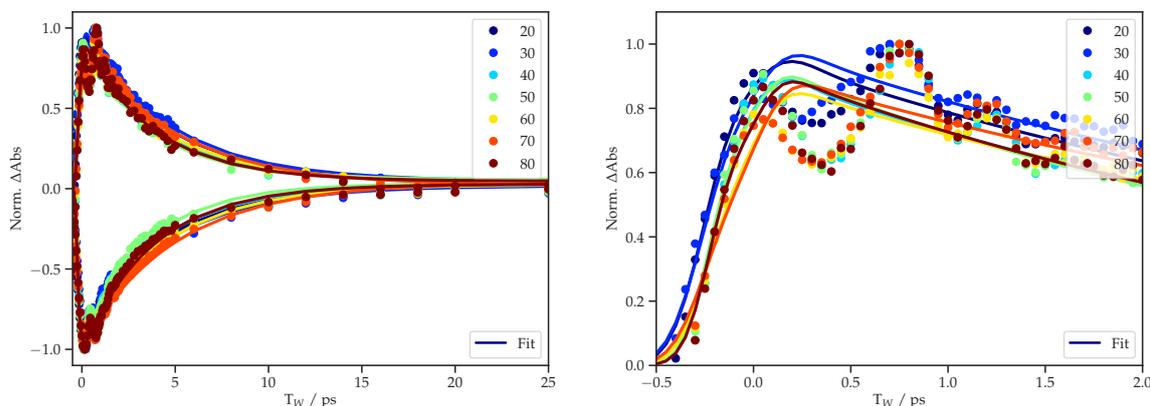
Water

2D-IR spectra of ferrocyanide in water was measured at a range of temperatures: 20, 30, 40, 50, 60, 70, 80 °C as a reference. The vibrational lifetime of both peaks in the spectra was measured by globally fitting (with a fast and slow factor *ca.* 0.3 and 4 ps) the peak intensities as shown in Figure 3.26. An Arrhenius plot was used to assess how these lifetimes varied with temperature (shown in Figure 3.27a) and no significant effect was observed for the long decay component. This was probably due to the solvent only having a minor effect which is likely driven mostly by IVR. At early waiting times, large oscillations were observed in the vibrational decay curves (Figure 3.26b). The fits of the decay curves were subtracted from the data and a FT was performed on the residual data, this is shown in Figure 3.27b where the power spectrum is displayed. There are two main regions of intensity at *ca.* 60 and 125 cm^{-1} which do not vary in frequency over the range of temperatures. It is therefore assumed that these are not due to solvent fluctuations, and are instead related to IVR of lower energy vibrational modes of ferrocyanide. Very similar behaviour was observed for both types of carrageenan.

Vibrational decay times were not analysed further for the carrageenan samples due to the invariance with temperature. Instead, spectral diffusion was used as the main spectral probe.

Spectral diffusion provides information about the solvent environment and fluctuations occurring. There are multiple methods for measuring spectral diffusion, in this case ivCLS was found to be the best method due to the noise and scatter artefacts in the spectra. Figure 3.28 shows the ivCLS plotted on top of the 2D-IR spectra (specifically the $\nu_{0 \rightarrow 1}$ peak) at various waiting times and two temperatures 20 and 80 °C. The gradient can be observed decreasing as the waiting time of the spectrum increases. Figure 3.30 is a plot of the gradient of the ivCLS versus waiting time with a respective exponential decay fit at each temperature. Although it is difficult to observe changes between specific temperatures there is a definite trend of the decay times getting shorter as the temperature increases. This is consistent with solvent fluctuations becoming more frequent at higher temperatures, since more energy is available to break H-bonding within the solvent network.

Using the Arrhenius equation (eq. 3.1), the rate of decay was plotted against the inverse temperature and a linear fit was performed, which is shown in Figure 3.29. The gradient of the linear fit is equivalent to the activation energy (E_a) of the process and the inverse of the pre-exponential factor (A) can be used to estimate the time taken to cross energy barrier. E_a was determined to be 7.6 kJ mol^{-1} which is within but on the slightly lower side of literature values of between 6.3 and 20.5 kJ mol^{-1} .^{121,122} The lower estimation, compared with a previous value calculated using 2D-IR, could be due using ferrocyanide as the probe molecule. The large negative charge of ferrocyanide could distort the local structural fluctuations leading to a disruption of the local H-bonding network and therefore an underestimation.^{123,124} Using the pre-exponential factor, the timescale to cross the energy barrier is estimated to be 32 fs which is close to the 40 fs libration mode of water.¹²⁵



(a) Long waiting times for both peaks with respective fits. (b) Short waiting times showing oscillations in intensity of the $\nu_{1 \rightarrow 2}$ peak.

Fig. 3.26 Vibrational decay curves for the $\nu_{0 \rightarrow 1}$ and $\nu_{1 \rightarrow 2}$ peaks and the corresponding global fits as a solid lines and all at a range of temperatures from 20 - 80 °C. Oscillations were observed at early times. These were analysed further by calculating power spectra which is shown later.

$$\ln(1/\tau) = \ln(A) - \frac{E_a}{RT} \quad (3.1)$$

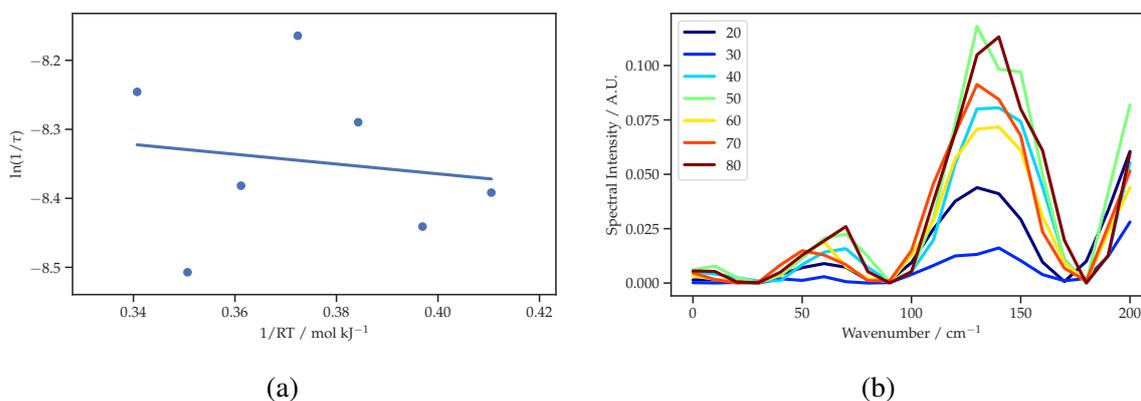


Fig. 3.27 (a) A plot of $\ln(1/\tau)$ versus $1/RT$ showing the temperature dependence of the vibrational lifetime for ferrocyanide in water. No Arrhenius behaviour was observed for the vibrational lifetimes and so this method was not used. (b) Power spectrum of the oscillations observed in the $\nu_{1 \rightarrow 2}$ peak intensity at early T_W and at a range of temperatures from 20 - 80 °C. Again no obvious temperature dependence was observed with the power spectra and so this was not investigated further.

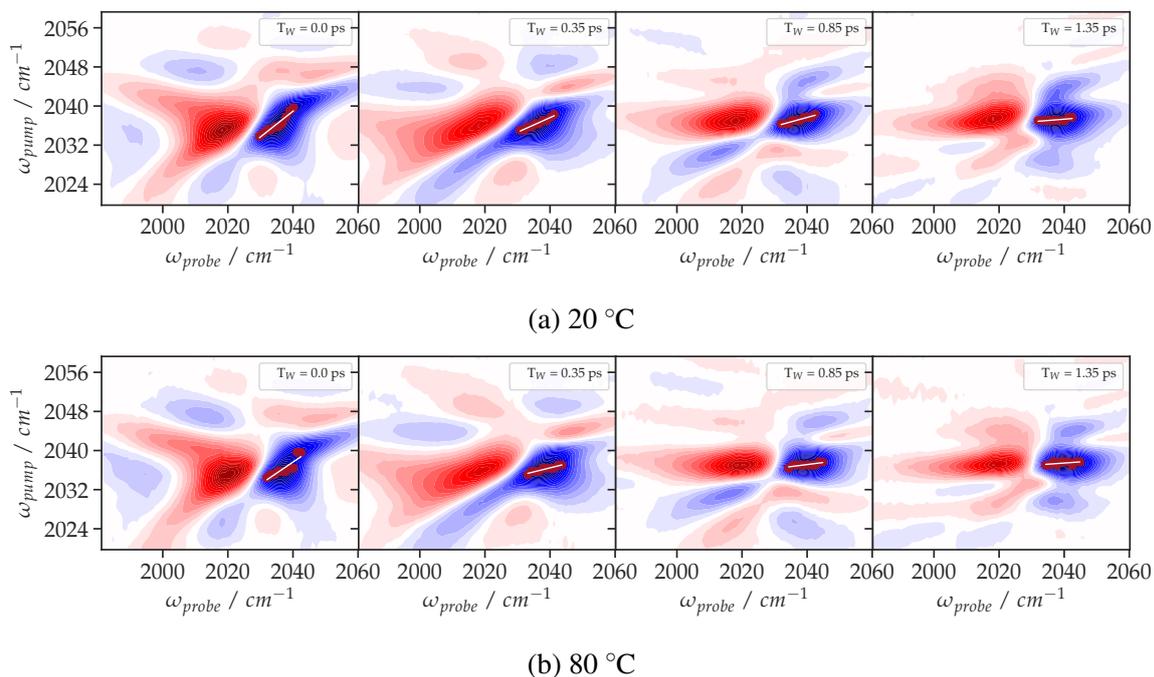


Fig. 3.28 2D-IR spectra of ferrocyanide in water at 20 and 80 °C and at different waiting times with the ivCLS plotted in red and the corresponding fit as a white line. The gradient of the white line is used as a measure of spectral diffusion.

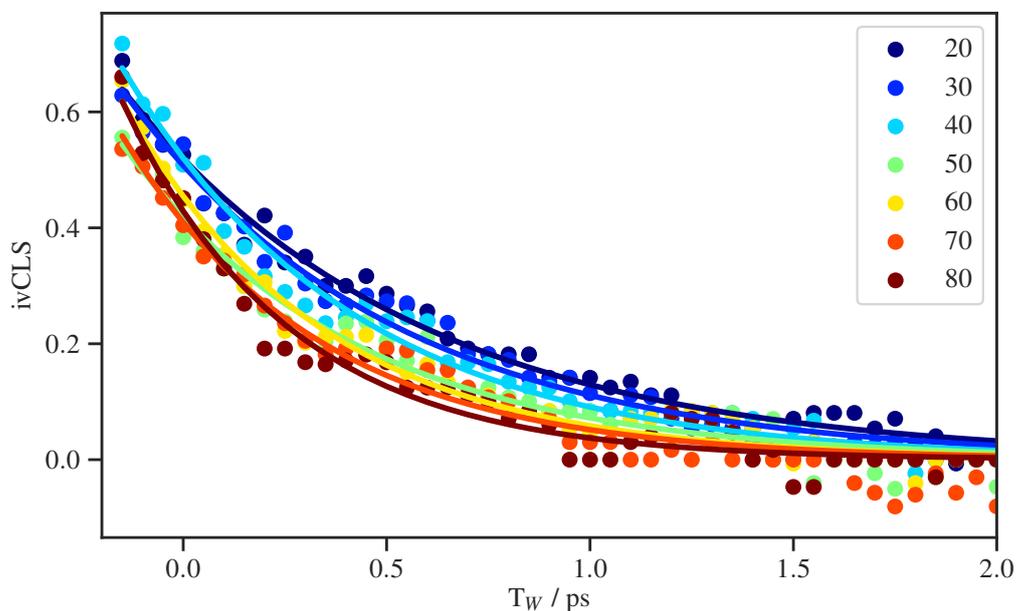


Fig. 3.29 Plots of the ivCLS gradient versus waiting time, T_W , at a range of temperatures along with their respective biexponential decay fits for ferrocyanide in water. The rate of spectral diffusion increases with increasing temperature.

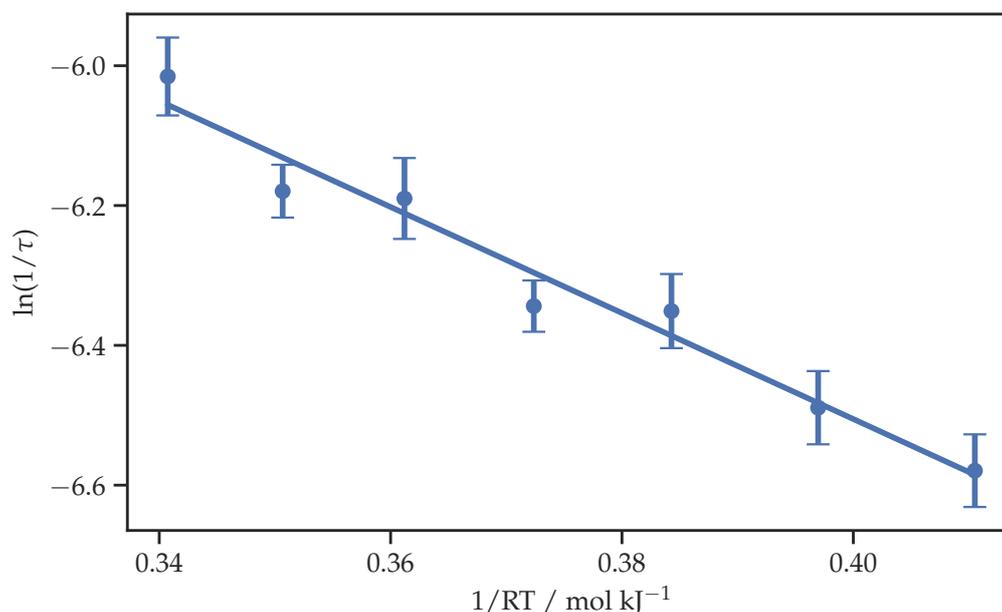


Fig. 3.30 A plot of $\ln(1/\tau)$ versus $1/RT$ showing the temperature dependence of the spectral diffusion rate constant for ferrocyanide in water. The error bars are propagated from the ivCLS fits. The linear fit was used to calculate the Arrhenius behaviour.

***t*-carrageenan**

The ivCLS measured at various waiting times for ferrocyanide in 4 % *t*-carrageenan looks similar to that of water, and is shown in Figure 3.31. The gradient of the ivCLS versus T_W is plotted with respective exponential decay fits in Figure 3.32 for each temperature from 20 - 80 °C. Although the data is noisier there is still a general trend of the decay times getting faster as the temperature increases. This is evident in the Arrhenius plot which is shown in Figure 3.33 where Arrhenius behaviour is displayed. Ideally this would have been fitted using a Vogel-Fulcher type equation (eq. 3.2) to analyse the gel-sol transition,¹²⁶ however, the noise makes this difficult, coupled with the fact that there is a high gel-sol transition temperature (estimated from FTIR) of *ca.* 60 °C and low temperature resolution which

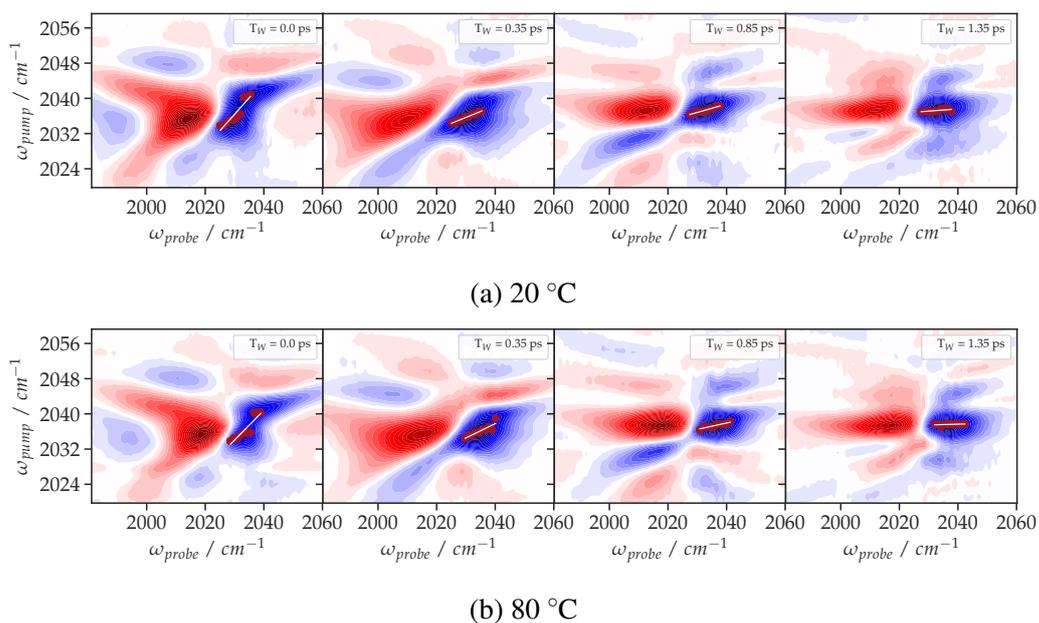


Fig. 3.31 2D-IR spectra of ferrocyanide in ι -carrageenan at 20 and 80 °C and at different waiting times with the ivCLS plotted in red and the corresponding fit as a white line, which are used to calculate the spectral diffusion rate.

impedes this. The E_a calculated from the Arrhenius plot was determined to be 4.0 kJ mol^{-1} which is approximately half that of what was calculated for water. The barrier crossing time was determined to be substantially longer than water at 123 fs.

$$\ln(1/\tau) = B e^{-E_A/R(T-T_0)} \quad (3.2)$$

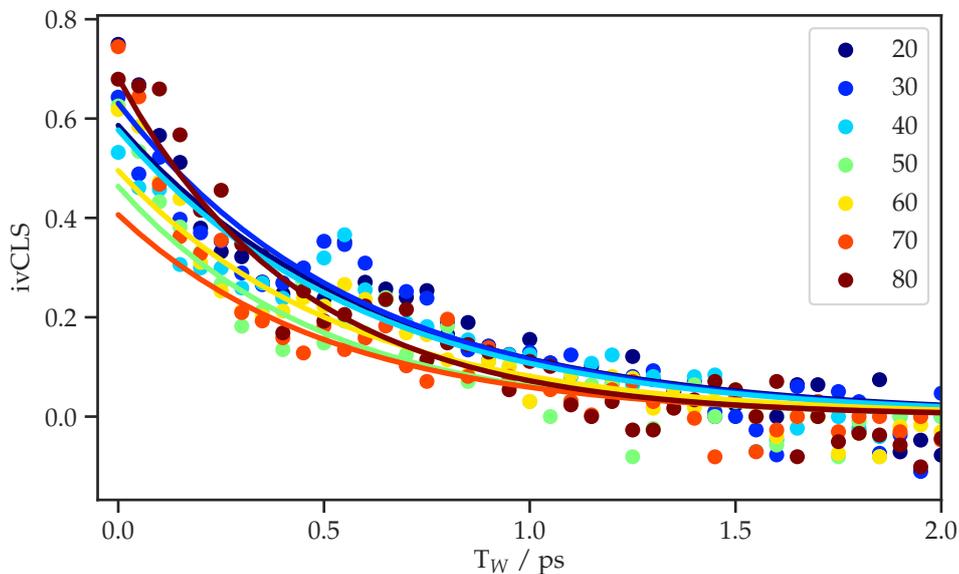


Fig. 3.32 Plots of the ivCLS gradient versus waiting time, T_W , at a range of temperatures along with their respective exponential decay fits for ferrocyanide in t -carrageenan. The temperature dependence of the spectral diffusion rate is less clear than that of ferrocyanide in water, however, there is still a general trend of increasing spectral diffusion rate with temperature.

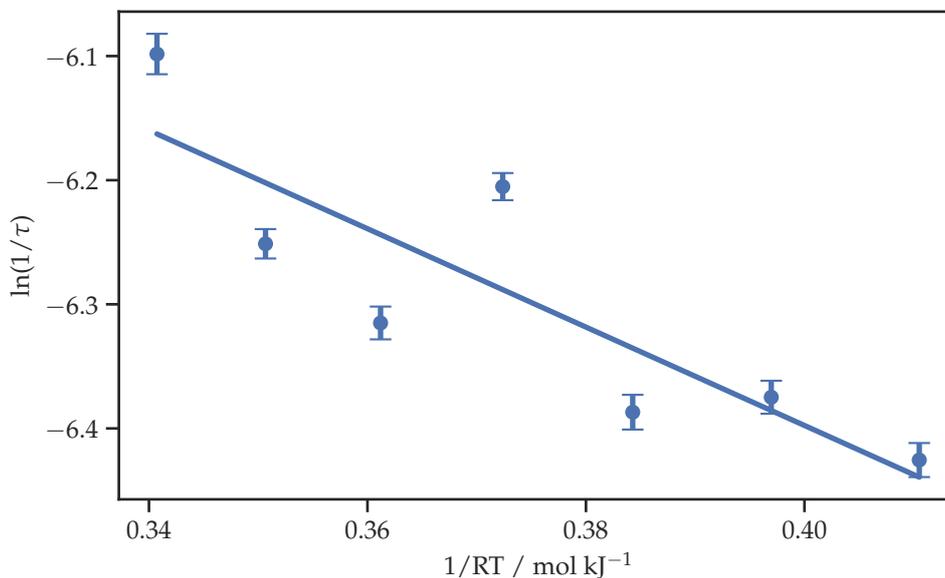


Fig. 3.33 A plot of $\ln(1/\tau)$ versus $1/RT$ showing the temperature dependence of the spectral diffusion rate constant for ferrocyanide in t -carrageenan. The error bars are propagated from the ivCLS fits. The linear fit was used to calculate the Arrhenius behaviour.

κ -carrageenan

Looking at κ -carrageenan, Figure 3.34 shows the ivCLS measured at various waiting times for ferrocyanide in κ -carrageenan. The ivCLS gradient versus T_W is plotted at a range of temperatures with respective exponential decay fits in Figure 3.35. The Arrhenius plot of the spectral diffusion against temperature is shown in Figure 3.36, which is very similar to ι -carrageenan with a substantial amount of noise but still clearly displaying Arrhenius behaviour. The E_a was again determined to be approximately half that of water with a value of 3.8 kJ mol^{-1} . The barrier crossing time was calculated to be approximately the same as ι -carrageenan with a value of 147 fs.

All calculated values relating to the Arrhenius plots is summarised in Table 3.3.

Table 3.3 The fitted parameters obtained from fitting the Arrhenius equation to the temperature dependent plots for each of the samples.

Sample	1/A / fs	E_a / kJ mol^{-1}
Water	32 ± 1	7.6 ± 0.7
ι -carrageenan	123 ± 1	4.0 ± 1.1
κ -carrageenan	147 ± 1	3.8 ± 1.4

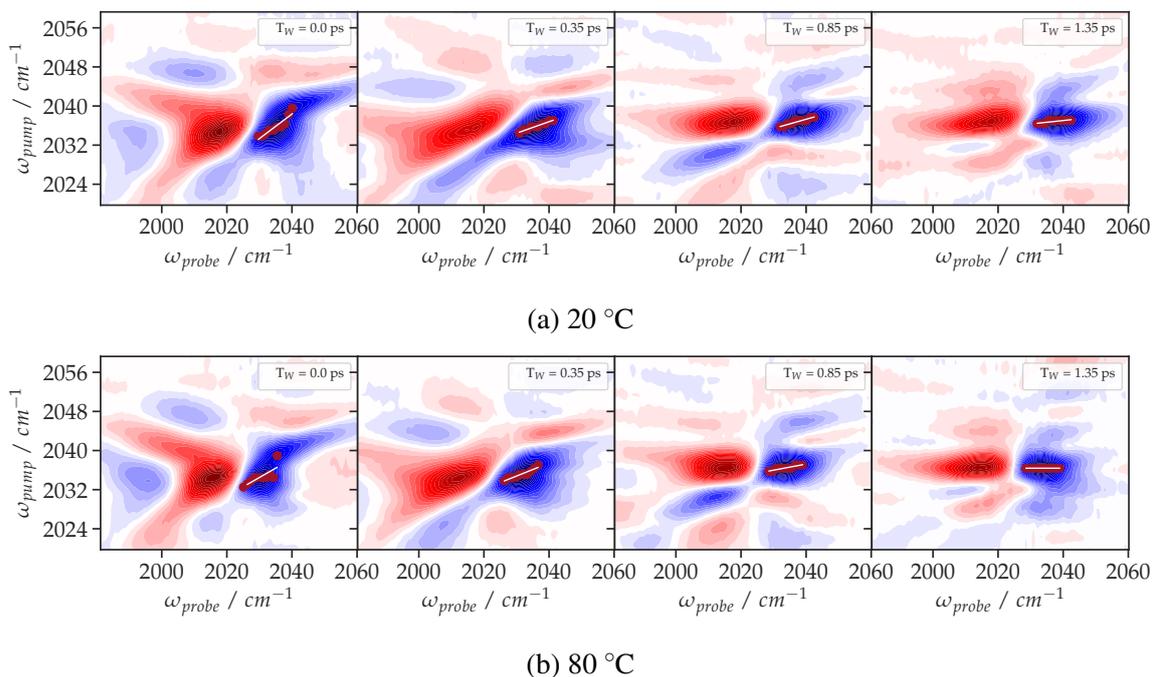


Fig. 3.34 2D-IR spectra of ferrocyanide in κ -carrageenan at 20 and 80 °C and at different waiting times with the ivCLS plotted in red and the corresponding fit as a white line, which are used to calculate the spectral diffusion rate.

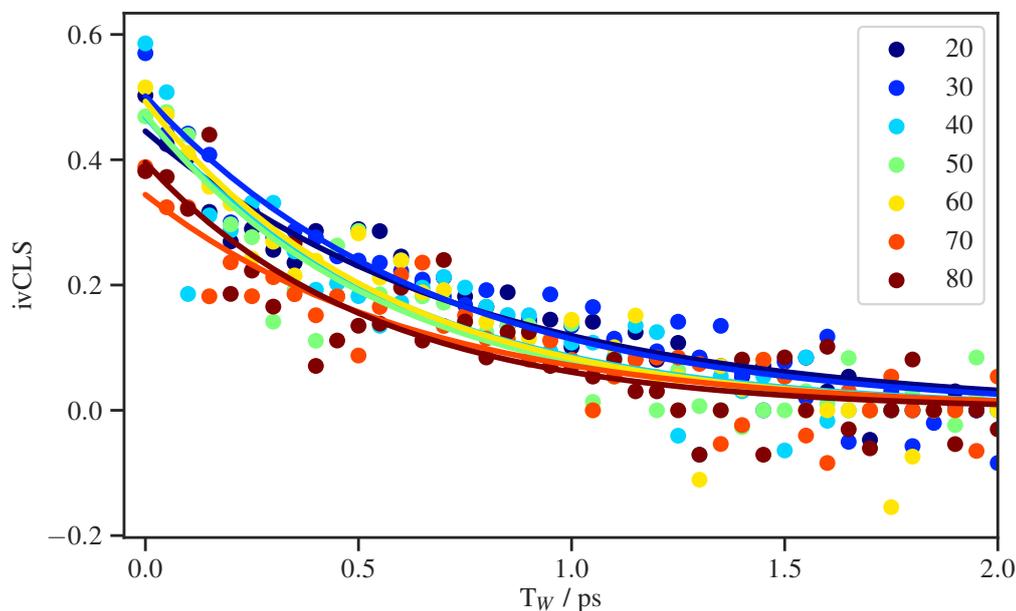


Fig. 3.35 Plots of the ivCLS gradient versus waiting time, T_W , at a range of temperatures along with their respective exponential decay fits for ferrocyanide in κ -carrageenan. The temperature dependence appears similar to that of ι -carrageenan.

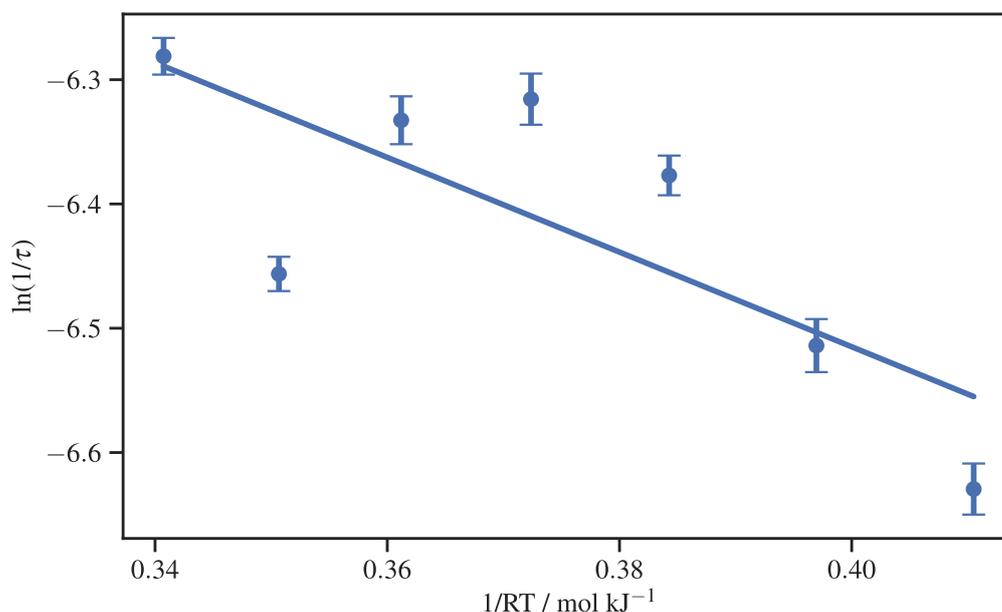


Fig. 3.36 A plot of $\ln(1/\tau)$ versus $1/RT$ showing the temperature dependence of the spectral diffusion rate constant for ferrocyanide in κ -carrageenan. The error bars are propagated from the ivCLS fits. The linear fit was used to calculate the Arrhenius behaviour.

3.3 Discussion

3.3.1 Carrageenan gelation

κ -carrageenan

Looking specifically at κ -carrageenan 2D-IR spectra, as the temperature decreases there is a general narrowing of the band as well as an overall redshift in the frequency. The narrowing is indicative of a reduction in the large range of environments that the organosulfate functional groups are in. An overall redshift indicates a weakening of the S=O bond strength, which is likely related to increased interactions between the organosulfate groups and stabilising cations as the temperature decreases. The cations would result in a reduction in the electron

density around the functional groups.¹¹⁶ Apart from the overall narrowing, a small peak appears and becomes more prominent as the temperature decreases. This suggests that there is a small subset of organosulfate groups which exist in a significantly different environment from the rest.

Tako *et al.* proposed cation interactions based on rheological experiments and Makshakova *et al.* further supported this with DFT calculations.^{85,116} The interaction with cations are summarised in Figures 3.5 and 3.37 where cations are shown to interact with more than one functional group and between chains. The small peak, significantly redshifted, observed could possibly be due to exclusive interaction between a cation and organosulfate group. This would explain the large frequency shift.

Spectral slices show little change for κ -carrageenan between 45 and 22 °C suggesting limited molecular changes are occurring during this temperature decrease. During this temperature window super helical structures are thought to form. The lack of change occurring could be because IR spectroscopy is not sensitive to super helices forming, which is possible because the major molecular rearrangements are occurring earlier on the gelation process when the helices are formed and not when these helices aggregate.

***ι*-carrageenan**

Very similar overall trends to κ -carrageenan are observed for *ι*-carrageenan with narrowing of peaks along with a slight redshift in frequency. This again suggests a smaller range of environments and interaction of ions. Unlike κ -carrageenan, cross peaks become very evident

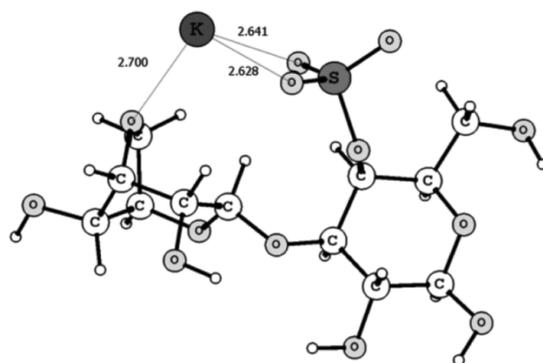


Fig. 3.37 Structure of κ -carrageenan monomer unit with a K^+ cation interacting with the organosulfate and ether functional groups. Reproduced from [116]

at 22 °C which indicates as the gelation occurs the organosulfate groups begin to interact more and because these are cross peaks it suggests that these interacting groups are from intertwined strands of carrageenan. This interaction could occur from two different sources, either intra-monomer or inter-monomer vibrational energy transfer. Like the mechanism proposed by Tako *et al.* for κ -carrageenan, a similar mechanism was proposed for ι -carrageenan except these experiments were performed with Ca^{2+} cations which is thought to strengthen the gel formed by stabilising the higher density of charge on each monomer unit.⁸⁵ In this case the cation is predominantly K^+ but it is assumed to have a similar overall effect. The inter-monomer energy transfer could be mediated via the cation. Alternatively, the intra-chain energy transfer would be mediated via the covalent bonding present within the monomer unit. While it is likely that both routes contribute, the intra-monomer energy transfer would likely be the dominant one.

The cross peaks are particularly evident when spectral splices are extracted at each temperature. At 22 °C, three peaks are clearly evident which suggest three main environments for

the organosulfate groups to be present. Based on a technique called 2D-COS, Calum Welsh assumed there were only two main environments, however, based on the evidence presented above it is clearly three.¹¹³ It is assumed each environment is defined by the interaction with cations. These have been tentatively assigned. Starting at lowest wavenumber is an organosulfate group interacting exclusively with one cation, then two organosulfate groups both interacting with a cation and finally a group of organosulfate in a small region interacting with multiple cations without specific binding but with the charge density more diffuse.

3.3.2 Solvent environment

Spectral diffusion (frequency fluctuations) provide insight into local solvation structures/shells and their dynamics. The magic angle condition was used to measure spectra. This minimises the rotational component of the chromophore in the dynamics. It is possible to extract this rotational component by calculating the anisotropy by collecting both parallel and perpendicular spectra. This allows it to be separated from the solvent fluctuations. For these experiments it was assumed that the rotational time for the ferrocyanide was long enough that it did not sufficiently contribute to the FFCF, especially, considering spectra were measured at the magic angle.¹²³

Water

For water, the activation energy determined is at the lower end of the range stated in the literature. This could be due to using a different probe molecule which couples to different

solvent modes and fluctuations and the large charge of ferrocyanide. There is also the concentration of salt in solution which would also affect the solvent fluctuations but is unlikely to have such a large effect at these concentrations.¹²⁷ Nonetheless the activation energy is reasonable considering the data.

Vibrational decay lifetimes in 2D-IR spectra are generally a much less suitable measure for Arrhenius behaviour because they are significantly less sensitive. This is because they are impacted by processes such as IVR which can still be dominant in solution.¹²¹ In this particular case, vibrational lifetimes are a bad measure considering the noise levels and scatter artefacts in the spectra along with the IVR processes that take place which are likely dominant initially.

Carrageenan

When assessing the results of the solvent state experiments, viscosity is an obvious consideration that must be taken into account, since carrageenan solutions are significantly more viscous than water. The timescales observed in carrageenan solutions when compared with water do not show any major differences. Often, with higher viscosity, it is expected vibrational dynamics slow down, which was not the case here.¹²⁶ This is possibly explained because the viscosity for gels is a macroscopic property. In this case we used a probe molecule which is measuring the microscopic solvent environment. Carrageenan gels can be thought of as a heterogeneous system where one part is the carrageenan strands that provide the structure of the gel and the other parts are ‘pockets’ of solvent in between entangled or

aggregated carrageenan strands. The charge of the probe molecule likely causes it to ‘sit’ in these pockets of solvent and so does not ‘experience’ the macroscopic viscosity property. Therefore, because of the gels heterogeneous nature, the viscosity of the gel is not taken into account.[‡]

In water, the timescale of H-bond reorientation is predominantly due to large angular jumps where new H-bonds are formed. This is the basis of the H-bond jump model which is responsible for the main solvent fluctuations in water.¹²⁸ There is also a minor component where H-bonds slowly rotate intact. The pre-exponential factors calculated and converted to barrier crossing timescales above are summarised in Table 3.3. They are significantly slower for carrageenan than water. This suggests that the process of reforming H-bonds takes longer because the reorientation time is much longer. This could possibly be due to the lack of available H-bond donors in the solvent environment because of a more structured H-bonding network. However, the activation energy of carrageenan solutions were determined to be approximately half that of water which would discourage this theory as the energy required to disrupt a more structured H-bond network would surely be greater and not less than that of water. Instead, the disruption of the H-bonding network is more likely the cause. Interfacial water molecules are known to have both, slower and weaker H-bonding.¹²⁹ This could explain both the barrier crossing time increasing and the weaker H-bonds observed in carrageenan solutions. The nanostructure of the carrageenan gelation networks could result in small regions of confined water. These confined regions would have disrupted H-bonding networks due to an increase in the number of interfacial water molecules. Going

[‡]Although at high temperatures (>60 °C) the sample is in the sol state where the heterogeneity is blurred. However, in this case there are too few data points from the low temperature resolution.

forward these experiments could act as a benchmark for future experiments, where the frozen carrageenan solutions are investigated looking specifically at the behaviour of ice crystals and how carrageenan affects this.

3.4 Conclusion and future work

3.4.1 Conclusion

2D-IR spectroscopy has been utilised to investigate the gelation process of κ -carrageenan and ι -carrageenan by directly probing the organosulfate groups. As the temperature is reduced from 70 to 22 °C, there is a narrowing and overall redshift of peaks which suggest a reduction in the range of environments occupied by carrageenan molecules, along with a greater interaction between the functional groups with stabilising cations. In ι -carrageenan specifically, there are clear cross peaks observed which firstly suggest there are three main conformations for the organosulfate groups and secondly that these different environments interact – most likely within each monomer unit. This is not observed clearly with κ -carrageenan which further supports the intra-monomer vibrational energy transfer for ι -carrageenan. At low temperatures for both types of carrageenan, a small separate peak is observed at low frequency. This is thought to be a result of exclusive interactions between one organosulfate group and a cation.

Using a combination of FTIR spectroscopy and 2D-IR spectroscopy with the probe molecule ferrocyanide, the effect of carrageenan on water dynamics was investigated. Arrhenius plots

were generated by measuring the spectral diffusion of ferrocyanide at a range of temperatures. This resulted in the activation energy for water molecule H-bonding in carrageenan to be approximately half of that of pure water, indicating that weaker H-bonds are formed in the presence of carrageenan which is probably due to the disruption of large H-bonding networks. No difference was discernable between the two types of carrageenan. This was limited by the errors associated with these measurements. The timescales of H-bond reformation was estimated to be *ca.* 130 fs while for water it was 30 fs which suggests reforming H-bonds is difficult. This could be due to the H-bonding networks being confined as a result of disruption from the carrageenan chains.

3.4.2 Future work

Ideally, going forward, the organosulfate IR band would be investigated using the time-domain method at a higher temperature resolution. This would provide two advantages, firstly it would provide a significantly higher time resolution which would allow the kinetics of these functional groups and cross peaks to be probed. The cross peak kinetics could provide insight into which pathway the functional groups are interacting e.g. intra or inter-molecularly. The spectral diffusion could also be probed in order to provide an understanding as to how the solvent environment around the chains change as gelation occurs. Secondly, at a higher temperature resolution, the gelation process could be tracked more closely resulting in an even greater understanding as to how the molecular rearrangements track with the macroscopic properties of the gelation process.

Salt exchange experiments could be performed to gain a greater insight into the precise role that cations perform in relation to the organosulfate groups as gelation occurs. The carrageenan samples by default contain a large amount of K^+ ions, it would be interesting to replace these with divalent cations and ions of different ionic radii, as well as varying the overall concentration. It would be interesting to see how the anionic organosulfate groups would interact differently given their large anionic charge. This would link up nicely with previous rheological experiments.¹³⁰

Unfortunately due to time and rule limits, further experiments investigating the solvent environment of carrageenan solutions with a probe molecule were not possible. Given more time, these experiments would be repeated at a lower concentration and looking specifically at the sol-gel transition. This would allow results from some of the previous experiments to be tied to specific molecular changes. The transition would in theory be more visible as it would occur in the middle of the temperature range.

Temperature-jump 2D-IR (T-jump) experiments would theoretically be possible with this system and would be advantageous given its precise temperature control with repetitive measurements. However, classical T-jump experiments have typically involved only a 10-15 °C rise in temperature to look at protein folding dynamics which occur on the millisecond timescale. This temperature jump would firstly be too small to observe significant changes in the carrageenan. Secondly, the microsecond timescale would also be too fast to detect such large macroscopic changes occurring.^{131,132}

Recently a new T-jump experiment has been designed and implemented which utilises a continuous wave laser modulated by an AOM.¹³³ This allows a much larger temperature increase to be probed at significantly longer timescales, past seconds. It would be very interesting to study the gelation of carrageenan utilising this method because many sol-gel transitions can be probed during the measurement providing much more reliable data on the process.

Ultimately these experiments have been working towards studying carrageenan at sub-zero temperatures using 2D-IR spectroscopy. Various challenges need to be overcome first, e.g. a low temperature transmission cell. Of specific interest, however, is how carrageenan interacts with ice crystals. This would link in with FTIR experiments performed by Calum Welsh,¹¹³ who found that the ι -carrageenan concentrated together as the temperature was held at -12 °C. The solvent environment in those samples would be of particular interest because there could be a small amount of phase separation occurring, which could be detected by the probe molecule. The organosulfate groups would also be in a more confined space and therefore the vibrational energy pathways could possibly change, which would again be observable using 2D-IR spectroscopy.

Chapter 4

The interactions of mucins and theaflavins

4.1 Introduction

The body produces between 0.75 and 1.5 litres of saliva a day. It serves many essential functions including: the initial process of digestion, protection of teeth and lubrication of food, allowing it to be swallowed. The majority of saliva is composed of water, but also includes enzymes, antimicrobial agents, antibodies and glycoproteins.¹³⁴ Saliva has been shown to be a non-Newtonian fluid despite it consisting mostly of water. This behaviour has been attributed to gel-forming glycoproteins.¹³⁵ Mucins are a family of large polymeric glycoproteins found in both saliva and the gastrointestinal tract, which can cross-link to form gels (although there are also non-gel forming mucins).¹³⁶ There are five gel-forming mucins:

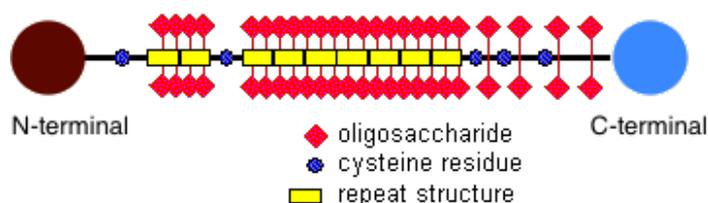


Fig. 4.1 Cartoon representation of mucin with the C and N-terminal domains labelled and the central region with glycosidic bridges. Reproduced from [138].

MUC5B, MUC5AC, MUC6, MUC2 and MUC19. They help protect the body from bacteria by aggregating around them, immobilising the bacteria via a mechanism called mucociliary clearance.¹³⁷

Figure 4.1 shows a schematic representation of a basic mucin structure. The N- and C-terminal domains (start and end sections of the protein) are shown by brown and blue circles respectively. In between the terminals are various sections made up of two main types of domain: cysteine rich and glycosidic domains. The cysteine rich domains (small purple dots) interrupt the glycosidic domains, where oligosaccharide chains (pink diamonds) are attached to the protein backbone.

4.1.1 MUC5B

MUC5B is the main gel forming mucin found in saliva and is very large in size (*ca.* 2 - 100 MDa).¹³⁹ The protein can change the visco-elasticity of saliva by varying the degree of cross-linking.¹⁴⁰ This cross linking is mediated by the cysteine rich N- and C-terminal domains by forming disulfide linkages. Using SAXS and electron microscopy, both of these domains have been shown to be globular proteins and using TERS-Raman alongside ROA

the secondary structure was found to be β -sheet.¹⁴¹ The NT5B domain has been shown to exist in both a monomeric and dimeric state in solution. The cross-linking of MUC5B is highly dependent on the pH and calcium ions, it has been shown that high concentrations of Ca^{2+} and a low pH encourage aggregation of NT5B.¹⁴² The central region of MUC5B is rich in proline, threonine and serine,¹⁴³ the last two residues are the sites of the O-linked glycosidic chains, which attach various polysaccharide chains to the protein backbone. These glycan chains have been shown to have many diverse structures via techniques such as NMR and mass spectrometry.¹⁴⁴

4.2 Tea polyphenols

Tea is an extremely popular refreshment around the world and there are many varieties. Two of the main varieties (among others) are green and black tea. They differ mainly in their harvesting and manufacturing process. For green tea, immediately after harvesting, the tea leaves are steamed and dried, this process prevents the fermentation of the leaves. Whereas for black tea, the leaves are left for a period of time in order to allow them to undergo fermentation.¹⁴⁵ Green tea contains a class of compounds called polyphenols and more specifically catechins, the structure of which are shown in Figure 4.2. During the fermentation process these catechins undergo an enzymatic oxidation process (more specifically a heterodimerisation reaction).¹⁴⁶ This results in a family of compounds called theaflavins which are characterised by the benzotropolone core (shown in Figure 4.3). Theaflavins are

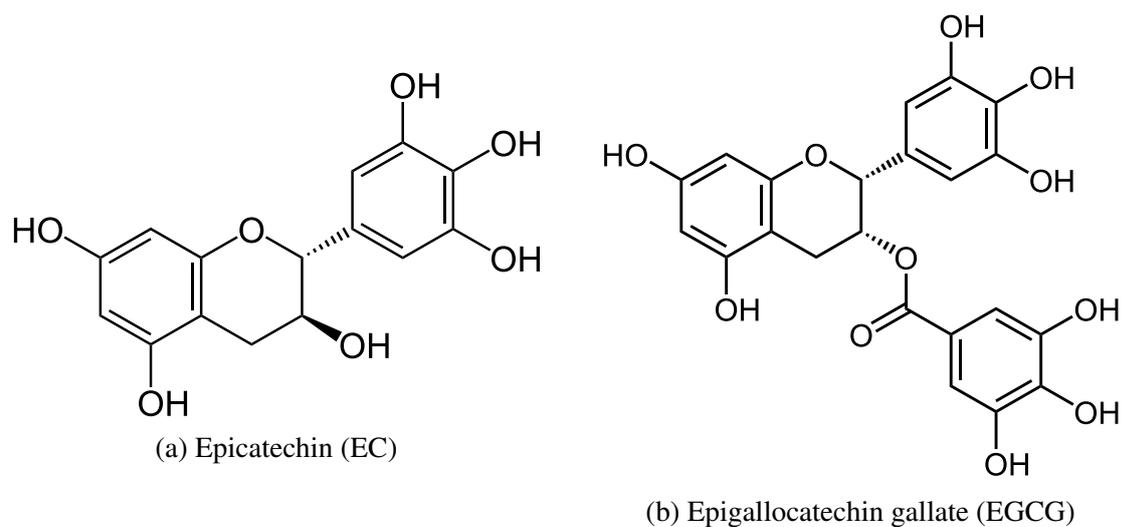


Fig. 4.2 The chemical structure of green tea catechins: Epicatechin and Epigallocatechin gallate.

a dark red colour and are one of the main reasons for tea appearing darker in colour after fermentation.¹⁴⁷

Catechins and theaflavins are secondary metabolites that act as a defence mechanism for the plants designed to discourage consumption of them,¹⁴⁸ as a result, tea has been shown to produce a bitter taste. This has been attributed to an astringency thought to be caused by the polyphenols of the tea interacting with salivary mucins and other proline rich proteins (PRP) present in saliva. It is believed that this interaction causes the proteins to cross-link and precipitate leading to a drying sensation within the mouth.¹⁴⁹

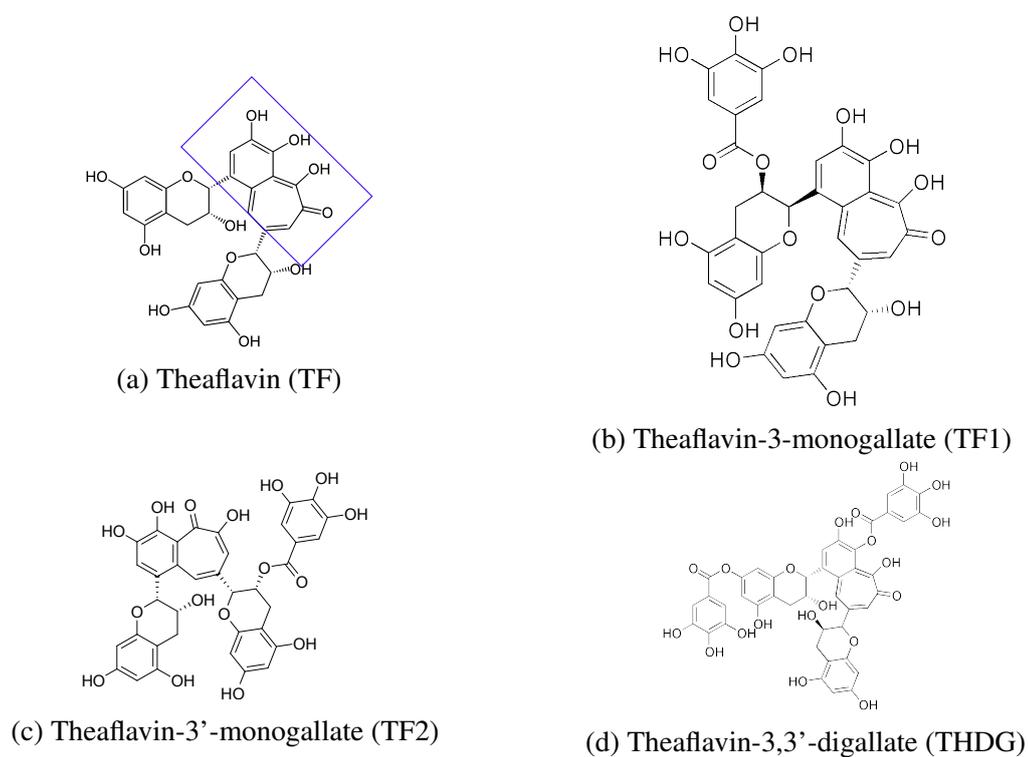


Fig. 4.3 The chemical structure of black tea theaflavins: theaflavin (TF), theaflavin-3-monogallate (TF1), theaflavin-3'-monogallate (TF2) and theaflavin-3,3'-digallate (THDG). The benzotropolone core is highlighted by the blue box.

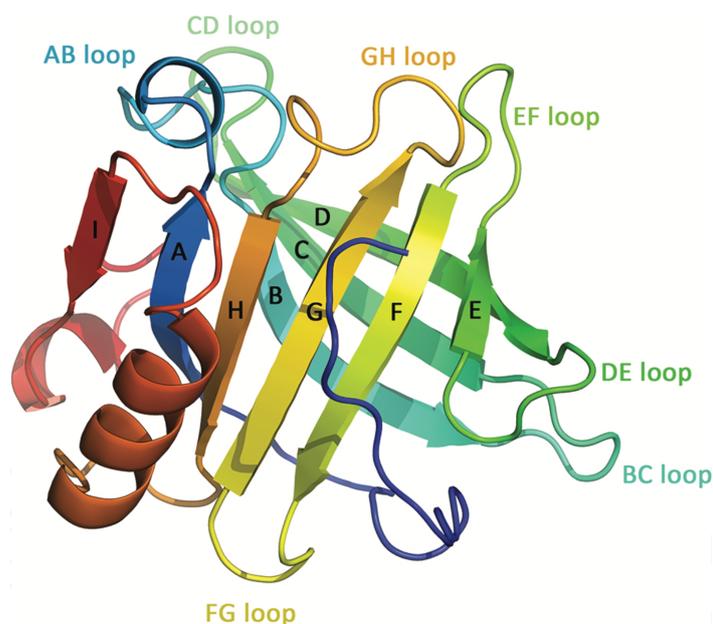


Fig. 4.4 The secondary structure of β -LG showing both the α -helix and β -sheet regions. Reproduced from [150].

4.3 Interactions between β -lactoglobulin and tea polyphenols

Milk is an ingredient which is commonly combined with tea and it has been shown that it changes the mouth feel and perceived astringency of tea.¹⁵¹ Therefore, there is great interest in the interactions between milk and tea polyphenols. For example, Ye *et al.* showed significant changes in the secondary structure of milk proteins when black tea polyphenols or green tea polyphenols were bound. Increases in β -sheet and α -helix conformations were observed and it was shown black tea polyphenols resulted in a larger change.¹⁵² β -lactoglobulin (β -LG) is the main whey protein in milk and is a small globular protein which exists as a mixture of monomers and dimers at neutral pH.¹⁵³ The secondary structure of β -LG is well known and consists of a β -sheet barrel, with a small α -helix structure outside (shown in Figure 4.4). Three separate binding sites have been determined, the internal cavity of the β -sheet barrel, the outer surface of the barrel between that and the α -helix structure and finally one on the outer surface between two amino acids Trp and Arg.¹⁵⁴ The interaction between β -LG and various polyphenols has been extensively studied.^{155–157} Kanakis *et al.* used FTIR, CD and fluorescence spectroscopy to show changes occurred in the protein structure due to interactions with epicatechin (EC), epigallocatechin gallate (EGCG) and other derivatives. Increases in β -sheet and no change in α -helix conformations were observed suggesting the polyphenols stabilised the secondary structure, while EGCG was shown to interact more with the protein when compared with EC.¹⁵⁵

4.4 Evidence of interactions between polyphenols and mucins

There is already significant evidence for the interactions of catechins with mucins.^{136,158} Originally, research was conducted using pig intestinal mucins (MUC2 and MUC5AC). The addition of EGCG to solutions of MUC5AC resulted in an increase in the viscosity of solution and small angle neutron scattering (SANS) showed larger order structures were also formed. This indicated that EGCG acts a cross-linker. When EC was used instead, very little change was observed in the viscosity and formation of larger order structures. This suggests that it is the gallate ring that is key in the interaction between polyphenols and mucins.¹⁵⁸

Salivary mucins and their interactions with green tea polyphenols were investigated by Davies *et al.*¹³⁶ Both MUC5B and MUC7 (a non gel-forming mucin) were investigated with EC and EGCG. Similar results compared to the intestinal mucins were observed. For example a large increase in viscosity was measured for EGCG and MUC5B but very little change was observed for EC and MUC5B. Rate-zonal centrifugation was used to determine structural changes of MUC5B and MUC7 when EC and EGCG were added, by essentially separating samples into many fractions based on the size of structures. For example large cross-linked units would be observed at a large fraction number compared with smaller molecules observed at very low fraction numbers. Mixtures of EGCG with MUC5B or MUC7 were both observed at larger fraction numbers compared with mixtures of EC and mucins where they were only

observed at small fraction numbers. This again reinforces the idea that the gallate ring is essential in the interaction between mucins and tea polyphenols.¹³⁶

The interactions between gastrointestinal mucins and black tea extracts were also investigated. Increases in the viscosity and large structure formation were observed for the black tea extract. This extract contained both theaflavin (TF) and its gallated derivatives, although this could not be conclusively confirmed, it was thought likely that it was the gallated derivatives that interacted with the mucin.¹⁵⁸

Recent work carried out by Elena Owen has shown evidence for interactions between four different theaflavins and salivary mucins, MUC5B and MUC7.¹⁵⁹ Rate-zonal centrifugation was again used to show structural changes in MUC5B and MUC7 when gallated theaflavins are present. The interaction between the mucins and theaflavin digallate (THDG) showed large cross-linked units forming. Very little change was seen when only theaflavin was present.¹⁵⁹

FTIR spectroscopy was also used in this case to investigate the effect MUC5B had on the spectrum of TF and THDG. A small shift to lower wavenumber was observed for the peak at *ca.* 1250 cm⁻¹ for THDG after it had been incubated with the mucins. This peak has been assigned to the galloyl ring -OH bend/CH bend. No significant shift was observed for the theaflavin spectrum post-incubation. This further supported the idea of the gallate groups being critical to the interaction.¹⁵⁹

There is also specific evidence suggesting that the main interaction between mucins and polyphenols occurs at the N-terminal domain. This has been observed using TEM, where

the N-termini doubled in size with the addition of EGCG.¹⁶⁰ Rate-zonal centrifugation experiments looking at the oligosaccharide rich regions of MUC5B showed no change when in the presence of EGCG.¹³⁶ It is therefore likely that the theaflavins are also interacting with the N-terminal domain as well.

N-terminal domains of MUC5B (NT5B) were purified. The effect of these on the IR spectra of TF and THDG was investigated. Once again a small but significant peak shift at *ca.* 1250 cm^{-1} was observed for the THDG after incubation with NT5B. No shift was observed for TF. These IR spectra again demonstrated that the theaflavin derivatives interact significantly and induce changes in the mucin organisation. This is also consistent with that observed with the green tea catechin derivatives and underlines the conclusions that a gallate ring is essential in the interaction with mucins.¹⁵⁹

Figure 4.5 shows a representation of the proposed gelation mechanism between gastrointestinal mucins and EGCG. The mucins initially exist in solution with only minor interactions (a and c). When EGCG is added it binds to the C-termini, internal cys domains and the N-termini which results in the N-termini doubling in size (b). The EGCG molecules then act as a 'glue' causing the N-termini to aggregate together and the C-termini and cys domains to cross-link. Ultimately this results in a viscoelastic gel (b and d).¹⁵⁸

4.4.1 Aims

From previous research it is known that theaflavins and salivary mucins interact significantly, however the exact effect polyphenols have on the protein structure is still unknown. It has

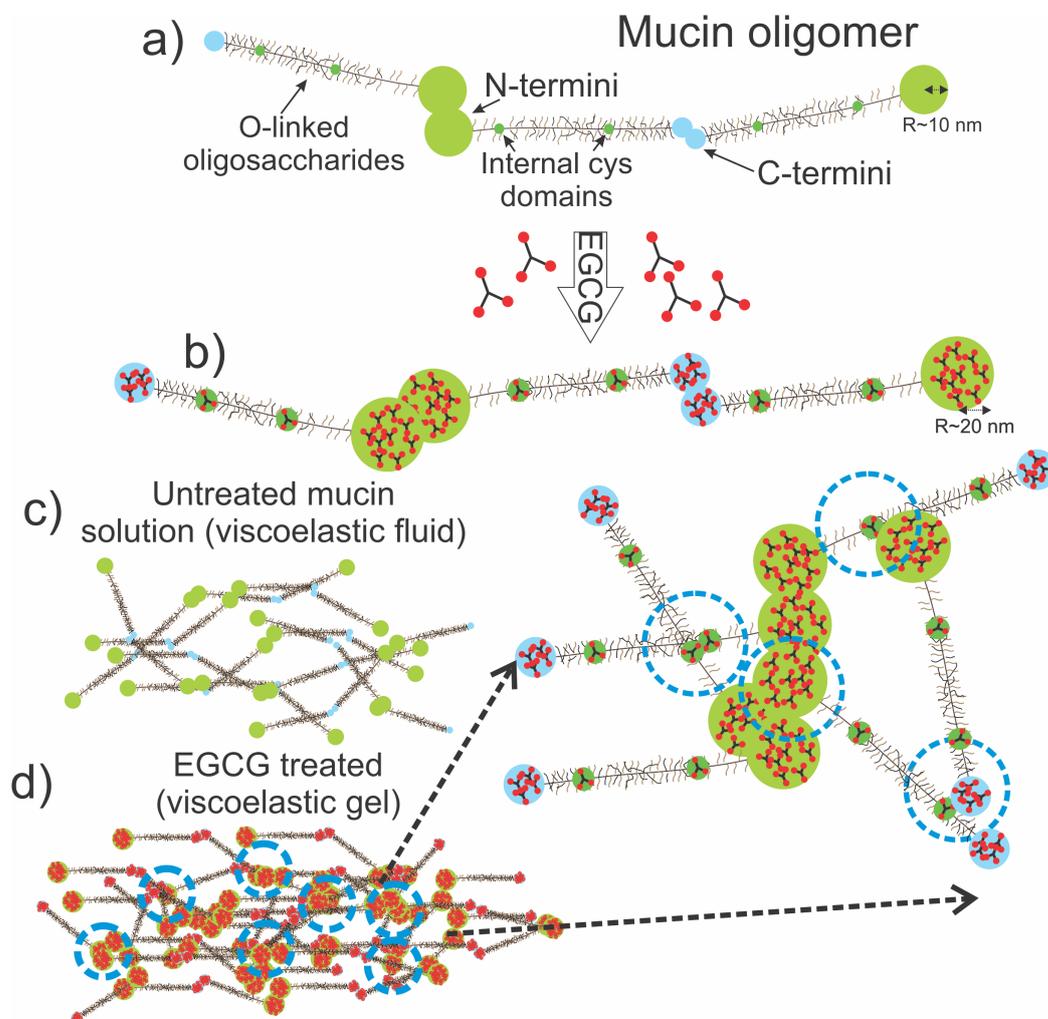


Fig. 4.5 A schematic representation of how mucins gel in the presence of EGCG. a) Shows the basic structure of mucin with the green and blue circles representing N- and C-termini respectively, internal cys domains are represented by smaller green circles and EGCG molecules shown by three small red circles. b) Represents EGCG binding and swelling the N-termini in particular as well as the C-termini and internal cys domains. c) Shows the mucins in solution. d) The gelled mucins with EGCG, the blue circles highlight the aggregation and cross linking between mucin chains. Reproduced from [158].

been shown that the glycosylated regions of MUC5B have limited interactions with these molecules and it is the C- and N-terminal and cysteine rich domains which do. The aim of this work was to determine what effect theaflavins had on the structure of the N-terminal domain protein, NT5B, using 2D-IR spectroscopy. The next section presents FTIR and 2D-IR spectra of two types of protein, with and without a ligand. The first protein is β -LG in the presence of the green tea extract EGCG. This system has been extensively studied previously using various methods including FTIR, CD and fluorescence spectroscopy, however, this investigation presents the first results obtained using 2D-IR spectroscopy.¹⁵⁵ This was used as a test system from which another, more unknown protein could be investigated. The second protein investigated was NT5B, the interactions between NT5B and THDG were specifically looked at here. Unfortunately, due to the limited supply of NT5B, it was only possible to investigate one theaflavin. Earlier work by Elena Owen indicated that THDG interacted the most, therefore this was selected as the theaflavin used.

4.5 Results

FTIR and 2D-IR spectra are displayed below for β -LG and NT5B. Spectra were collected for each protein with and without a polyphenol ligand. The amide I band was used to specifically investigate the changes in the secondary structure of the proteins as a result of interaction with the polyphenols. For the 2D-IR spectra, diagonal slices of the spectrum for the fundamental transition were extracted along with pump spectral slices to calculate the anharmonicity.

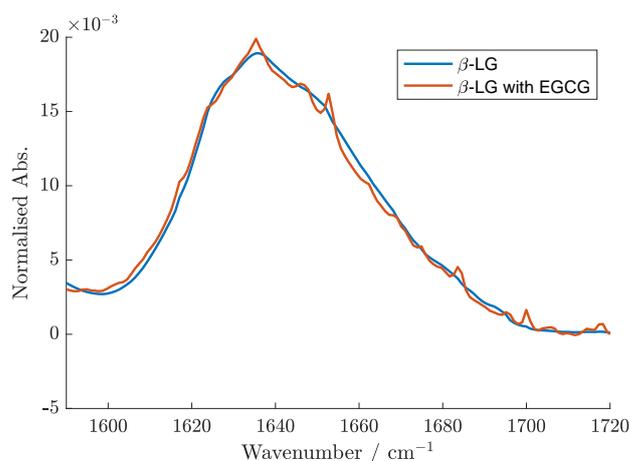


Fig. 4.6 FTIR spectra of the amide I band of β -LG (blue) and β -LG with the presence of EGCG (orange). No substantial shift in the amide I band is observed, indicating no significant change in the secondary structure.

4.5.1 β -LG

FTIR spectroscopy

The FTIR spectrum of β -LG with and without EGCG present is shown in Figure 4.6. Interestingly, there appears to be very little change in the amide I band. This indicates only small differences in secondary structure occur due to the interaction with EGCG. This is surprising considering FTIR spectra in [155] showed large changes in the amide I band. Peak fitting of the amide I band was performed to calculate the composition of secondary structure. The resulting peak fits are shown in Figure 4.7 and summarised in Table 4.1. The peak fitting suggested adding EGCG resulted in an increase of β -sheet content, while the α -helix content decreased.

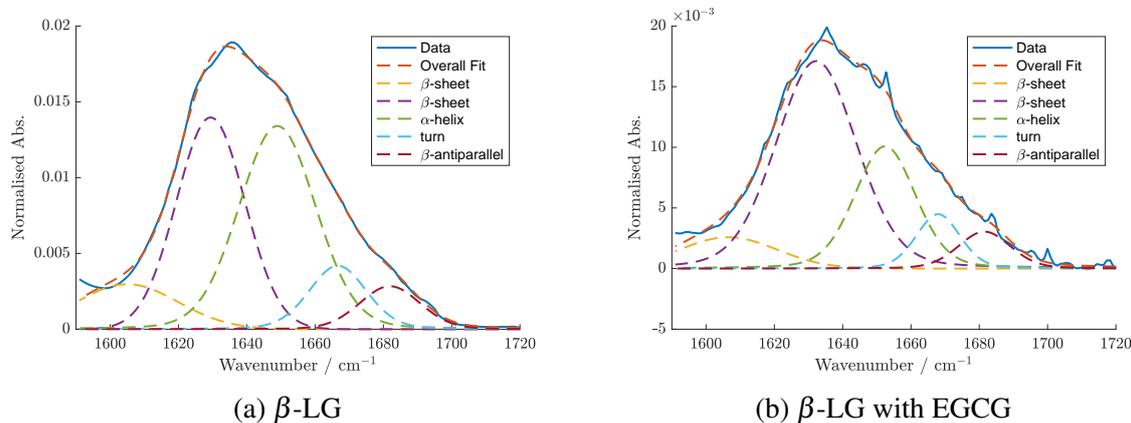


Fig. 4.7 Peak fits of FTIR spectra of β -LG and β -LG with EGCG present. The fits of each pseudo-voigt are represented by the dotted lines and assigned to a particular secondary structure type.

Table 4.1 Peak fitting results from amide I band of β -LG with and without EGCG with percentage values for the estimated secondary structure composition. These were calculated based on the area of each fitted peak.

	β -sheet	α -helix	turn	β -antiparallel
β -LG	46	39	10	6
β -LG with EGCG	63	24	7	6

2D-IR spectroscopy

2D-IR spectra of the amide I band of β -LG are shown in Figure 4.8. Two different spectra are displayed, one with EGCG (4.8b) and one without (4.8a), both at a waiting time of 0 ps. The spectra display clear peaks for different types of secondary structure when compared to the broad single peak of the amide I band in the FTIR spectrum. Peaks are clearly observed for β^\perp , α -helix and β^\parallel vibrational modes for pure β -LG, this is consistent with the known secondary structure of β -LG which mainly consists of β -sheet with small α -helix segments.¹⁶¹ The fourth peak is assigned to β -sheet as well, labelled β^{ISC} (inter-strand coupling). This is the same vibrational mode as β^\perp , except more coupling occurs which is due to closer strands in the β -sheet, as well as a slightly more parallel angle of the amide modes, resulting in it being redshifted.

The only difference observed in the 2D-IR spectra of β -LG when in the presence of EGCG compared to the absence of EGCG, is a slight change in the intensity of the various peaks relative each other. In order to investigate these changes in more detail, diagonal slices of the 2D-IR spectrum along the negative fundamental transition peaks were extracted and are shown in Figure 4.9. This is similar to comparing FTIR spectra, however, 2D-IR spectra resolve the different secondary structure better, thanks to narrower linewidths and amide I coupling. The region assigned to β^{ISC} gains significant intensity (*ca.* 1605 cm^{-1}). There is also a slight redshift observed in the β^\perp peak (*ca.* 1625 cm^{-1}). The α -helix mode (*ca.* 1640 cm^{-1}) appears to decrease in intensity as well as the β^\parallel peak (*ca.* 1665 cm^{-1}).

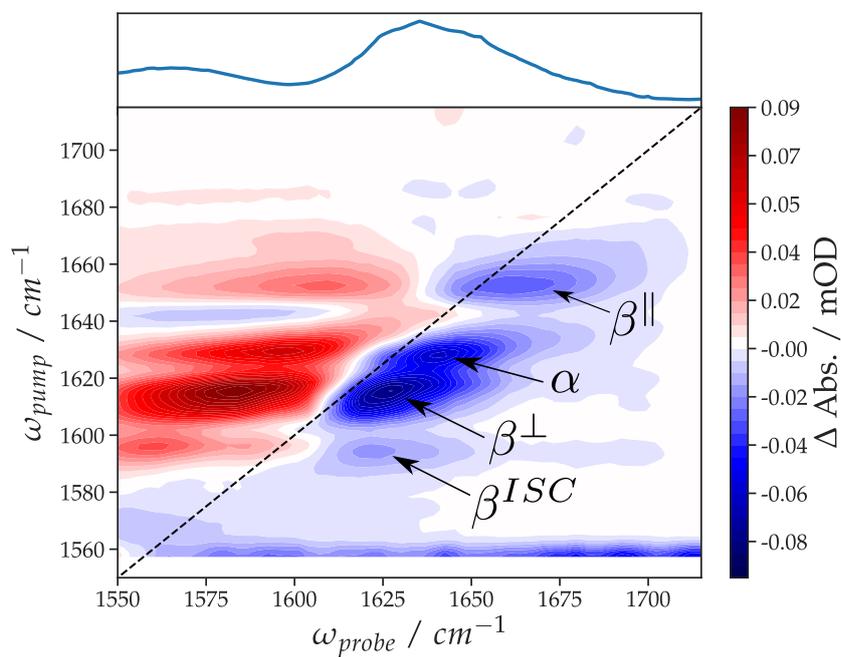
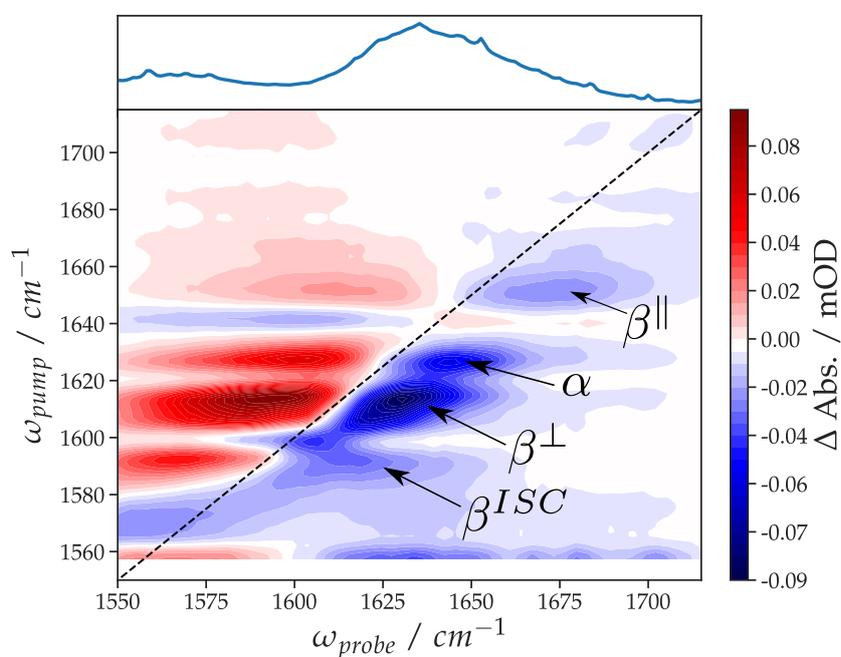
(a) β -LG(b) β -LG with EGCG

Fig. 4.8 2D-IR spectra of the amide I band of β -LG with (b) and without (a) EGCG present. The negative, blue, peaks represent the $\nu_{0 \rightarrow 1}$ transitions and the positive, red, peaks represent the $\nu_{1 \rightarrow 2}$ transitions. The different secondary structure amide I peaks are highlighted with labels.

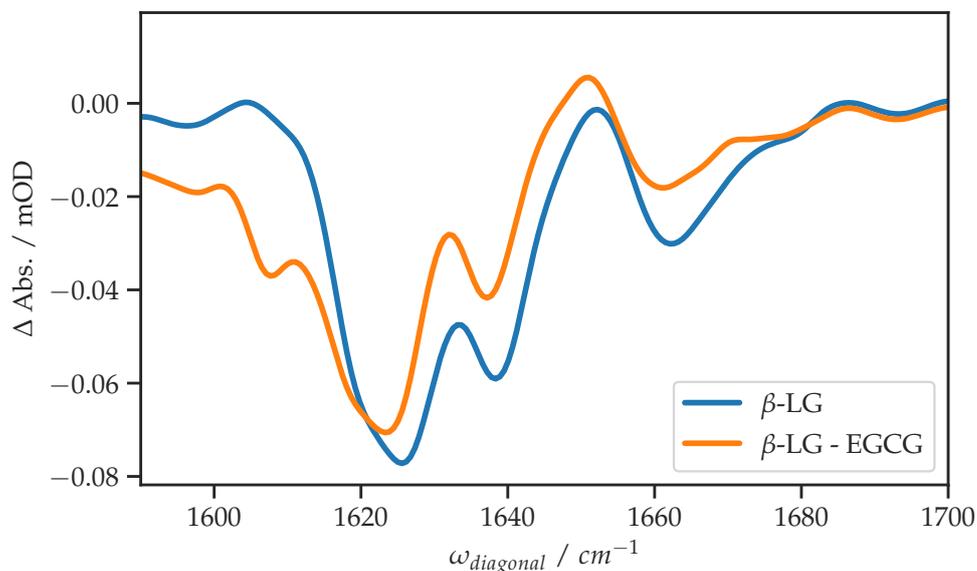


Fig. 4.9 Extracted diagonal spectra from the 2D-IR spectra of β -LG with and without EGCG. A small change at *ca.* 1605 cm^{-1} is observed between the two spectra.

Collectively these differences suggest changes in the secondary structure of β -LG with EGCG present.

Pump slices along each peak were extracted in order to calculate the anharmonicity of each vibrational mode of the amide I band, in both the absence and presence of EGCG. This was done by fitting Gaussian peaks to each spectral slice. In some cases, spectral slices required more than two peaks to provide a sufficient fit, due to cross peak intensity present in the slices. In particular, the β -LG slice of the β^{ISC} is clearly formed of 4 peaks as a result of coupling to vibrational modes with significantly larger intensity. These spectral slices are displayed in Figure 4.10 and the calculated anharmonic values are summarised in Table 4.2. The values do appear to be significantly larger than expected but most values are consistent with each other.¹⁶²

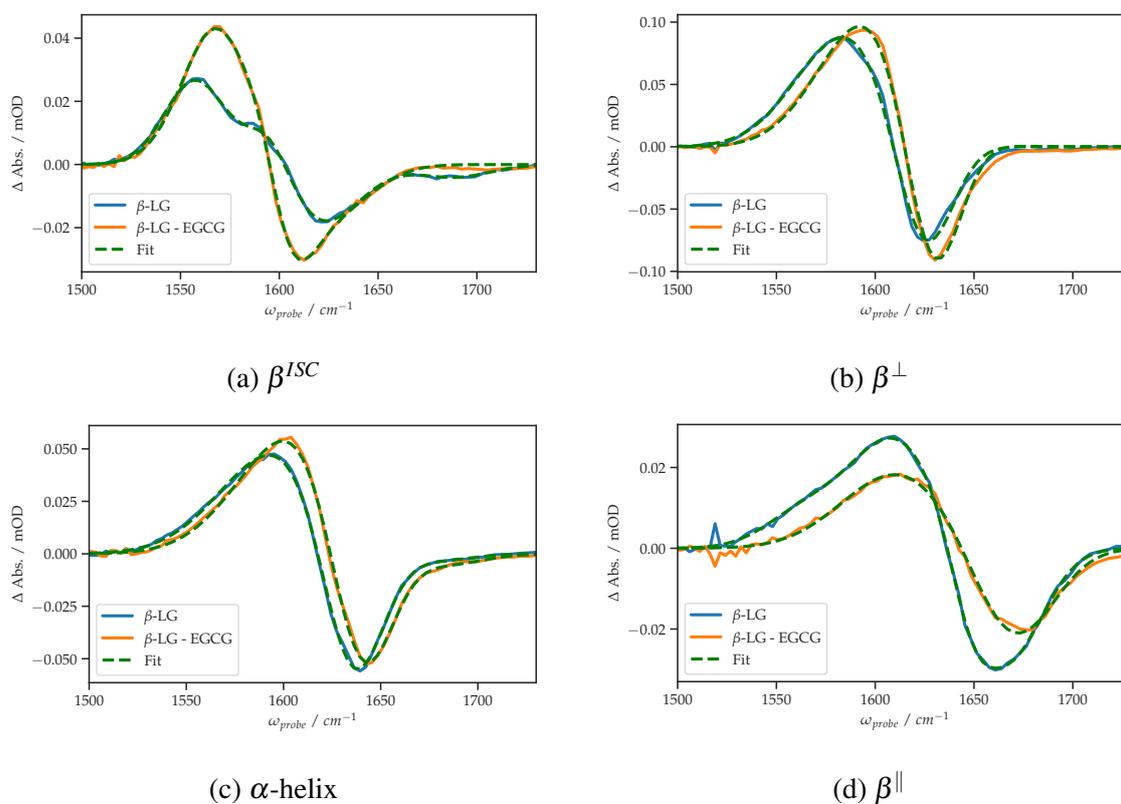


Fig. 4.10 2D-IR spectral pump slices of β -LG with and without EGCG at each of the main vibrational modes of the amide I band. Each spectral slice has been fitted with multiple Gaussian peaks (green dashed line) so that the anharmonicity values could be extracted.

For the α -helix mode, the anharmonicity decreases. Which suggests an increase in α -helix content since the two are inversely proportional. The β^\perp mode decreases in anharmonicity too which suggests an increase in β -sheet thanks to an increase in the de-localisation of the vibrational modes. Conversely, the β^\parallel anharmonicity increases slightly, suggesting the modes along the β -strands are becoming more localised. However, as the change is only 3 cm^{-1} this could instead be due to broadening of the $\nu_{1\rightarrow 2}$ peak.

The β^{ISC} anharmonicity also decreases significantly, however, the spectral slices shown in Figure 4.10 are more complex. For β -LG, there are clearly two modes coupled (β^\perp at 1625 cm^{-1} and β^\parallel at 1670 cm^{-1}) and because these modes are significantly more intense than the β^{ISC} mode, they end up dominating the spectrum. Therefore the calculated shift was judged to be an artefact. For β -LG with EGCG, there is no evidence of cross peaks between the β^{ISC} and β^\perp modes, however, there is a significant shoulder at *ca.* 1650 cm^{-1} indicating the β^{ISC} and α -helix modes are coupling (Figure 4.10a). Small cross peaks are observed for the β^\perp spectral slice (minor deviations of the fit at *ca.* 1650 cm^{-1}) suggesting very minor coupling of the β^\perp and α -helix modes (Figure 4.10b). No coupling was observed with the α -helix slice, the cross peaks expected here could be obscured by the two main peaks (Figure 4.10c). Finally, the β^\parallel slice does not show any cross peaks for the β -LG-EGCG sample. However, for the β -LG sample, there is one positive cross peak present at *ca.* 1565 cm^{-1} indicating a small amount of coupling between the β^\perp and β^\parallel modes (Figure 4.10d).

Transition dipole analysis was conducted in order to assess the structural changes further. This utilises the fact that the intensity in FTIR spectroscopy is $\propto |\mu|^2$, whereas in 2D-IR

Table 4.2 Anharmonicities calculated from fitting spectral slices for each vibrational mode. The β^{ISC} anharmonicity for β -LG appears to be very large, however this is an artefact due to other modes coupling. This is discussed further in the body of the text.

Vibrational mode	Δ / cm^{-1}	
	β -LG	β -LG with EGCG
β^{ISC}	63	21
β^\perp	39	32
α	37	32
β^\parallel	56	59

spectroscopy it is $\propto |\mu|^4$.²⁹ This process normally requires a calibration standard, however, in this case the two spectra are only compared relative to each other. Therefore the transition dipole spectrum calculated is relative as opposed to absolute.³⁰ This is done by dividing the normalised 2D-IR diagonal intensity by the normalised FTIR intensity (eq. 4.1). Figure 4.11 shows the calculated transition dipole spectrum. This confirms the β^{ISC} increase, while the minor β^\parallel change is also evident, as well as the slight redshift of the β^\perp mode. The α -helix mode does not appear to significantly change, confirming that the α -helix structure remains (random coil modes also appear in the same region as the α -helix modes but the constant, or near constant, transition dipole shows that it remains an α -helix structure).

$$d_{rel}(\omega) = \frac{I_{2D-IR \text{ Diag.}}(\omega)}{I_{FTIR}(\omega)} \quad (4.1)$$

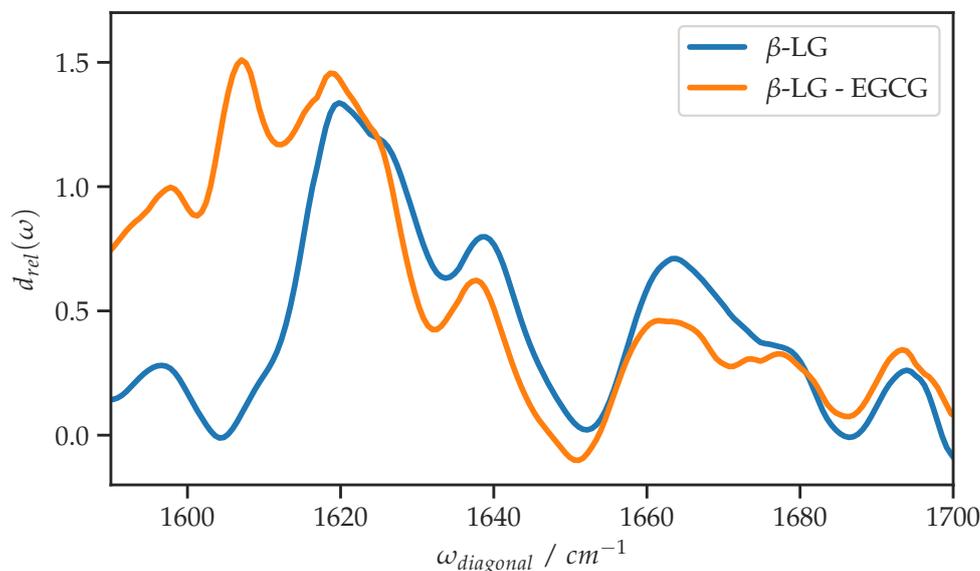


Fig. 4.11 Calculated relative transition dipole spectrum for β -LG with and without EGCG. A large increase in the transition dipole spectrum at *ca.* 1605 cm^{-1} is observed when EGCG was added. This indicates that the β -sheet structure is altered.

4.5.2 NT5B

FTIR spectroscopy

The FTIR spectra of NT5B with and without THDG present in the amide I region are shown in Figure 4.12.* Here a significant blueshift of the amide I band was observed. The overall profile of the band also broadened, suggesting a larger range of secondary structures. Peak fitting was performed on the amide I band showing an approximate composition of 60 % β -sheet to 40 % α -helix (Table 4.3). When THDG was added, the proportion of α -helix remained unchanged with only the distribution of β -sheet modes altered, suggesting a change in the β -sheet structure.

*NT5B spectrum was particularly noisy due to a malfunctioning purge system.

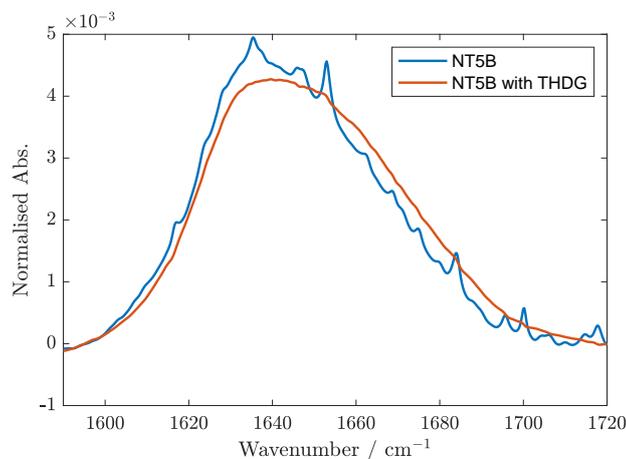


Fig. 4.12 FTIR spectra of the amide I band of NT5B (blue) and NT5B with the presence of THDG (orange). A slight blueshift is observed when THDG was added indicating a change in the secondary structure.

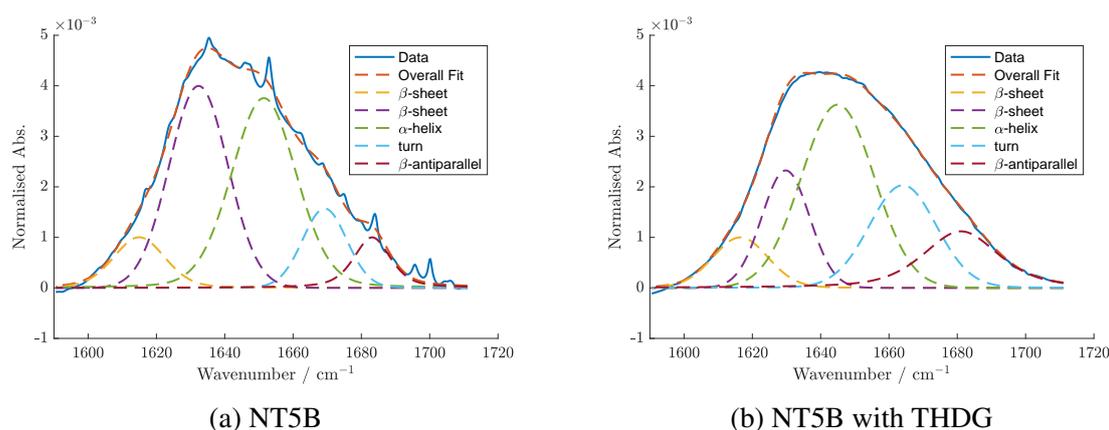


Fig. 4.13 Peak fits of FTIR spectra of NT5B with and without THDG present. The fits of each pseudovoigt are represented by the dotted lines and assigned to a particular secondary structure type. Differences in the fitted peaks can be observed between the two samples.

Table 4.3 Peak fitting results from amide I band of NT5B with and without THDG with percentage values for the estimated secondary structure composition. These were calculated based on the area of each fitted peak.

	β -sheet	α -helix	turn	β -antiparallel
NT5B	44	38	11	7
NT5B with THDG	25	40	21	14

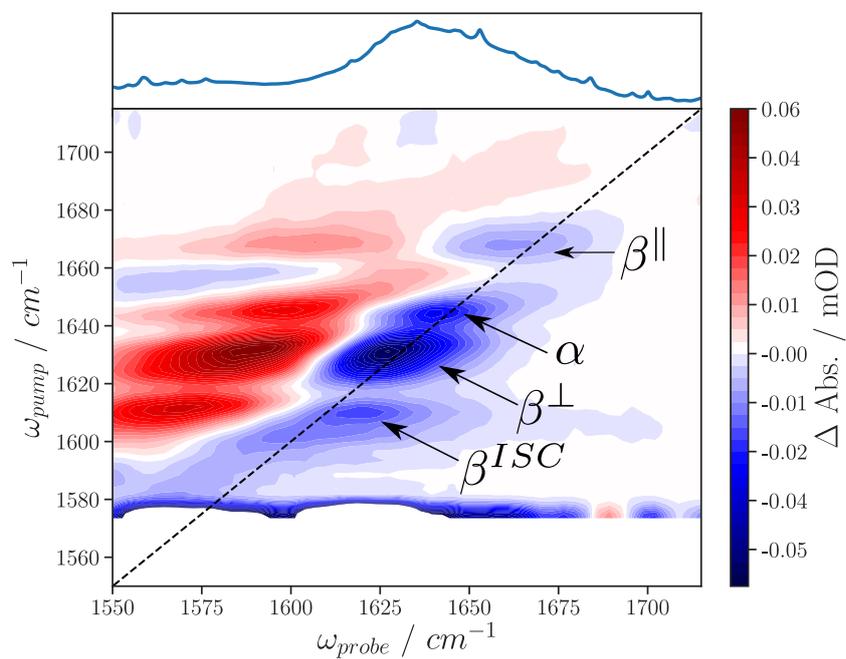
2D-IR spectroscopy

2D-IR spectra, in the amide I region, of NT5B with and without THDG are shown in Figure 4.14. Without THDG present, the amide I region comprises of clear peaks indicating a mixture of secondary structures. Peaks assigned to β^{ISC} , β^\perp , α -helix and β^\parallel vibrational modes are visible. In the presence of THDG, the 2D-IR spectrum changes significantly, suggesting the secondary structure also changes. 2D-IR spectroscopy makes this much clearer than FTIR spectroscopy. There is no longer a peak assigned to the β^{ISC} vibrational mode, while the β^\perp and α -helix modes appear to have merged and the β^\parallel peak has a stronger intensity relative to the other peaks in the spectrum.

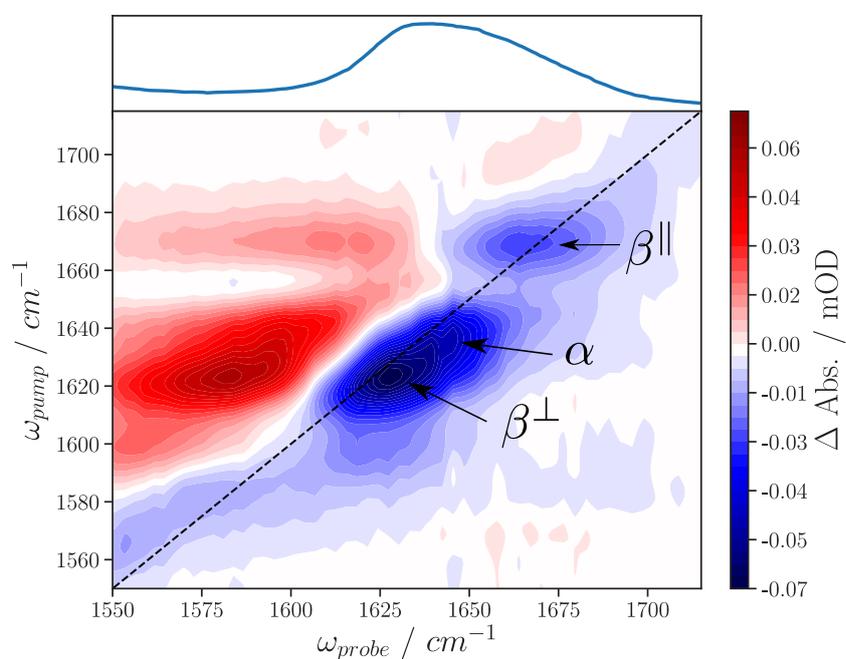
Diagonal slices of each 2D-IR spectrum were extracted in order to make the changes in diagonal peaks clearer (Figure 4.15). With respect to adding THDG, the β^{ISC} and β^\perp peaks (1610 and 1630 cm^{-1} , respectively) appear to merge into one larger, broader peak (1625 cm^{-1}) assigned to β^\perp , with a shoulder visible at higher wavenumber. The α -helix mode appears at the same frequency and the β^\parallel mode is slightly more intense and broader.

Spectral slices along the probe axis were extracted for each vibrational mode identified to analyse the spectral changes further. Gaussian peaks were again fitted to the $\nu_{0\rightarrow 1}$ and $\nu_{1\rightarrow 2}$ peaks and the anharmonicity calculated. The spectral slices are shown in Figure 4.16 and the anharmonicities summarised in Table 4.4.

The β^{ISC} slice shows that there is still a small peak present in the NT5B with THDG spectrum; the intensity is considerably smaller and it is evident from the 2D-IR spectrum that it is a shoulder of the main β^\perp peak. The anharmonicity considerably increases with the addition



(a) NT5B



(b) NT5B with THDG.

Fig. 4.14 2D-IR spectra of the amide I band of NT5B with (b) and without (a) THDG present. The negative, blue, peaks represent the $\nu_{0 \rightarrow 1}$ transitions and the positive, red, peaks represent the $\nu_{1 \rightarrow 2}$ transitions. The different secondary structure amide I peaks are highlighted with labels.

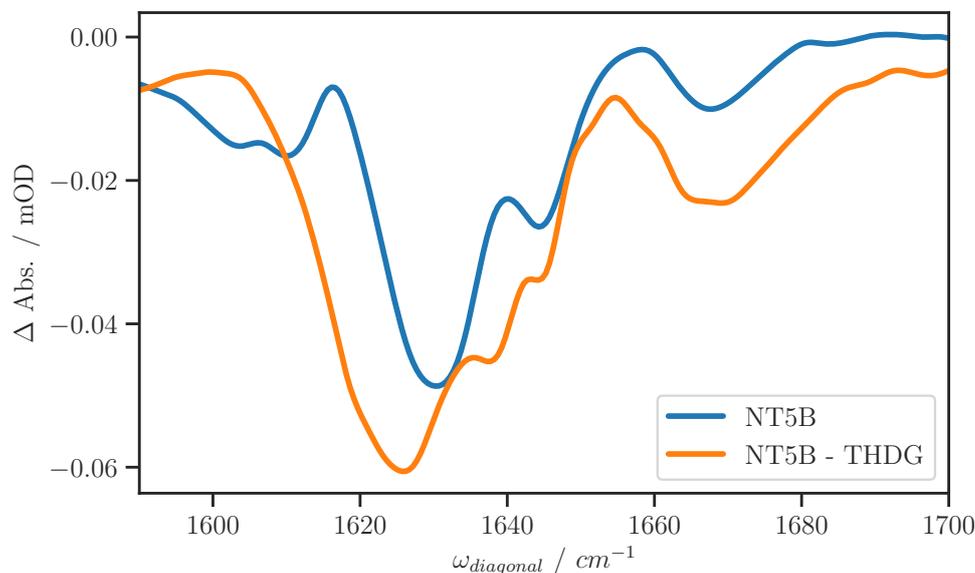


Fig. 4.15 Extracted diagonal spectra from the 2D-IR spectra of NT5B with and without THDG. A change between *ca.* 1605 and 1630 cm^{-1} is observed between the two spectra, indicating a change in the β -sheet structure of NT5B.

of THDG, indicating a large reduction in the de-localised modes and a loss of secondary structure. The anharmonicity also increases for the β^\perp and β^\parallel modes, albeit by a much smaller degree (6 and 17 cm^{-1} vs 35 cm^{-1}), which suggests a reduction in the intra-strand coupling. There is also an overall peak broadening effect, indicating a loss of well defined secondary structure. The α -helix mode also increases in anharmonicity, although this could be due to overlapping with the large β^\perp peak.

Cross peaks are evident in the spectral slices shown in Figure 4.16. There is one observed in the β^{ISC} slice for both samples coupling to the peak *ca.* 1670 cm^{-1} . Cross peaks are also observed with the NT5B - THDG sample with the β^\perp (small) and α -helix spectral slices indicating a coupling to the β^\parallel mode.

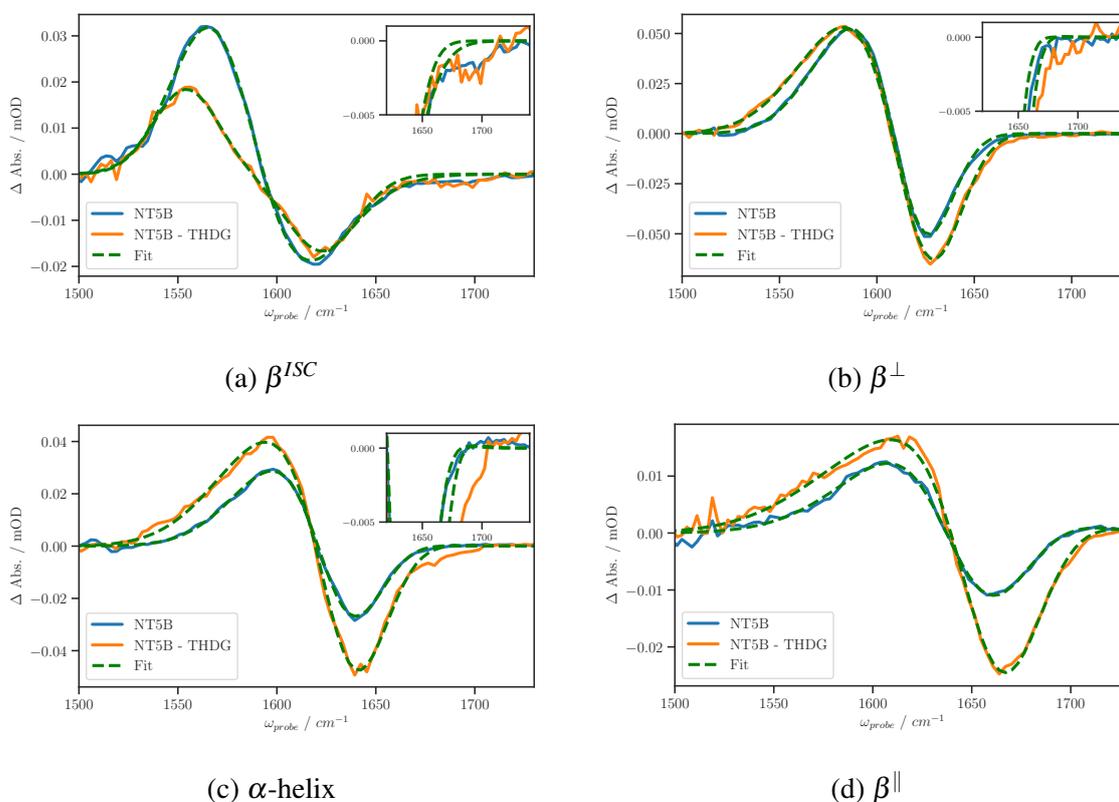


Fig. 4.16 2D-IR spectral pump slices of NT5B with and without THDG at each of the main vibrational modes of the amide I band. Green dashed lines show the Gaussian fits and the insets show cross peaks in a zoomed in region of the spectrum. The peak positions of the fitted Gaussian peaks were used to calculate the anharmonicities.

Table 4.4 Anharmonicities calculated from fitting spectral slices for each vibrational mode. The large β^{ISC} anharmonicity value is due to other larger modes coupling and is also discussed further in the body of the text.

Vibrational mode	Δ / cm^{-1}	
	NT5B without THDG	NT5B with THDG
β^{ISC}	30	65
β^{\perp}	31	37
α	29	37
β^{\parallel}	27	44

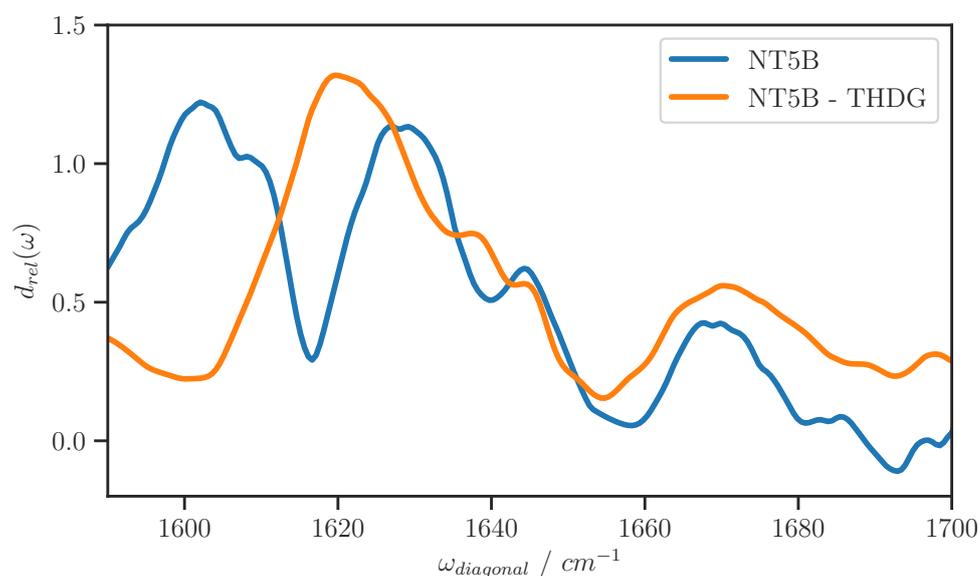


Fig. 4.17 Relative transition dipole strength spectra calculated for NT5B with and without THDG. A significant change in the transition dipole spectrum between *ca.* 1600 and 1630 cm^{-1} is observed when THDG was added. This indicates that the β -sheet structure is altered.

As with β -LG, the relative transition dipole was calculated for both NT5B samples. This provides insights into the secondary structure assignments and is shown in Figure 4.17. As with the diagonal spectrum, both β^{ISC} and β^{\perp} appear to merge into one broader mode centred at 1620 cm^{-1} . The shoulder of this peak at 1638 cm^{-1} is also assigned to β^{\perp} . The transition dipole strength of the α -helix mode appears to remain constant, indicating little change in the α -helix content as a result of adding THDG. There is a slight increase in the intensity of the β^{\parallel} mode, however, this does not appear to be significant.

4.6 Discussion

4.6.1 β -LG with EGCG

FTIR spectroscopy showed very little change in the amide I band of β -LG when EGCG was added. These results were unexpected as previous work had shown EGCG interaction with β -LG resulted in large amide I changes. Specifically, work by Kanakis *et al.* showed large changes in β -sheet vibrational modes which were not reflected in the above FTIR spectra despite the concentrations being similar.¹⁵⁵ The discrepancy could be due to the purity and variant of protein used. The β -LG A variant was used by Kanakis *et al.*, however, a mixture of β -LG A and B was used here. There is a possibility that the B variant signal offsets the change from the A variant, although this is unlikely given A and B only differ by a few specific amino acids which should not impact significantly on the secondary structure. Another possibility is that A and B variants are forming oligomers, changing the tertiary and quaternary structures of the protein solution. Having said this, the peak fitting conducted on the FTIR amide I band did show an increase in β -sheet content which is consistent with [155]. However, a decrease in α -helix content was also observed, which is inconsistent with [155], where it was observed to remain constant. It is well established that peak fitting the amide I band can obtain varying results depending on what peak fitting parameters are used.²⁶ 2D-IR spectroscopy is a more powerful technique in investigating the secondary structure of proteins, because it is sensitive to the coupling between groups of amide I modes. This can provide more detailed information about the structure of the protein.

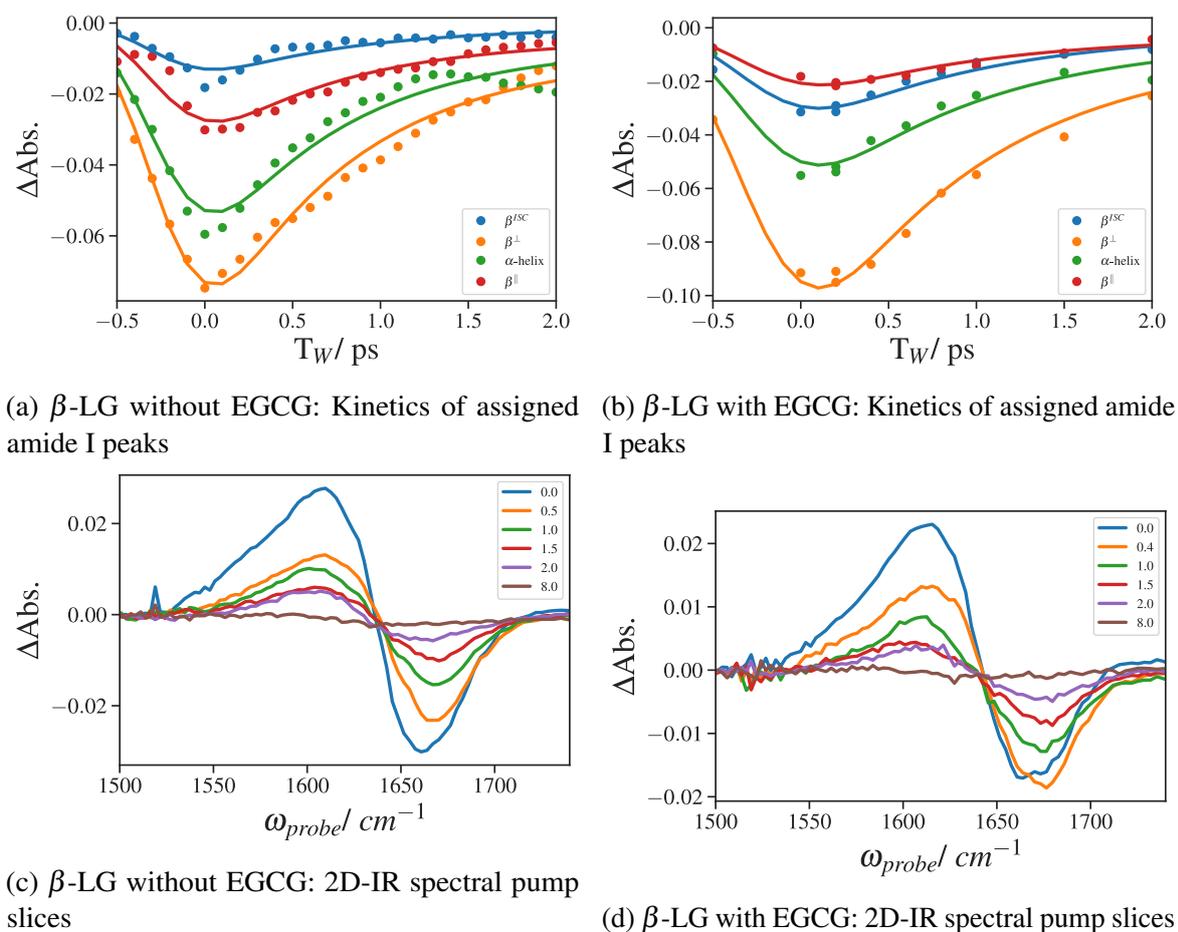


Fig. 4.18 (a) and (b): kinetic traces of the assigned amide I $\nu_{0 \rightarrow 1}$ peaks in 2D-IR spectrum for both β -LG samples. Each set of kinetic traces have been globally fitted with one exponential decay resulting in values of 0.91 ± 0.14 ps and 1.00 ± 0.09 ps, respectively. (c) and (d) 2D-IR spectral pump slices at 1652 cm^{-1} for a range of waiting times for both β -LG samples. These show ‘standard’ decays of excited peaks without any extra peaks forming at later time delays, which indicates an absence of H_2O in the sample.

The 2D-IR spectra, in comparison with FTIR spectra, display clearer signatures of the different secondary structures present in β -LG. Spectral features assigned to β^\perp , β^\parallel and α -helix were all observed, as well as a vibrational mode assigned as β^{ISC} . The vibrational mode assigned to β^{ISC} was observed at an unusually low frequency for an amide I vibration. The residue arginine does have a vibrational mode that is active in this region, however, this was discounted as a possibility given its low abundance in β -LG (1.7 %).⁷ Solvent exposure could also redshift the β^\perp mode, this would have to be significant though.

Another possibility is that it results from residual H₂O in the sample, this is also unlikely though for the following reasons: Firstly, kinetics of the peak assigned to β^{ISC} would have to be significantly different, which is not the case since all peaks decay with a *ca.* 1 ps lifetime (shown in Figures 4.18a and 4.18b). Secondly, there would also have to be evidence of a water peak arising at longer time delays (*ca.* 5 ps) and ‘growing’ in as the amide I signal decays due to solvent heating.¹⁶³ Figures 4.18d and 4.18c show 2D-IR pump spectral slices of β -LG with and without EGCG, which both show the $\nu_{0\rightarrow 1}$ and $\nu_{1\rightarrow 2}$ peaks decaying as the T_W is increased without any additional peaks appearing. Finally, for experimental reasons it also unlikely to be due to H₂O. This is because the β^{ISC} peak was observed for NT5B but disappeared when THDG was added to the exact same sample therefore it would be highly unusual for H₂O to escape the sample in the short time THDG was added. For these three reasons it was thought that water is unlikely to be the cause of the peak at *ca.* 1605 cm⁻¹, however, this cannot be ruled out entirely.

The β^{ISC} vibrational mode was observed in β -amyloid fibrils by Lomont *et al.* and is a variant of the β^\perp mode.¹⁶² It was assigned to a smaller inter-strand distance with a slightly more parallel angle of the amide modes within a β -sheet structure. The anharmonic coupling was also observed to be slightly larger (*ca.* 5 cm^{-1}) than the typical β^\perp mode. Suggesting, the vibrational modes are more de-localised across the structure. Cross peaks from β^\perp and β^\parallel modes dominate the spectral slice for β -LG. This indicated that the structure from the β^{ISC} mode forms a small part of β -LG and is attached to the main β -sheet structure.

When spectra are compared between β -LG with and without EGCG present, initially there does not appear to be significant differences. However, changes were observed when diagonal slices were extracted, for example a large increase in β^{ISC} peak intensity was observed. The intensity ratio between the β^\perp and β^\parallel modes is an indicator for the size of β -sheet structures.^{164,165} This ratio changed from *ca.* 3 to 5, indicating larger β -sheets were formed. The frequency difference between these two modes is also an indicator of β -sheet size. A slight redshift was observed for the β^\perp peak while the β^\parallel peak remained constant. This again suggests slightly larger β -sheet structures are formed by adding EGCG to β -LG. The anharmonicity provides a measure of the de-localisation of a vibrational mode. In this particular case, the anharmonicity decreased by 7 cm^{-1} for the β^\perp mode, again indicating larger β -sheet conformations.

The α -helix anharmonicity measurement indicated a small decrease which suggests a larger α -helix was formed. However, this was contradicted by a small decrease in the relative transition dipole strength (by adding EGCG) which indicated a slightly less ordered or

smaller helix. In FTIR spectroscopy it is difficult to distinguish between α -helix and random coil structures given the close frequency ranges and large peak widths. The transition dipole analysis has ruled out any large change from an ordered helix to a random coil configuration because there is no large change in dipole strength.³⁰ There was no shift observed in the frequency of the α -helix mode which suggests there has been little-to-no change in the helical structure. Given the conflicting information provided, it was assumed the α -helix secondary structure remained unchanged. This is supported by the CD experiments performed by Kanakis *et al.* which indicated very little change for α -helix when β -LG was interacting with EGCG.¹⁵⁵

Putting the 2D-IR observations together, no significant change in the α -helix conformation was observed as a result of EGCG. For β -sheet, the anharmonicity, redshift and dipole strength all indicate an increased de-localisation of the β -sheet vibrational modes. This could be as a result of larger β -sheet structures forming or similarly sized but slightly different ‘tighter’ β -sheets which have a large degree of coupling. Since β -LG is unlikely to form larger β -sheet structures, this is thought to be due to increased coupling between strands. These tighter interactions could occur inside the β -barrel which is a known docking site. The EGCG molecule could be stabilising the various side chains within the β -sheet barrel allowing for tighter binding between the β -strands and greater coupling of the local modes. Cross peaks present in spectral slices indicated significant interactions between β^{ISC} and α -helix modes with EGCG present. This could be caused by two factors: an EGCG molecule binding at the known site between the α -helix structure and β -barrel,¹⁵⁴ or, dimerisation of β -LG, resulting in α -helix and β -sheet structures directly interacting with each other.¹⁵⁰

Docking studies have indicated the β -barrel entrance is the most favourable site.^{154,155} It is therefore possible that dimerisation of β -LG in the presence of EGCG could explain this coupling.

4.6.2 NT5B

Contrary to the FTIR spectra for β -LG, the NT5B spectra showed changes in the amide I band when THDG was added. The peak fitting analysis performed indicated NT5B was mainly composed of β -sheet and α -helix. With THDG present, the α -helix content remained unchanged while the β -sheet distribution changed.

The 2D-IR spectra also indicated NT5B was formed of β -sheet and α -helix conformations. Davies *et al.* concluded that the N-terminal domain was mainly composed of β -sheet using TERS-Raman.¹⁴¹ However, the spectra here indicated the presence of α -helix too. The mixture of secondary structures present is also supported by the von Willebrand factor (similar to NT5B) which contains regions of both α -helix and β -sheet.¹⁶⁶

In the case of β -sheet modes, two distinct modes were observed (β^{ISC} and β^\perp) which suggest two different β -sheet conformations are present. It is possible these are two different conformations on one protein molecule, or alternatively, the result of the protein in two different conformations in solution. There is only one small cross peak observed in the spectral slices, this is between the β^{ISC} and β^\parallel modes. No cross peaks are observed between β^\perp and β^\parallel modes, which indicates two different protein conformations. It is known that NT5B exists as a dimer in solution, therefore it is possible that each monomer forming the

dimer has a different β -sheet conformation.¹⁴² However, further work would be required to determine this.

NT5B with THDG

The 2D-IR spectrum for NT5B with THDG showed large changes when compared with the NT5B spectrum. The main change was the absence of the β^{ISC} peak in the spectrum which appears to be a minor shoulder of the larger β^\perp peak. The extracted diagonal spectrum shows the ‘merging’ of the two modes into a broader peak. This is also reflected in the relative transition dipole spectrum. It suggests the THDG molecules disrupt the β -sheet structure, resulting in a reduction in the de-localisation (indicated by the anharmonicity increase). A reduction in the size of the β -sheet was also indicated by the reduced ratio between the β^\perp and β^\parallel peaks, from *ca.* 5 to 3. There is also a shoulder that appears at 1638 cm^{-1} in the $d(\omega)$ spectrum. Even though this is close to the characteristic frequency for α -helix conformations, it is thought to be due to a blue shifted β^\perp vibrational mode as a result of smaller β -sheet structures. This is because there is already a clear α -helix peak at 1642 cm^{-1} which remains unchanged in transition strength and frequency, suggesting the α -helix conformation remains unchanged as a result of THDG (the anharmonicity change is thought to be as a result of the overlapping broad β^\perp mode).

Cross peaks were observed in the spectral slices between the β^\parallel mode and β^\perp (small), α -helix (large) and β^{ISC} (medium) shoulder, indicating significant coupling between all the different amide I modes. This could potentially be a result of the THDG molecules mediating

the energy transfer between the different structural domains in the protein. Alternatively, this could be a result of the THDG molecules causing aggregates to form in solution with the different monomeric proteins assembling together (e.g. α -helix of monomer A bound to β -sheet structure of monomer B).⁷ There is previous evidence showing a tea polyphenol (EGCG) interacting with the N-terminal domain of mucins (MUC2 and MUC5A) results in the aggregation of the N-termini.¹⁵⁸ Rate zonal diffusion experiments also showed evidence aggregation of NT5B in the presence of THDG.¹⁵⁹ Further work would be required to resolve this.

4.7 Conclusion and future work

4.7.1 Conclusion

Using 2D-IR spectroscopy it has been shown EGCG causes changes in the secondary structure of β -LG. The FTIR spectra showed very little change between β -LG with and without EGCG. The 2D-IR spectra detected subtle changes in the β -sheet structure. The α -helix conformation did not appear to change significantly. This was consistent with previous literature, however, 2D-IR spectroscopy was able to detect two different β -sheet structures present when EGCG binds to β -LG which was not observed in previous studies.

2D-IR spectroscopy was also used to determine that the secondary structure of NT5B is composed of β -sheet and α -helix (which was in agreement with FTIR results). It is possible that the protein has two different β -sheet conformations in solution which are linked to the

dimer formation. Adding THDG changes the secondary structure of the protein. This results in reduced sizes of β -sheet, while α -helix conformations remain unchanged. The large β^{\perp} peak observed could indicate the protein aggregating and this being mediated by the THDG molecules, however, further work is required to confirm this. This shows that the β -sheet conformation in NT5B is potentially important in the cross-linking process of gel-forming mucins.

4.7.2 Future work

The β -LG spectra showed interesting features. Given that the structure of the protein is well known it would be useful to simulate the 2D-IR spectra of β -LG with and without EGCG bound.¹⁶⁷ This would allow the assignments and changes observed here to be confirmed. Since no significant changes were observed in the FTIR spectra, contrary to literature, it would also be helpful to perform this experiment again with the β -LG A variant specifically, in order to assess what differences the two variants have interacting with EGCG. Finally, the interactions between β -LG and theaflavins should also be investigated. This would provide further insight into the specific interactions and structural changes that occur in a cup of tea when milk is added. Ultimately, this would be working towards looking at the interactions between milk proteins, salivary mucins and tea polyphenols.

There is significant potential in investigating the effect other theaflavin molecules could have on the structure of NT5B. For example both theaflavin-monogallate molecules which also show interactions with NT5B, as well as theaflavin itself which does not show a significant

interactions with NT5B. It would be interesting to assess what effect these molecules had on the β -sheet conformation of the protein.

Unfortunately, due to equipment problems and time pressures, experiments with the C-terminal domain (CT5B) were unable to be carried out. However, it would be useful to assess if THDG has a similar effect on CT5B as these are also thought to be important in the cross-linking of the mucins. NMR could also be utilised to investigate the binding of THDG molecules to the protein. This could provide important information about how the structure of the theaflavin molecule changes upon binding and whether it exists as a bridge between protein monomers. NMR would also be useful for assessing the role disulphide linkages play between monomers.¹⁶⁸

Ultimately, a 3D atomistic structure of NT5B would be useful in this particularly case because simulated 2D-IR spectra can be calculated using this. This could provide confirmation with the interpretation of the spectra. This could be obtained either using either NMR or X-ray crystallography. An alternative method, however, is using the homology of the protein to find a known similar structure, and substitute in the primary structure of NT5B. This would allow one of the most 'active' parts of the MUC5B protein to be studied computationally, allowing many different interactions to be screened beforehand and in order to experimentally study those shown to be most significant.

Chapter 5

Experimental

5.1 2D-IR spectroscopy setups

Two ultrafast spectrometers were used to perform the experiments described in this thesis. The first one was the ULTRA spectrometer which is a powerful and flexible setup designed to perform spectroscopies such as TRIR, TA, 2D-IR and T-2D-IR. This spectrometer was used to perform the T-2D-IR experiments shown in Chapter 2 as well as the initial 2D-IR experiments of the organosulfate band for Chapter 3. ULTRA employs the double resonance technique in order to obtain 2D-IR and T-2D-IR spectra. The advantages and disadvantages of this technique were discussed previously in Chapter 1. The second spectrometer used to obtain 2D-IR spectra was the Nottingham 2D-IR spectrometer which is a PHASETECH 2D-IR spectrometer. This utilised the pulse-shaping technique, which was also discussed in Chapter 1. Unfortunately, laser tuning issues meant that this spectrometer could not be

used at *ca.* 1250 cm^{-1} . As a consequence, it could only be used to obtain spectra in the amide I region for Chapter 4 and spectra of the probe molecule ferrocyanide in Chapter 3. A description of each spectrometer is followed by the data processing and analysis methods and ends with details of the experimental sample preparation.

5.1.1 ULTRA spectrometer and laser system

The ULTRA spectrometer was used to collect 2D-IR spectra of carrageenan, using the double resonance technique, and T-2D-IR spectra of $\text{Fe}(\text{CO})_5$. The set-up has been described in detail elsewhere,⁶ but briefly: A 65 MHz Ti:Sapphire oscillator seeds two synchronised regenerative amplifier lasers at a repetition rate of 10 kHz. Two trains of pulses are produced which have a pulse durations (Δt) of *ca.* 1-2 ps and 50 fs and a wavelength of 800 nm. The double resonance technique utilises a tunable narrow-band pump pulse ($\Delta t = 1\text{-}2\text{ ps}$) and a time-delayed broadband probe pulse ($\Delta t = 50\text{ fs}$) which has been generated by pumping separate optical parametric amplifiers (TOPAS, Light Conversion) followed by difference frequency generators to produce two tunable mid-IR pulses. The pump and probe pulses are overlapped both spatially and temporally, the probe pulse is then directed in a spectrograph followed by two linear MCT detectors (resolution = $2\text{-}3\text{ cm}^{-1}$). 2D-IR spectra are generated by scanning the pump frequency and recording the dispersed pump-probe spectra for a given time delay. The pump pulses are amplitude modulated to obtain pump-on-pump-off difference spectra.⁶ For T-2D-IR spectroscopy, the same set-up as above was used except

a third UV/visible pulse was introduced to the system. A difference spectrum was then obtained by collecting spectra alternatively with and without the excitation pulse.

5.1.2 Nottingham 2D-IR spectrometer and laser system

The 2D-IR spectrometer at the University of Nottingham was used to collect spectra of the carrageenan - ferrocyanide systems and the two protein systems: β -LG with EGCG and the salivary mucin, NT5B, with THDG. The laser system used in the spectrometer setup is the Newport Spectra-Physics Solstice Ace. The regenerative amplifier is seeded by a Ti:Sapphire MaiTai laser and pumped by a Q-switched Ascend laser. The output is 6 mJ, *ca.* 100 fs pulses centred on 800 nm with a repetition rate of 1 kHz. The output is used to pump a TOPAS prime optical parametric amplifier which effectively generates two NIR frequencies of light using sapphire and BBO crystals. The beam is then directed into a TOPAS prime difference frequency generator which produces tunable mid-IR pulses by combining these two NIR frequencies using a AgGaS₂ crystal. These pulses feed the PHASETECH 2DQuick IR 2D-IR spectrometer which is shown in Figure 5.1. Inside the spectrometer, the pump and probe pulses are first separated using a ZnSe beam splitter (BS). The probe pulse is then further split using a CaF₂ vertical beam splitter (VBS) into probe and reference pulses which are then incident on the sample. Reference pulses are used to account for shot-to-shot variations in intensity. The pump pulse enters the germanium pulse shaper (which has been described in more detail in Chapter 1) where it is split into two pulses separated by a time τ , the desired waiting time is then achieved by adjusting the delay stage to change the length of

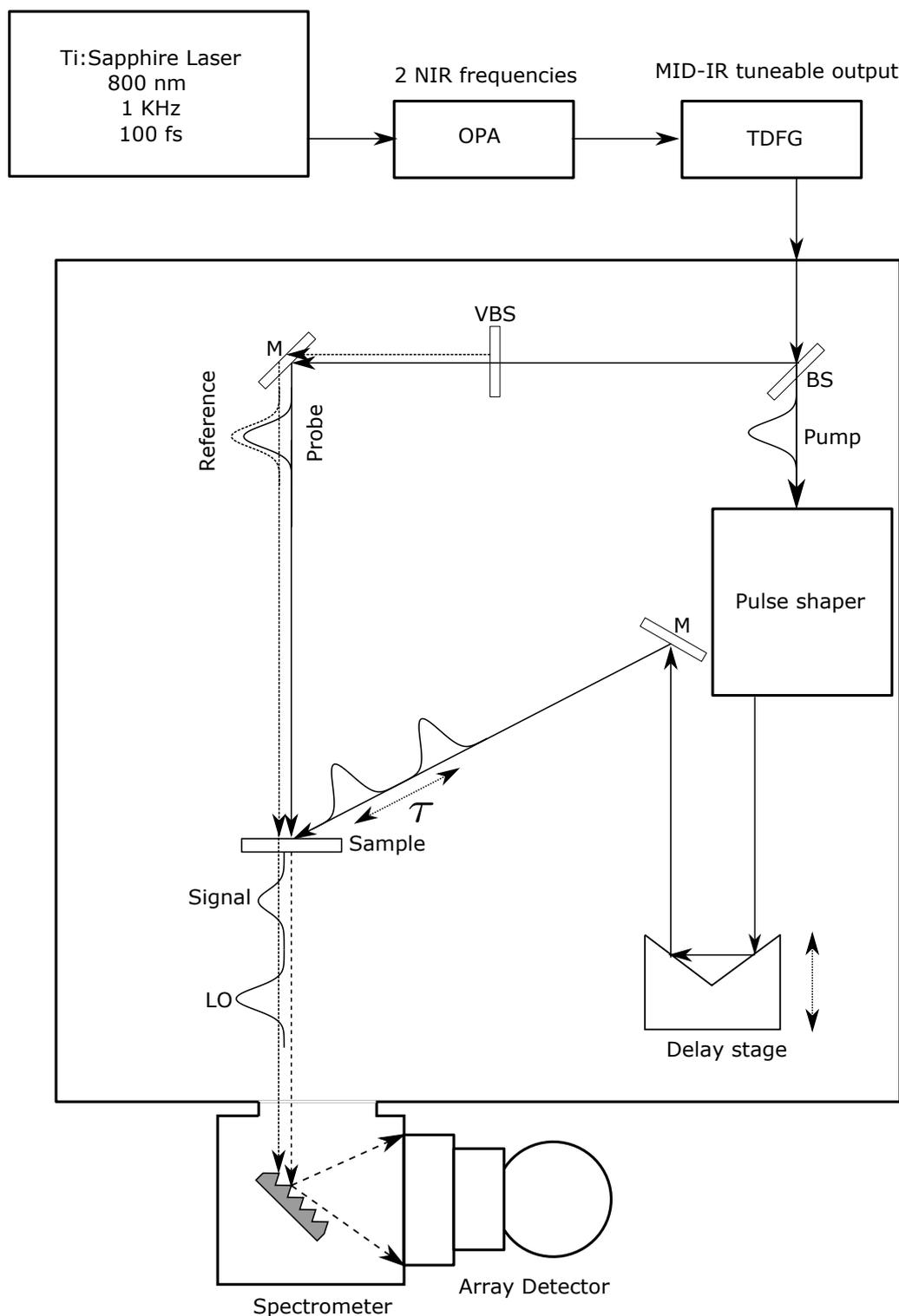


Fig. 5.1 A diagram of the Nottingham 2D-IR setup showing the pulsed laser feeding the OPA and DFG which produce tunable mid-IR ultrafast pulses. This feeds the PhaseTech 2DQuick IR spectrometer. The pump and probe beams are initially split using a ZnSe beam splitter (BS). The pump pulse then enters the Ge pulse shaper which applies the relevant masks and results in two pump pulses separated by a time τ which is then incident on the sample. The probe follows a different path which first splits it into probe and reference pulses using a vertical beam splitter (VBS). The probe pulse is then incident on the sample separated by a waiting time controlled by a delay stage in the pump path. The signal is then projected onto an MCT detector using a spectrograph.

the optical path for the pump pulses. A signal is then generated after the two pump pulses are incident on the sample which is then followed by the probe pulse after a set waiting time (assuming the the pump and probe pulses are spatially overlapped). The resultant signal, which is collinear with the probe pulse, is then directed into a spectrograph with a grating of either 30 or 75 g/mm (resolution = *ca.* 3 or 1 cm^{-1} respectively) and finally onto a focal plane array MCT detector. The variable τ is then scanned over a range of values (typically 0-4000 fs) and the phase of the two pump pulses are cycled in order to collect an entire 2D-IR spectrum. The operating procedures for the spectrometer are provided in Appendix A.

5.2 2D-IR spectroscopy data analysis and processing

5.2.1 Frequency domain

Data processing was performed using python (v3.7). A pump calibration was first performed in order to calculate the excitation frequency and intensity. Reference pump spectra were obtained through a pinhole aperture at each frequency. A Gaussian peak was fitted to each pump frequency and the central frequency and total area obtained from this was used to normalise the 2D-IR spectra obtained subsequently.

For 2D-IR spectra, the reference beam was first subtracted from the difference spectrum, this accounts for shot-to-shot intensity variations. The negative delay was then subtracted from the positive delays, and lastly a first order baseline correction was applied to the spectra before being normalised by the pump pulse intensity. Finally, the pump axis was calibrated

using a polystyrene standard. Contour plots with a pump and probe axis were then produced to display the 2D-IR spectra.

5.2.2 Time domain

Data processing was again performed using python (v3.7). The software, QuickControl, output 2D-IR spectral data with the pump axis in the time-domain and the probe axis in the frequency domain. A background correction was applied to the spectra at each delay in the pump sequence, an apodizing filter (Hamming window) was then applied before finally being Fourier transformed in the pump axis. A calibration of the probe axis was performed as above using a polystyrene standard and the pump axis calibration was achieved using a linear fit of the spectrum diagonal versus the calibrated probe axis. Contour plots with a pump and probe axis were then produced to display the 2D-IR spectra.

Exponential decay or growth fitting

Kinetic fits of data were performed using equation 5.1 where a is the amplitude, τ is the time constant, t is the time delay and i the number of exponential functions in the fit.* The function curvefit was used from the SciPy toolbox (v1.5.4), which is a non-linear least squares algorithm, to optimise the fit.

$$f(t) = \sum_i a_i e^{-t/\tau_i} \quad (5.1)$$

*One exponential fit required an additional parameter acting as a baseline as the data did not decay to zero.

Global fitting

Global fitting of spectra was performed using equation 5.2 where $H(t)$ is the Heavside step function, $G(t)$ is the Gaussian function, a is the amplitude, τ is a time constant, i the number of exponential functions in the fit and j is the number of kinetic traces in the fit. The L-BFGS-B algorithm was used as part of the SciPy minimize function (v1.5.4) to optimise the fit.

$$f(t) = H(t) * \left(G(t) \otimes \sum_{i,j} a_{i,j} e^{-t/\tau_i} \right) \quad (5.2)$$

Peak fitting

Gaussian peaks were fitted to 2D-IR spectral slices using the LMFIT (v1.0.1) package in python.¹⁶⁹ Anharmonicity values were calculated by subtracting the peak maximum from the minimum (equation 5.3).

$$\Delta = \omega_{01} - \omega_{12} \quad (5.3)$$

Spectral diffusion analysis

Spectral diffusion of samples was measured by first identifying a spectral region with one positive and negative peak and is summarised in Figure 5.2. The respective maxima and minima were calculated to find the centre of these peaks (yellow dots). For the central line

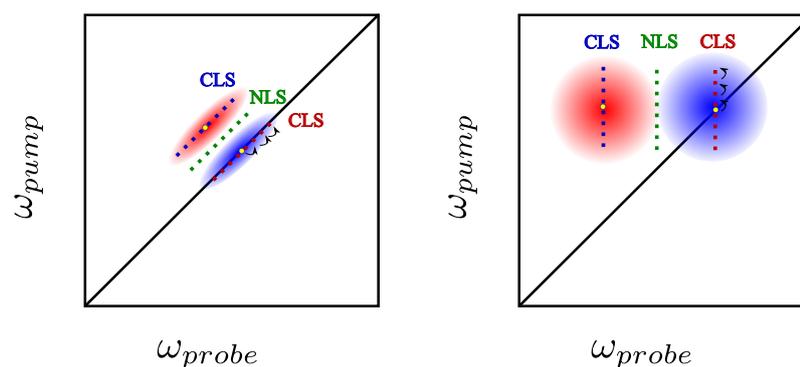


Fig. 5.2 A diagram showing how CLS and NLS are calculated. Left: $T_W = 0$. Right: $T_W > 0$. The yellow dot represents the maximum and minimum of each peak, The blue and red dotted lines represent the CLS calculated for each peak and the green dotted line presents the NLS.

slope (CLS), and taking the fundamental transition peak as an example (which is negative and blue), starting at the global minimum, the minimum is then calculated for the spectral slice one pixel above (represented by each dot in the red dotted line).[†] This is then repeated for multiple data points above and below the global minimum (curved arrows), typically 5-7 in total but is adjusted according to the sample since peak widths vary. Finally, a linear fit is calculated using these points and the gradient extracted as a result. A plot of the gradient versus waiting time is generated by applying this to each spectrum at different waiting times. This is also repeated for the $\nu_{1 \rightarrow 2}$ transition peak but in this case the maximum is of course used. The nodal line slope is defined as the zero region between the positive and negative peaks, since this region can be quite noisy it can be awkward to calculate it. Instead, using the points calculated for the CLS of both peaks, a cubic fit is calculated from the small pump spectral slices between the maximum and minimum points. The roots of this is calculated to determine the nodal points (green dotted line) and finally the gradient from a linear fit is

[†]Pixel being a data point in this case which could refer to an actual pixel in the probe axis or a data point in the pump axis.

extracted. The inverse CLS (ivCLS) is very similar to the CLS except the process is rotated 90° so that extrema are calculated from probe slices instead, and the inverse gradient is used to calculate the spectral diffusion.[‡]

5.3 FTIR spectroscopy

FTIR spectra were recorded using a Nicolet is50 spectrometer using an MCT detector with a KBr beam splitter. All spectra were recorded with a resolution 2 cm^{-1} and the number of scans and specific detail is stated below in the sample preparation section. Figure 5.3 shows a diagram of a Harrick cell which was used to collect transmission FTIR spectra and 2D-IR spectra of samples. The cell consists of a metal back plate which holds two O-rings, these sandwich and seal in two CaF_2 windows which themselves are separated by a Teflon spacer (typically $6\text{-}500\ \mu\text{m}$) and the sample solution. A flow system is also shown on the diagram where a sample solution reservoir is connected to the cell via inlet and outlet pipes and the solution is cycled through using a peristaltic pump. This setup was used for Chapter 2 where the solution needed to be continuously refreshed. A temperature controlled sample was required for Chapter 3, therefore a modified version of the Harrick cell where a heated steel block surrounded the main cell body was used. The temperature was controlled by flowing water through the steel block using a temperature controlled water bath.

[‡]The gradient was actually calculated for the dataset rotated by 90° . This was done because linear fits do not operate well when the gradient is close to 0.

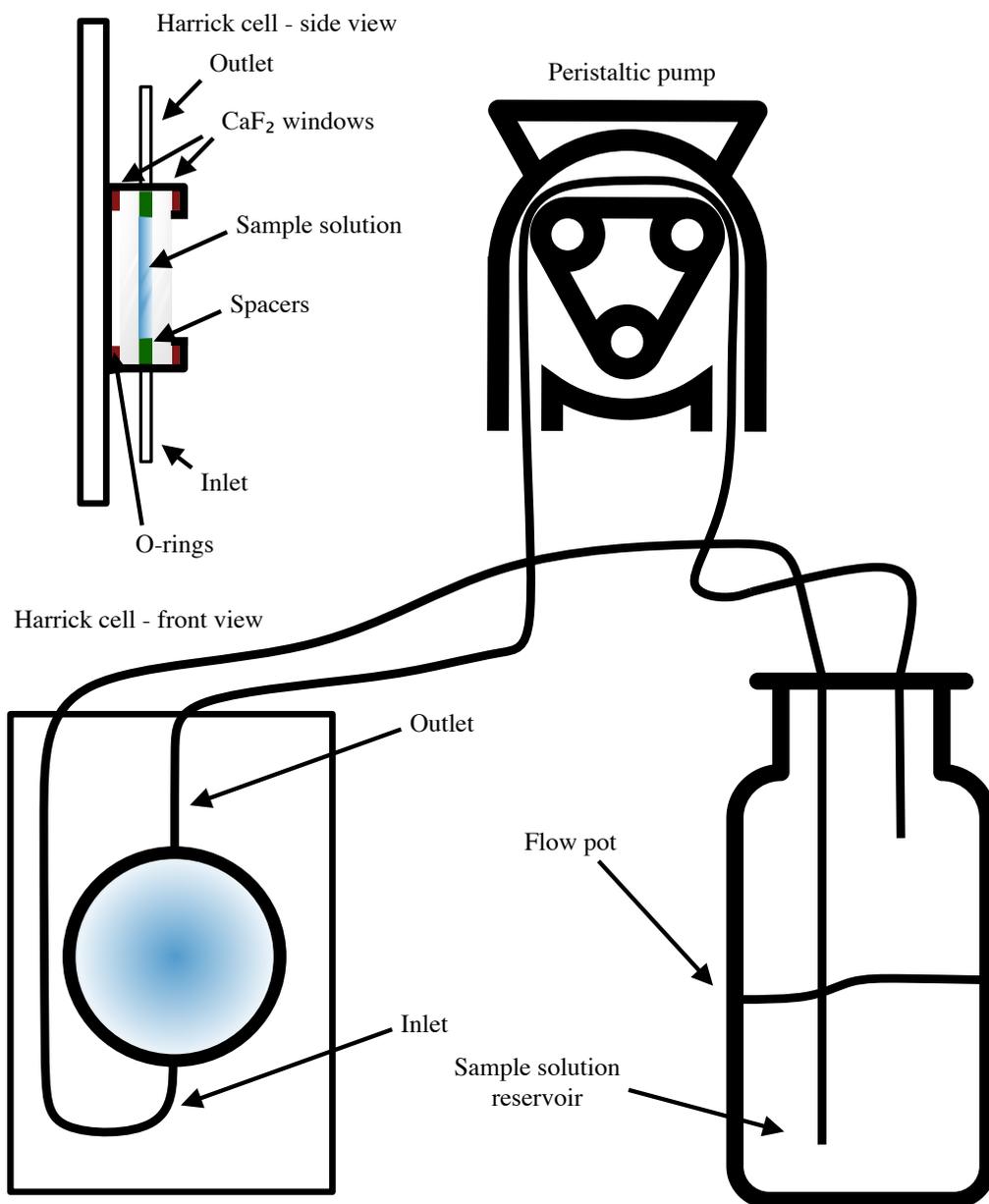


Fig. 5.3 A diagram showing the setup of a Harrick cell attached to a flow system. The flow system consists of an inlet and outlet pipe which link the sample solution reservoir in a flow pot to the Harrick cell via a peristaltic pump. The Harrick cell (with front and side view displayed) itself is formed of a back plate, where two O-rings (red) sandwich two CaF₂ windows, which themselves sandwich a Teflon spacer (green) and the sample solution.

5.3.1 MCR analysis

Multivariate curve resolution (MCR) analysis is a method used to decompose spectra into pure component spectra (components) and concentration profiles (scores). Equation 5.4 shows how spectra (D) are decomposed into components (C) and scores (S) with the residual data described by E .¹⁷⁰ MCR analysis was performed using PLS Toolbox from EigenVector software (v8.8.1). FTIR spectra were first preprocessed by performing a baseline routine followed by normalising the spectra by area in the region of interest. MCR was then performed and two factors was chosen to describe the data.

$$D = CS^T + E \quad (5.4)$$

5.3.2 Peak fitting

Pseudo-voigt peaks were fitted to data using Matlab (v2020a). Each peak was initialised with a peak amplitude, width, position and shape (a 0 to 1 scale how Gaussian or Lorentzian the peak is). Upper and lower limits were applied in some cases to limit and ensure peak size and position remained reasonable. The position of peaks for the amide I peak fitting were limited to $\pm 5 \text{ cm}^{-1}$ from their starting positions which were determined from [155]. The Non-linear Least Squares algorithm was used to perform the fit.

5.4 Sample preparation

5.4.1 Fe(CO)₅

Fe(CO)₅ was purchased from Sigma-Aldrich (CAS number: 13463-40-6) and used without further purification. Fe(CO)₅ was dissolved in *ca.* 20 ml of either dry heptane (CAS number: 142-82-5) or dry CH₂Cl₂ (CAS number: 75-09-2) such that each had an optical density (OD) of 0.5 and 0.2 respectively.[§] Each solution was prepared under an argon atmosphere in a flow system containing a Harrick cell (containing CaF₂ windows) with a sample path length of 300 μm. The sample was then flowed using a peristaltic pump.

T-2D-IR difference spectra were obtained by varying the time delay between the initial IR_{pump} and UV_{pump} pulses and finally probing the system following a waiting time using an IR_{probe} pulse. A UV excitation of 266 nm was used and the IR_{pump} was scanned from 1950 to 2040 cm⁻¹. Each spectrum was collected and averaged over 5 s for 3 cycles for a variety of time delays between the three pulses.

5.4.2 Carrageenan

5.4.3 Pure carrageenan samples

κ-carrageenan (CAS number: 1114-20-8) and ι-carrageenan (CAS number: 9062-07-1) was purchased from Sigma-Aldrich and used without further purification. Solutions of deionised

[§]Heptane was refluxed over CaH₂ under an inert argon atmosphere for 48 hours before use and dry CH₂Cl₂ was purchased from Sigma-Aldrich in a sealable container.

water and 2% w/w carrageenan were prepared at 80 °C by dissolving the carrageenan. Drops of the hot solution were then placed in a preheated variable temperature (modified Harrick) cell containing CaF₂ windows, with 12 μm spacers.

2D-IR spectra in the frequency domain were then obtained at temperatures of 70, 45 and 22 °C. Difference spectra were collected at -10 and 1 ps delays with a scanned pump range of 1100 - 1450 cm⁻¹ and a bandwidth of *ca.* 20 cm⁻¹. Each spectrum was collected and averaged over 3 s for 10 cycles.

5.4.4 Carrageenan and ferrocyanide

κ-carrageenan (CAS number: 1114-20-8) and ι-carrageenan (CAS number: 9062-07-1) was purchased from Sigma-Aldrich and used without further purification, along with potassium ferrocyanide (CAS number: 14459-95-1). Solutions of 4% w/w carrageenan with *ca.* 0.18 M potassium ferrocyanide were prepared by mixing them with deionised water at 80 °C. Drops of the hot solution were then placed in a preheated variable temperature (modified Harrick) cell containing CaF₂ windows, with 25 μm spacers.

FTIR spectra were obtained using a Nicolet is50 spectrometer, with a resolution of 2 cm⁻¹ and 1024 sample scans at temperatures of 20, 30, 40, 50, 60, 70, 80 °C. 2D-IR spectra in the time domain using the PhaseTech spectrometer were then obtained at the same range of temperatures in the CN stretching region of the spectrum. Four phase cycling with a pump sequence scanned from 0 - 4000 fs with steps of 16 fs was used. A waiting time range of -5 to 25 ps was collected. 10 spectra were averaged together at each waiting time.

5.4.5 Mucins and theaflavin

β -Lactoglobulin (β -LG)(CAS number: 9045-23-2), epigallocatechin gallate (EGCG)(CAS number: 989-51-5) and Tris-HCl (CAS number: 1185-53-1) buffer was purchased from Sigma-Aldrich and used without further purification. A solution of 0.25 mM β -LG and Tris-HCl was prepared in H₂O, the solution was then freeze dried and replaced with D₂O to exchange the labile protons. This was repeated a further two times. The exchanged β -LG solution was then measured using FTIR and 2D-IR spectroscopy in a Harrick cell with a pathlength of 50 μ m. Finally, EGCG was added to the exchanged β -LG solution to form a final concentration of 0.5 mM EGCG and 0.25 mM β -LG, which was again measured using FTIR and 2D-IR spectroscopy.

The NT5B protein was provided by Elena Owen at the University of Manchester and the theaflavin extracts were provided by Unilever. The protein extraction process is described in [159]. NT5B was exchanged with D₂O three times by freezing drying the sample. 2D-IR and FTIR spectroscopy was performed on the pure protein with the sample contained within a Harrick cell with a pathlength of 50 μ m. Theaflavin digallate (THDG) was then added to the NT5B protein solution to an approximate concentration of 0.5 mM.[¶] 2D-IR and FTIR spectroscopy was then performed with the NT5B and THDG sample.

FTIR spectra were obtained using a Nicolet is50 spectrometer, with a resolution of 2 cm⁻¹ and 64 sample scans. 2D-IR spectra were obtained in the time domain using the PhaseTech spectrometer. Two phase cycling and a pump sequence consisting of delays 0 - 4000 fs with

[¶]An exact concentration of THDG was very difficult to achieve because only a very small volume of NT5B was available. (250 μ l) and so the uncertainty in the concentration was determined to be large.

intervals of 20 fs. 50 spectra were averaged for each waiting time. Spectra were obtained for waiting times of -0.5 to 2 ps. Four phase cycling and a pump sequence consisting of delays 0 - 2544 fs with intervals of 24 fs was used for the NT5B - THDG sample because large scattering artefacts were present initially.

References

- [1] J. Cavanagh, W. J. Fairbrother, A. G. Palmer III and N. J. Skelton, *Protein NMR spectroscopy: principles and practice*, Academic Press, 1995.
- [2] H. Mantsch, *Infrared spectroscopy of biomolecules*, J. Wiley, 1996.
- [3] J. Zheng, K. Kwak and M. D. Fayer, *Acc. Chem. Res.*, 2007, **40**, 75–83.
- [4] P. Hamm and M. Zanni, *Concepts and Methods of 2D Infrared Spectroscopy*, Cambridge University Press, 2011, pp. 1–298.
- [5] M. Towrie, D. C. Grills, J. Dyer, J. A. Weinstein, P. Matousek, R. Barton, P. D. Bailey, N. Subramaniam, W. M. Kwok, C. Ma, D. Phillips, A. W. Parker and M. W. George, *Appl. Spectrosc.*, 2003, **57**, 367–380.
- [6] G. M. Greetham, P. Burgos, Q. Cao, I. P. Clark, P. S. Codd, R. C. Farrow, M. W. George, M. Kogimtzis, P. Matousek, A. W. Parker *et al.*, *Appl. Spectrosc.*, 2010, **64**, 1311–1319.
- [7] M. D. Fayer, *Ultrafast Infrared Vib. Spectrosc.*, CRC Press, 2013.
- [8] J. J. Turner, in *Bandwidths*, ed. J. J. Turner, John Wiley & Sons, Ltd, 2006, pp. 101–127.
- [9] J. R. Schmidt, N. Sundlass and J. L. Skinner, *Chem. Phys. Lett.*, 2003, **378**, 559–566.
- [10] W. Rock, Y.-L. Li, P. Pagano and C. M. Cheatum, *J. Phys. Chem. A*, 2013, **117**, 6073–6083.
- [11] L. P. DeFlores, R. A. Nicodemus and A. Tokmakoff, *Opt. Lett.*, 2007, **32**, 2966–2968.
- [12] S. Woutersen and P. Hamm, *J. Phys.: Condens. Matter*, 2002, **14**, R1035–R1062.
- [13] S.-H. Shim and M. T. Zanni, *Phys. Chem. Chem. Phys.*, 2009, **11**, 748–761.
- [14] S.-H. Shim, D. B. Strasfeld, E. C. Fulmer and M. T. Zanni, *Opt. Lett.*, 2006, **31**, 838–840.
- [15] S.-H. Shim, D. B. Strasfeld and M. T. Zanni, *Opt. Express*, 2006, **14**, 13120–13130.
- [16] A. Krummel, *Pulse shaper*, <http://sites.chem.colostate.edu/krummellab/research.html>, Accessed: 23-06-2020.

- [17] N. T. Hunt, *Chem. Soc. Rev.*, 2009, **38**, 1837–1848.
- [18] A. L. Le Sueur, R. E. Horness and M. C. Thielges, *Analyst*, 2015, **140**, 4336–4349.
- [19] C.-I. Brändén, J. Tooze and J. C. R. T. UK), *Introduction to Protein Structure*, Taylor & Francis, 1999.
- [20] J. M. Berg, J. L. Tymoczko and L. Stryer, *Biochemistry: International Edition*, W. H. Freeman, New York, NY, 7th edn., 2011.
- [21] A. Ghosh, J. S. Ostrander and M. T. Zanni, *Chem. Rev.*, 2017, **117**, 10726–10759.
- [22] T. Hayashi and S. Mukamel, *J. Mol. Liq.*, 2008, **141**, 149–154.
- [23] F. Mallamace, C. Corsaro, D. Mallamace, S. Vasi, C. Vasi and G. Dugo, *Comput. Struct. Biotechnol. J.*, 2015, **13**, 33–37.
- [24] A. Barth and C. Zscherp, *Q. Rev. Biophys.*, 2002, **35**, 369–430.
- [25] H. Yang, S. Yang, J. Kong, A. Dong and S. Yu, *Nat. Protoc.*, 2015, **10**, 382–396.
- [26] M. Jackson and H. H. Mantsch, *Critical Reviews in Biochemistry and Molecular Biology*, 2008.
- [27] N. Demirdöven, C. M. Cheatum, H. S. Chung, M. Khalil, J. Knoester and A. Tokmakoff, *J. Am. Chem. Soc.*, 2004, **126**, 7981–7990.
- [28] M. Grechko and M. T. Zanni, *The J. Chem. Phys.*, 2012, **137**, 184202.
- [29] E. B. Dunkelberger, M. Grechko and M. T. Zanni, *The J. Phys. Chem. B*, 2015, **119**, 14065–14075.
- [30] L. Minnes, D. J. Shaw, B. P. Cossins, P. M. Donaldson, G. M. Greetham, M. Towrie, A. W. Parker, M. J. Baker, A. J. Henry, R. J. Taylor and N. T. Hunt, *Anal. Chem.*, 2017, **89**, 10898–10906.
- [31] S. D. Moran, T. O. Zhang and M. T. Zanni, *Protein Sci.*, 2014, **23**, 321–331.
- [32] L. P. DeFlores, Z. Ganim, S. F. Ackley, H. S. Chung and A. Tokmakoff, *J. Phys. Chem. B*, 2006, **110**, 18973–18980.
- [33] D. Kraemer, M. L. Cowan, A. Paarmann, N. Huse, E. T. J. Nibbering, T. Elsaesser and R. J. D. Miller, *Proc Natl Acad Sci U S A*, 2008, **105**, 437–442.
- [34] F. Perakis and P. Hamm, *Phys. Chem. Chem. Phys.*, 2012, **14**, 6250–6256.
- [35] H.-S. Tan, I. R. Piletic, R. E. Riter, N. E. Levinger and M. D. Fayer, *Phys. Rev. Lett.*, 2005, **94**, 057405.
- [36] K. Jones, *Ph.D. thesis*, Massachusetts Institute of Technology, 2012.
- [37] N. T. Hunt, *Dalton Trans.*, 2014, **43**, 17578–17589.

- [38] A. I. Stewart, J. A. Wright, G. M. Greetham, S. Kaziannis, S. Santabarbara, M. Towrie, A. W. Parker, C. J. Pickett and N. T. Hunt, *Inorg. Chem.*, 2010, **49**, 9563–9573.
- [39] L. M. Kiefer, J. T. King and K. J. Kubarych, *Acc. Chem. Res.*, 2015, **48**, 1123–1130.
- [40] J. Bredenbeck, J. Helbing and P. Hamm, *J. Am. Chem. Soc.*, 2004, **126**, 990–991.
- [41] C. E. Wayne, *Photochemistry / Carol E. Wayne, Richard P. Wayne.*, Oxford : Oxford University Press, Oxford, 1996.
- [42] M. Poliakoff and J. J. Turner, *Angew. Chem. Int. Ed.*, 2001, **40**, 2809–2812.
- [43] L. J. G. W. van Wilderen and J. Bredenbeck, *Angew. Chem. Int. Ed.*, 2015, **54**, 11624–11640.
- [44] M. Delor, P. A. Scattergood, I. V. Sazanovich, A. W. Parker, G. M. Greetham, A. J. H. M. Meijer, M. Towrie and J. A. Weinstein, *Science*, 2014, **346**, 1492–1495.
- [45] M. Delor, T. Keane, P. A. Scattergood, I. V. Sazanovich, G. M. Greetham, M. Towrie, A. J. H. M. Meijer and J. A. Weinstein, *Nat. Chem.*, 2015, **7**, 689–695.
- [46] D. Kern-Michler, C. Neumann, N. Mielke, L. J. G. W. van Wilderen, M. Reinfelds, J. von Cosel, F. Santoro, A. Heckel, I. Burghardt and J. Bredenbeck, *J. Am. Chem. Soc.*, 2018, **140**, 926–931.
- [47] N. Leadbeater, *Coord. Chem. Rev.*, 1999, **188**, 35–70.
- [48] M. Poliakoff and J. J. Turner, *J. Chem. Soc., Dalton Trans.*, 1974, 2276–2285.
- [49] A. J. Ouderkerk, P. Wermer, N. L. Schultz and E. Weitz, *J. Am. Chem. Soc.*, 1983, **105**, 3354–3355.
- [50] P. Portius, J. Yang, X.-Z. Sun, D. C. Grills, P. Matousek, A. W. Parker, M. Towrie and M. W. George, *J. Am. Chem. Soc.*, 2004, **126**, 10713–10720.
- [51] D. M. Hayes and E. Weitz, *The J. Phys. Chem.*, 1991, **95**, 2723–2727.
- [52] G. T. Long and E. Weitz, *J. Am. Chem. Soc.*, 2000, **122**, 1431–1442.
- [53] M. Poliakoff and J. J. Turner, *J. Chem. Soc., Dalton Trans.*, 1973, 1351–1357.
- [54] J. K. Burdett, *J. Chem. Soc., Faraday Trans. 2: Molecular and Chem. Phys.*, 1974, **70**, 1599–1613.
- [55] T. J. Barton, R. Grinter, A. J. Thomson, B. Davies and M. Poliakoff, *J. Chem. Soc., Chemical Communications*, 1977, 841–842.
- [56] C. Daniel, M. Benard, A. Dedieu, R. Wiest and A. Veillard, *The J. Phys. Chem.*, 1984, **88**, 4805–4811.
- [57] M. Poliakoff, *J. Chem. Soc., Dalton Trans.*, 1974, 210–212.
- [58] L. A. Barnes, M. Rosi and C. W. Bauschlicher, *The J. Chem. Phys.*, 1991, **94**, 2031–2039.

- [59] A. J. Ouderkirk and E. Weitz, *The J. Chem. Phys.*, 1983, **79**, 1089–1091.
- [60] T. A. Seder, A. J. Ouderkirk and E. Weitz, *The J. Chem. Phys.*, 1986, **85**, 1977–1986.
- [61] R. J. Ryther and E. Weitz, *The J. Phys. Chem.*, 1991, **95**, 9841–9852.
- [62] R. J. Ryther and E. Weitz, *The J. Phys. Chem.*, 1992, **96**, 2561–2567.
- [63] P. C. Engelking and W. C. Lineberger, *J. Am. Chem. Soc.*, 1979, **101**, 5569–5573.
- [64] S. A. Trushin, W. Fuss, K. L. Kompa and W. E. Schmid, *J. Phys. Chem. A*, 2000, **104**, 1997–2006.
- [65] P. Wernet, T. Leitner, I. Josefsson, T. Mazza, P. S. Miedema, H. Schröder, M. Beye, K. Kunnus, S. Schreck, P. Radcliffe, S. Düsterer, M. Meyer, M. Odellius and A. Föhlich, *J. Chem. Phys.*, 2017, **146**, 211103.
- [66] M. Bergt, T. Brixner, B. Kiefer, M. Strehle and G. Gerber, *J. Phys. Chem. A*, 1999, **103**, 10381–10387.
- [67] S. P. Church, F.-W. Grevels, H. Hermann, J. M. Kelly, W. E. Klotzbücher and K. Schaffner, *J. Chem. Soc., Chemical Communications*, 1985, 594–596.
- [68] P. T. Snee, C. K. Payne, K. T. Kotz, H. Yang and C. B. Harris, *J. Am. Chem. Soc.*, 2001, **123**, 2255–2264.
- [69] V. Bachler, F.-W. Grevels, K. Kerpen, G. Olbrich and K. Schaffner, *Organometallics*, 2003, **22**, 1696–1711.
- [70] S. C. Nguyen, J. P. Lomont, M. C. Zoerb, A. D. Hill, J. P. Schlegel and C. B. Harris, *Organometallics*, 2012, **31**, 3980–3984.
- [71] P. Wernet, K. Kunnus, I. Josefsson, I. Rajkovic, W. Quevedo, M. Beye, S. Schreck, S. Grübel, M. Scholz, D. Nordlund, W. Zhang, R. W. Hartsock, W. F. Schlotter, J. J. Turner, B. Kennedy, F. Hennies, F. M. F. de Groot, K. J. Gaffney, S. Techert, M. Odellius and A. Föhlich, *Nature*, 2015, **520**, 78–81.
- [72] J. F. Cahoon, K. R. Sawyer, J. P. Schlegel and C. B. Harris, *Science*, 2008, **319**, 1820–1823.
- [73] M.-C. Heitz and C. Daniel, *J. Am. Chem. Soc.*, 1997, **119**, 8269–8275.
- [74] J. F. Cahoon, K. R. Sawyer, J. P. Schlegel and C. B. Harris, *Science*, 2008, **319**, 1820–1823.
- [75] Q. Guo, P. Pagano, Y.-L. Li, A. Kohen and C. M. Cheatum, *The J. Chem. Phys.*, 2015, **142**, 212427.
- [76] P. Portius, J. Yang, X.-Z. Sun, D. C. Grills, P. Matousek, A. W. Parker, M. Towrie and M. W. George, *J. Am. Chem. Soc.*, 2004, **126**, 10713–10720.
- [77] W. Thomas, in *Thickening and gelling agents for food*, Springer, 1997, pp. 45–59.

- [78] V. Gaukel, A. Leiter and W. E. L. Spieß, *J. Food Eng.*, 2014, **141**, 44–50.
- [79] M. Tako, *J. Appl. Glycosci.*, 2000, **47**, 49–53.
- [80] D. Lecacheux, R. Panaras, G. Brigand and G. Martin, *Carbohydr. Polym.*, 1985, **5**, 423–440.
- [81] J. Necas, L. Bartosikova *et al.*, *Vet Med*, 2013, **58**, 187–205.
- [82] M. Bosco, A. Segre, S. Miertus, A. Cesaro and S. Paoletti, *Carbohydr. Res.*, 2005, **340**, 943–958.
- [83] C. Viebke, L. Piculell and S. Nilsson, *Macromolecules*, 1994, **27**, 4160–4166.
- [84] S. Janaswamy and R. Chandrasekaran, *Carbohydr. Res.*, 2002, **337**, 523–535.
- [85] M. Tako, *Advances in Bioscience and Biotechnology*, 2015, **6**, 22–36.
- [86] P. Wu and M. Imai, *Advancing Desalination*, 2012.
- [87] C. Rochas and M. Rinaudo, *Biopolymers*, 1984, **23**, 735–745.
- [88] A. Michel, M. Mestdagh and M. Axelos, *InterNatl. J. (Wash.) of Biological Macromolecules*, 1997, **21**, 195–200.
- [89] C. Viebke, J. Borgström, I. Carlsson, L. Piculell and P. Williams, *Macromolecules*, 1998, **31**, 1833–1841.
- [90] M. Robal, T. Brenner, S. Matsukawa, H. Ogawa, K. Truus, B. Rudolph and R. Tuvikene, *Food Hydrocolloids*, 2017, **63**, 656–667.
- [91] M. R. Mangione, D. Giacomazza, D. Bulone, V. Martorana, G. Cavallaro and P. L. San Biagio, *Biophys. Chem.*, 2005, **113**, 129–135.
- [92] H. Grasdalen and O. Smidsroed, *Macromolecules*, 1981, **14**, 1842–1845.
- [93] C. A. Running, R. Falshaw and S. Janaswamy, *Carbohydr. Polym.*, 2012, **87**, 2735–2739.
- [94] S. Bayley, *Biochim. Biophys. Acta*, 1955, **17**, 194–205.
- [95] N. S. Anderson, J. W. Campbell, M. M. Harding, D. A. Rees and J. W. B. Samuel, *Journal of Mol. Biol.*, 1969, **45**, 85–97.
- [96] S. Arnott, W. E. Scott, D. A. Rees and C. G. A. McNab, *Journal of Mol. Biol.*, 1974, **90**, 253–267.
- [97] L. Schefer, J. Adamcik, M. Diener and R. Mezzenga, *Nanoscale*, 2015, **7**, 16182–16188.
- [98] M. Gobet, M. Mouaddab, N. Cayot, J.-M. Bonny, E. Guichard, J.-L. Le Quéré, C. Moreau and L. Foucat, *Magn. Reson. Chem.*, 2009, **47**, 307–312.
- [99] V. Meunier, T. Nicolai and D. Durand, *Macromolecules*, 2000, **33**, 2497–2504.

- [100] D. Yang, S. Gao and H. Yang, *Food Hydrocolloids*, 2020, **99**, 105317.
- [101] M. O. Eleya and S. Turgeon, *Food Hydrocolloids*, 2000, **14**, 245–251.
- [102] V. Morris and G. Chilvers, *Carbohydr. Polym.*, 1983, **3**, 129–141.
- [103] A.-M. Hermansson, E. Eriksson and E. Jordansson, *Carbohydr. Polym.*, 1991, **16**, 297–320.
- [104] E. M. Vilen, L. C. E. Lundqvist, D. Jouanneau, W. Helbert and C. Sandstrom, *Biomacromolecules*, 2010, **11**, 3487–3494.
- [105] M. V. C. Cardoso and E. Sabadini, *Carbohydr. Res.*, 2010, **345**, 2368–2373.
- [106] C. Naumann and P. William Kuchel, *Polym. Chem.*, 2010, **1**, 1109–1116.
- [107] P. Belton, R. Wilson and D. Chenery, *InterNat. J. (Wash.) of Biological Macromolecules*, 1986, **8**, 247–251.
- [108] P. S. Belton, B. J. Goodfellow and R. H. Wilson, *Macromolecules*, 1989, **22**, 1636–1642.
- [109] C. Rochas, M. Lahaye and W. Yaphe, *Botanica Marina*, 1986, **29**, 335–340.
- [110] M. Sekkal and P. Legrand, *Spectrochim. Acta Part A: Molecular Spectroscopy*, 1993, **49**, 209–221.
- [111] M. Sekkal, P. Legrand, J. P. Huvenne and M. C. Verdus, *J. Mol. Struct.*, 1993, **294**, 227–230.
- [112] T. Chopin and E. Whalen, *Carbohydr. Res.*, 1993, **246**, 51–59.
- [113] C. Welsh, *Ph.D. thesis*, University of Nottingham, 2017.
- [114] J. Prado-Fernandez, J. Rodriguez-Vazquez, E. Tojo and J. Andrade, *Anal. Chim. Acta*, 2003, **480**, 23 – 37.
- [115] I. T. Norton, D. M. Goodall, E. R. Morris and D. A. Rees, *J. Chem. Soc., Faraday Trans. 1: Physical Chemistry in Condensed Phases*, 1983, **79**, 2475–2488.
- [116] O. N. Makshakova, D. A. Faizullin and Y. F. Zuev, *Carbohydr. Polym.*, 2020, **227**, 115342.
- [117] J. Mo, R. D. Groot, G. McCartney, E. Guo, J. Bent, G. van Dalen, P. Schuetz, P. Rockett and P. D. Lee, *Crystals*, 2019, **9**, 321.
- [118] B. Ratajska-Gadomska and W. Gadomski, *The J. Chem. Phys.*, 2004, **121**, 12583–12588.
- [119] S. Kara, C. Tamerler, H. Bermek and Ö. Pekcan, *Journal of Bioactive and Compatible Polymers*, 2003, **18**, 33–44.
- [120] I. T. Norton, E. R. Morris and D. A. Rees, *Carbohydr. Res.*, 1984, **134**, 89–101.

- [121] R. A. Nicodemus, K. Ramasesha, S. T. Roberts and A. Tokmakoff, *The J. Phys. Chem. Lett.*, 2010, **1**, 1068–1072.
- [122] J. Teixeira, A. Luzar and S. Longeville, *J. Phys.: Condens. Matter*, 2006, **18**, S2353.
- [123] G. Prampolini, P. Yu, S. Pizzanelli, I. Cacelli, F. Yang, J. Zhao and J. Wang, *The J. Phys. Chem. B*, 2014, **118**, 14899–14912.
- [124] K. Ohta, H. Maekawa and K. Tominaga, *J. Phys. Chem. A*, 2004, **108**, 1333–1341.
- [125] T. Seki, K.-Y. Chiang, C.-C. Yu, X. Yu, M. Okuno, J. Hunger, Y. Nagata and M. Bonn, *The J. Phys. Chem. Lett.*, 2020, **11**, 8459–8469.
- [126] J. T. King, M. R. Ross and K. J. Kubarych, *Phys. Rev. Lett.*, 2012, **108**, 157401.
- [127] M. D. Fayer, D. E. Moilanen, D. Wong, D. E. Rosenfeld, E. E. Fenn and S. Park, *Acc. Chem. Res.*, 2009, **42**, 1210–1219.
- [128] D. Laage and J. T. Hynes, *Science*, 2006, **311**, 832–835.
- [129] I. Waluyo, D. Nordlund, U. Bergmann, L. G. M. Pettersson and A. Nilsson, *The J. Chem. Phys.*, 2009, **131**, 031103.
- [130] Y. Wang, C. Yuan, B. Cui and Y. Liu, *Carbohydr. Polym.*, 2018, **202**, 530–535.
- [131] H. S. Chung, M. Khalil, A. W. Smith and A. Tokmakoff, *Rev. Sci. Instrum.*, 2007, **78**, 063101.
- [132] X.-X. Zhang, K. C. Jones, A. Fitzpatrick, C. S. Peng, C.-J. Feng, C. R. Baiz and A. Tokmakoff, *The J. Phys. Chem. B*, 2016, **120**, 5134–5145.
- [133] B. Ashwood, N. H. C. Lewis, P. J. Sanstead and A. Tokmakoff, *The J. Phys. Chem. B*, 2020, **124**, 8665–8677.
- [134] W. Edgar, C. Dawes and D. O’Mullane, *Saliva and Oral Health*, British Dental Association, 2004.
- [135] S. J. Haward, J. A. Odell, M. Berry and T. Hall, *Rheol. Acta*, 2011, **50**, 869–879.
- [136] H. S. Davies, P. D. A. Pudney, P. Georgiades, T. A. Waigh, N. W. Hodson, C. E. Ridley, E. W. Blanch and D. J. Thornton, *PLoS One*, 2014, **9**, 1–13.
- [137] M. G. Roy, A. Livraghi-Butrico, A. A. Fletcher, M. M. McElwee, S. E. Evans, R. M. Boerner, S. N. Alexander, L. K. Bellinghausen, A. S. Song, Y. M. Petrova, M. J. Tuvim, R. Adachi, I. Romo, A. S. Bordt, M. G. Bowden, J. H. Sisson, P. G. Woodruff, D. J. Thornton, K. Rousseau, M. M. De la Garza, S. J. Moghaddam, H. Karmouty-Quintana, M. R. Blackburn, S. M. Drouin, C. W. Davis, K. A. Terrell, B. R. Grubb, W. K. O’Neal, S. C. Flores, A. Cota-Gomez, C. A. Lozupone, J. M. Donnelly, A. M. Watson, C. E. Hennessy, R. C. Keith, I. V. Yang, L. Barthel, P. M. Henson, W. J. Janssen, D. A. Schwartz, R. C. Boucher, B. F. Dickey and C. M. Evans, *Nature*, 2014, **505**, 412–416.
- [138] E. F. Gamarro, *Salivary Mucins*, <http://flipper.diff.org/app/items/255>, Accessed: 19-03-2020.

- [139] P. Verdugo, *Cold Spring Harbour Perspect. Med.*, 2012, **2**, a009597.
- [140] M. Aitken and P. Verdugo, *Symp. Soc. Exp. Biol.*, 1989, **43**, 73–80.
- [141] H. S. Davies, P. Singh, T. Deckert-Gaudig, V. Deckert, K. Rousseau, C. E. Ridley, S. E. Dowd, A. J. Doig, P. D. A. Pudney, D. J. Thornton and E. W. Blanch, *Anal. Chem.*, 2016, **88**, 11609–11615.
- [142] C. Ridley, N. Kouvatso, B. D. Raynal, M. Howard, R. F. Collins, J.-L. Dessey, T. A. Jowitt, C. Baldock, C. W. Davis, T. E. Hardingham and D. J. Thornton, *J. Biol. Chem.*, 2014, **289**, 16409–16420.
- [143] A. Zalewska, K. Zwierz, K. Zólkowski and A. Gindzieński, *Acta Biochim. Pol.*, 2000, **47**, 1067–1079.
- [144] K. A. Thomsson, A. Prakobphol, H. Leffler, M. S. Reddy, M. J. Levine, S. J. Fisher and G. C. Hansson, *Glycobiology*, 2002, **12**, 1–14.
- [145] H. N. Graham, *Prev. Med.*, 1992, **21**, 334 – 350.
- [146] J. V. Higdon and B. Frei, *Crit. Rev. Food Sci. Nutr.*, 2003, **43**, 89–143.
- [147] M.-C. Menet, S. Sang, C. S. Yang, C.-T. Ho and R. T. Rosen, *J. Agric. Food. Chem.*, 2004, **52**, 2455–2461.
- [148] L. K. Leung, Y. Su, R. Chen, Z. Zhang, Y. Huang and Z.-Y. Chen, *J. Nutr.*, 2001, **131**, 2248–2251.
- [149] E. Haslam, *Biochem. J.*, 1974, **139**, 285–288.
- [150] J. M. Crowther, G. B. Jameson and A. J. H. a. R. C. J. Dobson, *Milk Proteins - From Structure to Biological Properties and Health Aspects*, 2016.
- [151] D. Rossetti, G. E. Yakubov, J. R. Stokes, A. M. Williamson and G. G. Fuller, *Food Hydrocolloids*, 2008, **22**, 1068–1078.
- [152] J. Ye, F. Fan, X. Xu and Y. Liang, *Food Res. Int.*, 2013, **53**, 449–455.
- [153] B. Y. Qin, M. C. Bewley, L. K. Creamer, H. M. Baker, E. N. Baker and G. B. Jameson, *Biochemistry*, 1998, **37**, 14014–14023.
- [154] M. Sahihi, Z. Heidari-Koholi and A.-K. Bordbar, *Journal of Macromolecular Science, Part B*, 2012, **51**, 2311–2323.
- [155] C. D. Kanakis, I. Hasni, P. Bourassa, P. A. Tarantilis, M. G. Polissiou and H.-A. Tajmir-Riahi, *Food Chem.*, 2011, **127**, 1046–1055.
- [156] V. Carnovale, M. Britten, C. Couillard and L. Bazinet, *Food Res. Int.*, 2015, **77**, 565–571.
- [157] M. Stojadinovic, J. Radosavljevic, J. Ognjenovic, J. Vesic, I. Prodic, D. Stanic-Vucinic and T. Cirkovic Velickovic, *Food Chem.*, 2013, **136**, 1263–1271.

- [158] P. Georgiades, P. D. A. Pudney, S. Rogers, D. J. Thornton and T. A. Waigh, *PLOS ONE*, 2014, **9**, e105302.
- [159] E. Owen, *Ph.D. thesis*, University of Manchester, 2019.
- [160] Y. Zhao, L. Chen, G. Yakubov, T. Aminiafshar, L. Han and G. Lian, *J. Phys. Chem. B*, 2012, **116**, 13010–13016.
- [161] K. Sakurai, T. Konuma, M. Yagi and Y. Goto, *Biochim. Biophys. Acta (BBA) - General Subjects*, 2009, **1790**, 527–537.
- [162] J. P. Lomont, K. L. Rich, M. Maj, J.-J. Ho, J. S. Ostrander and M. T. Zanni, *The J. Phys. Chem. B*, 2018, **122**, 144–153.
- [163] S. Hume, G. M. Greetham, P. M. Donaldson, M. Towrie, A. W. Parker, M. J. Baker and N. T. Hunt, *Anal. Chem.*, 2020, **92**, 3463–3469.
- [164] C. Lee and M. Cho, *The J. Phys. Chem. B*, 2004, **108**, 20397–20407.
- [165] J. P. Lomont, J. S. Ostrander, J.-J. Ho, M. K. Petti and M. T. Zanni, *The J. Phys. Chem. B*, 2017, **121**, 8935–8945.
- [166] Y. J. K. Edwards and S. J. Perkins, *Journal of Mol. Biol.*, 1996, **260**, 277–285.
- [167] C. R. Baiz, B. Błasiak, J. Bredenbeck, M. Cho, J.-H. Choi, S. A. Corcelli, A. G. Dijkstra, C.-J. Feng, S. Garrett-Roe, N.-H. Ge, M. W. D. Hanson-Heine, J. D. Hirst, T. L. C. Jansen, K. Kwac, K. J. Kubarych, C. H. Londergan, H. Maekawa, M. Reppert, S. Saito, S. Roy, J. L. Skinner, G. Stock, J. E. Straub, M. C. Thielges, K. Tominaga, A. Tokmakoff, H. Torii, L. Wang, L. J. Webb and M. T. Zanni, *Chem. Rev.*, 2020, **120**, 7152–7218.
- [168] D. Sharma and K. Rajarathnam, *J. Biomol. NMR*, 2000, **18**, 165–171.
- [169] M. Newville, T. Stensitzki, D. B. Allen and A. Ingargiola, *LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python*, Zenodo, 2014.
- [170] A. de Juan, J. Jaumot and R. Tauler, *Anal. Methods*, 2014, **6**, 4964–4976.

Appendix A

2D-IR Spectrometer operating procedures

Using the 2D-IR Spectrometer

A.1 Turning on the 2D-IR spectrometer

1. Please refer to the Solstice manual for turning on and operating the Spectra Physics Solstice Ace Femtosecond Laser, ensuring the interlock is operated and all barriers are in place.
2. Turn on the recirculating chiller and ensure the temperature is set to 20 °C. The pulse shaper must be cooled otherwise it overheats and shuts off.
3. Turn on the RF pulse shaper unit.

4. Turn on the focal plane array detector and fill with liquid nitrogen.
5. Start the 2D-IR Quick software and Hardware centre on the computer.

A.2 Turning off the 2D-IR spectrometer

1. Please refer to the Solstice manual for turning off the Spectra Physics Solstice Ace Femtosecond Laser.
2. Close the 2D-IR Quick software and Hardware centre on computer.
3. Turn off the RF pulse shaper unit.
4. Turn off the focal plane array detector
5. Turn off the recirculating chiller.
6. Turn off the red diode alignment laser.

A.3 Basic alignment of the 2D-IR spectrometer

1. Wear appropriate laser goggles (for 1,160 nm to 10,000 nm) and gloves
2. Close the interlock on the TOPAS OPA.
3. Remove the sample laser lid and turn on the red diode alignment laser
4. Place the alignment tool in the sample holder and block the pump diode laser beam

5. Adjust the position of sample until probe diode laser beam intensity is observed.
6. Adjust until the intensity is maximised.
7. Unblock the pump diode laser beam and block the probe diode laser beam
8. Adjust the IR pump mirror until intensity from the pump laser beam is observed through the alignment tool.
9. Adjust until the intensity is maximised.
10. Remove alignment tool and replace with sample.
11. Open the interlock on the TOPAS OPA.
12. Adjust the Z-position of sample until 2D-IR signal is observed on software.
13. Replace laser lids carefully to avoid them dropping into the beam path.

A.4 Laser realignment of input beam

1. Wear appropriate laser goggles (for 1,160 nm to 10,000 nm) and gloves
2. Close the interlock on the TOPAS OPA.
3. Remove laser box lid for input beam
4. Visually inspect beam path and ensure that the expected beam path is safe.
5. Place the flip mirror in the alignment position and place thermochromic paper in front of alignment mirror.

6. Open the interlock on the TOPAS OPA.
7. Turn on the red diode alignment laser.
8. Adjust Mirror 1 until the diode laser and TOPAS output is centred.
9. Move thermochromic paper 0.5 m away.
10. Adjust Mirror 2 until the diode laser and TOPAS output is centred.
11. Repeat steps 7-9 until the beams overlap at both distances.
12. Replace the laser box lid for input beam carefully.
13. Please see basic alignment section for next steps.

A.5 Calibration of 2D-IR pump beam

1. Close the interlock on the TOPAS OPA.
2. Remove the sample laser lid.
3. Flip up the pump detector redirect mirror.
4. Replace the sample laser lid.
5. Open the interlock on the TOPAS OPA.
6. Adjust angle of gratings until pump intensity is maximised (gratings 1 and 2 should not need to be unlinked).

7. Follow the directions on the calibration tab of the software located within the settings page.
8. Once finished with calibration of the pump beam close the interlock on the TOPAS OPA.
9. Remove the sample laser lid.
10. Flip down the pump detector redirect mirror.
11. Replace the sample laser lid.
12. Open the interlock on the TOPAS OPA.

A.6 Collection of 2D-IR spectra

1. Assuming the input beam alignment is good, close the interlock on the TOPAS OPA.
2. Remove the sample laser lid.
3. Place the sample in the sample holder.
4. Replace the sample lid and open the interlock on the TOPAS OPA.
5. Set the laser wavelength and spectrograph according to the sample and region of interest.
6. Set the pump mask to basic and ensure 'chop' is selected over 'phase cycling'.

7. Set the pump delay to ~ 2 ps ensuring the pump pulse is incident before the probe pulse.
8. Set the detector to continuous collection and ensure approximately 40 - 100 spectra are collected on average.
9. Ensure pump-probe signal is observed in the spectrum. If it is not, then adjust the time delay both forwards and backwards until it is observed. If it is still not observed, then perform the basic alignment again routine again.
10. Adjust the time delay of the pump pulse until the peaks observed are maximised in intensity. Then set this time delay to 0.
11. Change to ACQUIRE mode and load '2D-IR two phase cycling' default settings. These should be appropriate for most applications (the four phase cycling method can be used if scatter artefacts are observed on the spectrum in control mode).
12. Load masks to the pulse shaper and enter in a list of time delays.
13. Set the number of spectra to average, ~ 10 -50 is appropriate depending on time available and signal to noise ratio from the sample.
14. Finally click 'Collect spectra' to start acquisition of 2D-IR spectra.
15. Check the output periodically to ensure pump-probe signal is still observed, as well as the overall laser intensity to ensure the laser remains in alignment.
16. When the spectra have been acquired the 2D-IR spectrometer can be switched off as described above and then the sample may be removed after that.

Appendix B

Tryptophan-to-Heme Electron Transfer in Myoglobins

Tryptophan-to-Heme Electron Transfer in Myoglobins

Roberto Monni,¹ Raphael Horvath,² Xue-Zhong Sun,² David Tiemessen,² Yang Zhao,¹
Michael Towrie,³ Michael W. George^{2,4*} and Majed Chergui^{1*}

¹Laboratoire de Spectroscopie Ultrarapide and Lausanne Centre for Ultrafast Science, École Polytechnique
Fédérale de Lausanne, FSB- ISIC, CH-1015 Lausanne, Switzerland

²School of Chemistry, University of Nottingham, Nottingham NG7 2RD, United Kingdom

³Central Laser Facility, CCLRC Rutherford Appleton
Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, United Kingdom

⁴Department of Chemical and Environmental Engineering, University of Nottingham Ningbo China, 199
Taikang East Road, Ningbo, 315100, China

Abstract:

The recent discovery of electron transfer (ET) from photoexcited Tryptophan to the heme in ferric myoglobins (Mbs), competing with resonant energy transfer raises questions about the coupling between these two processes, the pathway for electron transfer as well as the role of the final electron acceptor (the heme). Indeed, the latter changes with the oxidation state of the iron ion (ferric or ferrous), its spin state (planar low spin ligated vs deoxy high spin unligated form), and the type of distal ligand (CO, NO, H₂O, CN) or lack thereof (deoxyMb). To address these questions, we carried out ultrafast UV pump/infrared (IR) probe spectroscopy of deoxyMb, MbCN, MbCO and MbNO. Compared to previous studies using visible-UV probes, IR probing offers the advantage of a chromophore-specific monitoring both the heme and the ligands, along with the protein scaffold via its amide bands. A band shows up at 1720 cm⁻¹ upon excitation of tryptophan, which we identify as a marker of the reduced heme based on theoretical calculations. This band shows up in all Mb's, regardless of the oxidation and spin state and of the ligand, except NO. Heme reduction leads to release of the CO and CN ligands, accompanied by charge redistribution to the porphyrin macrocycle, making it the prime electron acceptor, while some electron density is also present on the Fe ion. In the case of MbNO, no reduction of the porphyrin is detected and from the data, we conclude that the electron ends up on the NO itself. In all cases, the ligand dissociates from the reduced heme in typically 200-250 ps, which we believe is driven by protein fluctuations, and this leads to a redistribution of charge on the macrocycle. With the ability to site-directly mutate its amino-acids, this study shows that Myoglobin is an ideal model system to investigate electron versus energy transfer in biological systems and to quantify the efficiency of the various amino-acids in the electron pathway to the acceptor.

Introduction

Electron transfer (ET) and Excitation Energy Transfer (EET) processes are widespread in biological systems where they play a crucial role in several fundamental processes.¹⁻⁹ The photosynthetic system is a very representative example, as it undergoes both energy and charge transfer processes during its function. Photoabsorption excites the light harvesting complexes, which undergo EET to the reaction centre. This then drives ET between the cytochromes c2 and the reaction centre, and the bc1 complex. Proton transfer then ensues, which triggers the ATPase. While only the first events are triggered by photoabsorption, in the blue copper protein, involved in electron transport in plant photosynthesis, photoinduced charge transfer occurs upon excitation in the visible region.¹⁰⁻¹¹ ET is also crucial for the electron transport chain in aerobic cellular respiration,¹² photosynthesis,¹³ DNA repair by photolyase,⁷ as well as in driving the proton pump of cytochrome c.¹⁴ It is well established that, within a protein, ET can occur on the ps-ns timescale over long distances ($> 10 \text{ \AA}$).^{1-3, 15-17} In parallel, EET plays a major role in energy transport in the form of excitons along DNA strands,¹⁸⁻¹⁹ and it has become the basis of the Fluorescence resonance energy transfer (FRET) commonly used in biological imaging.²⁰⁻²¹

In this respect, the amino-acid residue Tryptophan, plays a crucial role. Its absorption and emission bands are extremely sensitive to the environment, which strongly affects its fluorescence energy and quantum yield within proteins.²²⁻²⁵ For example, the fluorescence decay time of photoexcited Trp (*Trp) in Cytochrome c,²⁶⁻²⁷ bacteriorhodopsin (bR)²⁸⁻³⁰ and myoglobins (Mb)^{24, 31-33}, is significantly shorter (sub-ps to hundreds of ps), than its life-time in water ($\sim 3 \text{ ns}$).²² The lifetime quenching of Trp in proteins has generally been rationalized in the terms of FRET from *Trp to acceptor molecules that could be the cofactor or other amino-acid residues.^{24-25, 34-36} FRET is based on the Förster dipole-dipole coupling and its rate depends on the inverse sixth power of the donor-acceptor (D-A) distance and on the relative orientations of the donor and acceptor dipoles, as well as the overlap integral of the donor fluorescence band with the acceptor absorption band. The high sensitivity of the Trp spectra to the environment, along with the inverse-sixth power dependence on the D-A distance of the FRET rate, have made it a commonly used natural fluorescing probe (the so-called “molecular ruler”) in investigations of protein dynamics.^{21, 37-38}

Tryptophan is also known to undergo ET reactions in biosystems, such as the blue copper proteins,⁴ the azurins,³⁹ relay in Re-containing proteins³⁹⁻⁴¹ and heme reduction and oxidation

has been reported for proteins such as Cytochrome c,⁴²⁻⁴⁴ while recent UV-visible TA studies point to a FRET from Trp to the heme.²⁷ This raises the question of the extent to which the Trp fluorescence decay in hemoproteins is caused by ET or FRET and how these two processes are coupled, if at all. This issue started to be addressed recently in the case of myoglobin (Mb).⁴⁵⁻⁴⁶ The latter contains two Trp residues: Trp7, which is solvent exposed and lies ~21 Å (centre-to-centre) away from the heme and Trp14, which is embedded within the protein with a centre-to-centre distance to the heme of ~15 Å (Figure S1).²⁴ Several time-resolved studies of their fluorescence in Mb's^{24, 32-33, 47} have shown that excited Trp7 decays in 105 to 130 ps, while excited Trp14 decays in 16 to 28 ps in all Mbs (table S1). These decays have been modelled in terms of dipole-dipole resonance energy transfer (i.e. FRET), though with some discrepancies for the case of Trp14.^{32-33, 48} These discrepancies were explained by two-dimensional (2D) deep-ultraviolet (UV) Transient Absorption (TA) and UV pump/visible probe TA spectroscopic studies of the ferric Cyano-Mb (MbCN) and met-Mb (MbH₂O) by Consani et al,⁴⁵ which showed that the *Trp14 decay rate is not only due to FRET to the heme, but it competes with an ET with a yield of ~50%, while the *Trp7 decay is entirely due to FRET to the heme. In the case of ferric MbCN and metMb, formation of a ferrous deoxyMb, resulting from the ET, was witnessed by its spectral fingerprints in the region of the Q- and Soret absorption bands.⁴⁵ This study was followed by another on the ferrous deoxy-Mb,⁴⁶ which showed that a *Trp14-heme ET also occurs at a comparable rate as in ferric Mbs, even though the heme is in a high spin (HS) domed state, which is very different from the low spin (LS) ferric planar ligated form. It was also suggested that the ET pathway follows a route via the Leucine 69 (Leu69) and Valine 68 (Val68) amino acid residues (Figure S1). Later, Besley and co-workers⁴⁹ confirmed the proposed ET pathway using quantum chemical calculations of the Trp-mediated FRET and ET in Mb by density function theory (DFT) and time-dependent density functional theory (TDDFT), and compared them with those using the pathway model by Beratan and co-workers.⁵⁰⁻⁵² These calculations rationalised the experimental results for ferric Mbs,⁴⁵ and confirmed the proposed ET pathway.⁴⁶

The invariance of the Trp-heme ET rate in ferric MbCN and metMb⁴⁵ and in ferrous deoxyMb,⁴⁶ which have very different hemes, raises questions about the detailed nature of the ET process and its pathways, and calls for an investigation of other Mb species such as the ferrous ligated carboxy-Mb (MbCO) and Nitrosyl-Mb (MbNO). Given that the ET rates measured in the TA studies⁴⁵⁻⁴⁶ match well the fluorescence decay of Trp14 for a yield of ~0.5, this implies that the Trp14 decay is determined by ET and FRET to the heme at almost equal yields. Table S1

suggests that these processes are similar in all Mb species. In the case of the ET in Mb's the question as to the fate of the transferred electron is still open. For deoxyMb, it was concluded that it is localized on the porphyrin ring,⁴⁶ while for ferric Mb's it was concluded that it is on the iron ion⁴⁵. This calls for further questions, is there a back-electron transfer, or is it lost to other parts of the protein? What happens in the case of ligated ferrous Mb's, and what is the fate of the diatomic ligand?

The absorption by the two Trp's overlaps that of the heme in the 260-300 nm region (Figure S2) and therefore in addition to the FRET and ET to the heme due to photoexcited Trp's, direct excitation of the heme also occurs, which induces intra-porphyrin relaxation processes, including the dissociation of the distal ligand in ferrous Mb's. Given that ligand recombination spans similar timescales as the Trp decay (especially for MbNO⁵³⁻⁵⁶), the task of disentangling the processes triggered by direct excitation of the porphyrin from the Trp-mediated FRET and the ET processes cannot be achieved by visible TA.⁴⁵⁻⁴⁶ In order to circumvent this limitation, here we use UV pump/infrared (IR) probe spectroscopy, as the latter can unambiguously identify the various chromophores (ligand, porphyrin, Trp, etc.) via their marker bands. In addition, the stretch modes of the distal ligand have been shown to be good sensors of the oxidation state of the heme,⁵⁷⁻⁶⁰ and of the intra-protein electrostatic fields as demonstrated for CO and NO.⁶¹⁻⁶²

In this paper, we report a study of *Trp14-to-heme ET in ferric MbCN and ferrous deoxyMb, MbCO and MbNO by IR TA spectroscopy. Since the *Trp14-to-heme ET in MbCN and deoxyMb was already characterized using UV and visible probes,⁴⁵⁻⁴⁶ we revisit them here in order to establish the IR signatures of the ET, which are crucial for describing it in MbCO and MbNO. The experiments were carried out exciting the system with femtosecond pulses at 315 nm (exciting the heme porphyrin only) and at 290 nm (exciting the heme and the tryptophans simultaneously). The 315 nm pump is close enough to the 290 nm one, such that effects due to excess energy are minimized. We scrutinize the response of the system in the region of the diatomic ligand stretch modes (1550-2150 cm⁻¹ for NO, CO and CN) and in the fingerprint region of the porphyrin and the Trp modes (1300-1800 cm⁻¹). The latter strongly overlap the amide I bands, which are dominant. The infrared continuum probe is obtained by difference frequency generation.⁶³ The experiment ran at a repetition rate of 10 kHz and the cross-correlation of the pump and probe pulses was 200 fs and the spectral resolution of the detection was 3 cm⁻¹.

We identify an IR marker of the ET at $\sim 1720\text{ cm}^{-1}$, in all the studied Mb's except MbNO and conclude that it is a marker of the reduced heme. However, the electron distribution is delocalised to various degrees on both the metal and the porphyrin macrocycle depending on the Mb. Further processes, such as ligand detachment from the heme, and redistribution of charge onto the macrocycle occur on time scales of 200-250 ps, which we attribute to large scale conformational changes of the scaffold. In the case of MbNO, we conclude that the electron is localised on the ligand. The loss of the reducing electron occurs on timescales $>1\text{ ns}$. We confirm the pathway for ET previously suggested⁴⁶ and supported by theory.⁴⁹ Given its versatility in site-directed mutation of its amino-acid residues,²³ the present studies show that Mb is an ideal model system to describe the role of amino-acid residues in directing and enhancing ET processes in proteins and how this is balanced by competing energy transfer processes. Furthermore, they also underscore the power of ultrafast IR spectroscopy in recovering details of the electron transfer process with a high degree insight.

Results:

As already mentioned, in order to disentangle the ET processes from direct heme excitation, we carried out our study at two pump wavelengths: 315 nm, which exclusively excites the heme, and 290 nm, which excites both the heme and the 2 Trp's. By comparing the extinction coefficients of the Trp with those of myoglobin at the excitation wavelength of 290 nm (Figure S2),⁴⁵ we estimate that heme photoexcitation accounts for $\sim 64\%$ of the excited species, while each Trp's contributes for $\sim 18\%$. At 315 nm, the absorption coefficient of the heme is $\sim 80\%$ that at 290 nm.

In order to grasp the complexity of processes taking place upon Trp excitation of the various Mb's, figures S3 to S6 show charts of the anticipated processes in the various Mb's upon excitation at 315 nm and 290 nm. The processes induced by direct heme excitation in the visible and ultraviolet have been well studied in the past and are reported in these figures. MbNO (Figure S6) is more complex than the other Mb's due to the fact that the photodissociation quantum yield is 55% and the ligand recombination occurs over two time-scales: a rapid one (5-10 ps) due to nearby ligands recombining within the heme pocket, while a longer recombination time (150-250 ps) is due to ligands that have migrated further away from the

heme pocket and are presumably located in the Xe4 pocket.^{54, 64} The relaxation of the non-dissociating MbNO back to the ground state has been assumed to be purely thermal.⁶⁴

At 290 nm, direct heme excitation will lead to processes similar to those generated under 315 nm excitation. Excitation of Trp7 generates by FRET processes similar to those by direct heme excitation but with a rise time of 105-125 ps, corresponding to the Trp7 fluorescence decay (Table S1). Excitation of Trp14 leads to both FRET and ET to the heme at comparable rates.⁴⁵⁻⁴⁶ This implies that upon 290 nm excitation, only ~10 % of species are affected by ET. We will discuss in more detail Figures S3 to S6 when we deal with the different systems separately but suffices it to say at this point that a complex interplay of direct and FRET excitation occurs simultaneous to ET, which render the disentangling of the latter somewhat complex. As a consequence, fits of kinetic traces are purely phenomenological and will be discussed only as indicative of the on-going processes.

Not shown in the charts of Figures S3 to S6 are the photoreduction and photooxidation processes that have been reported for met-myoglobin and Cytochrome c⁴² and MbNO,⁶⁵ but whose origins are still debated. Although both processes have small cross-sections, they are comparable, and therefore we can assume that they balance out each other.

Deoxy-Myoglobin (deoxyMb) and the marker band of heme reduction

Deoxy-Mb is a HS (S=2) five-coordinated domed Fe^{II}-heme, whose photocycle was reported to occur in < 4 ps upon 405 nm excitation (in addition to a weak 15 ps component).⁶⁶ The recovery of the photoexcited system was ascribed to vibrational cooling in the ground electronic state.⁶⁷ Figure 1 shows a selection of transient IR spectra of deoxy-Mb in the 1600-1850 cm⁻¹ region at different time delays for excitation at 315 nm (A) and 290 nm (B). Figure S7 shows similar data over an extended frequency (1380-1800 cm⁻¹) range. It exhibits a rather complex pattern of positive (absorption) and negative (depletion or bleach) features, due to the responses of Trp and of the amide I (1650 cm⁻¹ region) and II (1540 cm⁻¹ region) bands of the protein.⁶⁸ These features reflect dynamics associated to conformational⁶⁹ and electrostatic changes affecting the entire protein scaffold. Note that the amplitude of the signal is typically 3 to 5 times larger at 290 nm compared to 315 nm excitation, the intensity ratio of the amide I and II bands changes and the amide I bands shift to higher frequencies in the 290 nm case. The detailed analysis of these signals is beyond the scope of this paper and will be discussed in a later publication. We will however partly come back to them when we discuss the MbNO case. For now, we focus on the 1600-1850 cm⁻¹ region (Figure 1).

The transients in Figure 1 differ significantly at the two pump wavelengths, in the sense that stronger changes appear in the $< 1700 \text{ cm}^{-1}$ region upon 290 nm excitation (Figure 1B), and a broad excited state absorption (ESA) band, centred around 1720 cm^{-1} , appears that is missing under 315 nm excitation (Figure 1A). Its time trace is shown in Figure S8. It exhibits an increase over several tens of ps, followed by a slow decay up to our detection limit of 3 ns. The fit of the kinetic trace yields a rise of $20 \pm 2 \text{ ps}$ (Tables 1 and S2) and a component of $\geq 3 \text{ ns}$ is used to account for the slow decay. The UV pump/visible probe TA studies of deoxyMb reported the growth of a broad ESA feature in the 500-600 nm region, on a 20 ps timescale, which was attributed to an Fe(II)-porphyrin π -anion radical, that then remained stable up to the detection limit of 1 ns.⁴⁶ The straightforward assignment would be that the 1720 cm^{-1} band is the IR marker of the reduced ferrous heme, generated through *Trp14-to-heme ET. However, this assignment is far from trivial and requires a more detailed discussion.

Indeed, the most probable candidates for the band 1720 cm^{-1} band are the Trp cation, a reduced amino-acid residue or the reduced porphyrin, as suggested above. Regarding Trp, its vibrational spectrum spans the region up to 1650 cm^{-1} .⁶⁰ Beyond and up to 2700 cm^{-1} it does not contain vibrational bands.⁷⁰⁻⁷¹ Very few studies have been performed on the Trp radical and these are only resonance Raman ones.⁷²⁻⁷³ They do not show evidence for a band at 1720 cm^{-1} . The same goes for the Histidine amino-acid, which does not have bands above $\sim 1650 \text{ cm}^{-1}$. Last but not least, and as will be seen when we discuss the MbNO case, if this band were associated with the Trp radical or a reduced amino-acid residue, it should also show up in UV-excited MbNO, which is not the case.

Concerning the heme, the high frequency cut-off of the IR spectrum of heme porphyrins is near $1700\text{-}1710 \text{ cm}^{-1}$ and is attributed to the symmetric stretching (ν_{COOH}) of the protonated carboxylic-COOH group, while the COO- band is near 1565 cm^{-1} .⁷⁴⁻⁷⁵ One could conclude that upon heme reduction, the COOH band shifts to 1720 cm^{-1} and/or increases in intensity. However, IR spectroelectrochemical studies of Cytochrome c and deoxyMb do not show evidence for such a band even though the visible part of the spectrum clearly showed changes upon reduction.⁷⁵⁻⁷⁶ Since no ligated Mb's had been so far studied, we repeated the IR spectrochemical studies in the case of MbCO. These are presented in § S2 and they also show no evidence of the 1720 cm^{-1} band showing up at either positive or negative potentials. The absence of this band in the spectroelectrochemical studies may be due to an insufficient sensitivity, especially considering that it is quite broad. We therefore resorted to quantum chemical calculations. These are described in § S3 and here we just present the main

conclusions. The peak at 1720 cm^{-1} can be assigned to the CO stretch of COOH. It is not clearly visible in the steady-state FTIR spectrum⁷⁵ as the number of COO⁻ modes far outweighs that of the COOH modes at pH 7 and the tail of the amide I band obscures whatever small intensity is left. The 1720 cm^{-1} band in the transients could arise either from an increase in population of the COOH modes through protonation or an increase in oscillator strength. However, it is unlikely to be due to protonation as the kinetics for this peak occur on the same timescale as the ET. Our calculations show that the COOH mode for the anion has an ~80% increase in oscillator strength compared to the normal Heme, with a negligible shift in frequency. Therefore, and while further studies are needed to confirm this attribution, we conclude that the 1720 cm^{-1} peak is due to heme reduction and we identify it as a marker band of the porphyrin radical anion.

Cyanomyoglobin (MbCN)

MbCN has a LS ($S=1/2$)⁷⁷ ferric hexacoordinated planar heme. Its photo-cycle occurs in $< 4\text{ ps}$ under 420 nm excitation, and it was attributed to a combination of thermal heme relaxation and cascading among metal-centred (i.e. HS) states leading to loosening, without dissociation, of the Fe-CN bond.⁷⁸ Figure 2 compares transient spectra in the region of the CN stretch at selected time delays upon 315 nm (A) and 290 nm (B) excitation, and Figure 3 shows the $1600\text{-}1850\text{ cm}^{-1}$ range. In both figures, the transients clearly differ at the two pump wavelengths. In Figure 2A, the transient IR spectra show three main features: the Ground State Bleach (GSB) of the CN stretching frequency (ν_{CN}) centred at 2125 cm^{-1} and two ESA bands centred at 2118 cm^{-1} (marked **1**) and 2095 cm^{-1} (marked **2**), all decaying in a few ps, after which the system has fully recovered. These results are consistent with previous studies on ferric MbCN under 420 nm excitation,⁷⁸ where the ESA features were attributed to vibrationally hot CN stretch modes (**1**) and to a CN loosely bound to the iron (**2**). In these studies, the latter appeared promptly around 2093 cm^{-1} and it gradually shifted towards 2100 cm^{-1} while decaying in a few ps. The kinetic trace of the GSB band for 315 nm excitation is shown in Figure S9A. It shows a largest (negative) amplitude at $t\approx 0$, due to the prompt excitation, that recovers exponentially with a time constant of $3.1 \pm 0.4\text{ ps}$, in good agreement with ref. ⁷⁸.

Figure 2B shows the transient IR spectra upon 290 nm excitation. Here again, the GSB and feature **1** appear promptly, as in Figure 2A, while feature **2** is not clearly distinguished. Instead, an additional ESA band centred at $\nu_{\text{CN}} \approx 2070\text{ cm}^{-1}$ grows over several tens of ps and decays over hundreds of ps. Its frequency is different from that of free CN (2046 cm^{-1}), HCN (2094

cm⁻¹), or even CN⁻ (2080 cm⁻¹).⁵⁷ Furthermore, the IR bands of free CN and CN⁻ are weak and therefore, difficult to observe. In ferrous Mb^{II}CN, $\nu_{\text{CN}} = 2057 \text{ cm}^{-1}$ (i.e. red-shifted by 68 cm⁻¹ with respect to Mb^{III}CN) at pH=8 with a shoulder appearing at 2078 cm⁻¹ at pH=5.6.⁵⁷ Based on their quantum chemical calculations, Reddy *et al.*⁵⁷ attributed the lower C-N frequency in Mb^{II}CN relative to Mb^{III}CN to an additional electron density on all anti-bonding CN orbitals. Furthermore, these calculations showed that σ -donation weakens the Fe-C bond, which was used to explain the spontaneous dissociation of CN⁻ in Mb^{II}CN at -5 °C. In our visible and 2D UV TA studies,⁴⁵ we concluded that the Trp14-to-heme ET yields Mb^{II}CN, which appears clearly in the transients in the Q- and Soret-band regions. The ESA band at ~2070 cm⁻¹ in Figure 2B appears at a frequency, which is intermediate between the value in Mb^{III}CN and Mb^{II}CN but closer to the latter (~10 cm⁻¹ higher in frequency).⁷⁹ Furthermore, its intensity seems much higher than that of the parent bleach as while the latter has almost vanished by 20 ps, the 2070 cm⁻¹ band is still growing. As a matter of fact, the oscillator strength of the CN stretch in Mb(II)CN is approximately 100 times larger than in Mb(III)CN.⁷⁹ Putting the latter two observations together with those from the 2D UV studies, suggests that even though most of the transferred electron density is localized on the Fe ion, part of it is also distributed on the porphyrin. This is further confirmed by Figure 3B, which clearly shows the appearance of the ESA band at 1720 cm⁻¹ as in deoxyMb (Figure 1B). Under 315 nm excitation (Figure 3A), this band is completely missing. It is however remarkable that in fig. 3B, its intensity is comparable to that of deoxyMb, even though only a partial charge is distributed over the heme. We also note it increases in intensity from 20 to 100 ps time delay before slowly decaying over nanoseconds. We will discuss these observations below.

Figures S9B-D show the kinetic traces of the various GSB and ESA bands observed in Figures 2B and 3B, along with their fits using multi-exponential functions. Just as for Figure S9A, in Figure S9B, the GSB appears promptly and recovers within a few ps. The fit of the trace yields the same timescale as in Figure S9A (~3 ps). According to figure S4, this trace is dominated (~64%) by the direct heme excitation (same as in figure S9A) with its bleach recovery of ~3 ps, but it should also contain a small rising component (~18 %) of the bleach due to the reduction of the porphyrin by ET from Trp14 and FRET from both Trp's, which depletes the parent Mb^{III}CN species on a 20 ps and 120 ps time scales. These should appear as rising components to the 3 ps component, but because they are rate-limiting steps to the latter, they should show up as decay components. However, these kinetic components are not easy to extract from figure S9B due to the noise level and we prefer a phenomenological fit minimising the number of

components. We note however that at the longest time delay, there is still ~10-15% of the species that have not recovered and are due to reduced hemes, as we will see below. The kinetic trace of the 2070 cm^{-1} band (Figure S9C) exhibits a growth in 26 ± 4 ps and a decay in 243 ± 15 ps and ≥ 3 ns (Table 1), while the kinetic trace of the 1720 cm^{-1} band (Figure S9D) shows, according to the fit, a biexponential rise in ~ 25 ps and ~ 240 ps and a decay in ≥ 3 ns. The additional rise time of ~ 240 ps accounts for the further increase in intensity observed in figure 3B. For both the 2070 and 1720 cm^{-1} bands, the ~ 25 ps rise reflects the time scale of the Trp14 decay (Table S1), associated with ET. As far as the ~ 240 ps component is concerned, the fact that it appears as a rise of the 1720 cm^{-1} band, a decay for the 2070 cm^{-1} band, and does not show up in the recovery of the 2125 cm^{-1} band, suggests loss of CN^- from the reduced heme, with a concomitant redistribution of electron density from the Fe-CN moiety to the porphyrin macrocycle.⁵⁷ This is in line with the reaction cycle $\text{Mb(II)CN} \rightarrow \text{Mb(II)} + \text{CN}^-$ reported in ref. ⁵⁷, leaving a reduced deoxyMB and we provide here the time scale for its appearance. In other words, the ligand dissociates from the reduced heme leaving an ~ 10 -15% depletion of the initial population, while charge redistributes within the unligated reduced heme.

Interestingly, the timescale of the ligand release from the reduced heme (~ 240 ps) is identical to that of the large amplitude motion of the protein scaffold.^{54, 80-82} This motion may be triggered by the ET and/or by thermal fluctuations, but in any case, it drives the detachment of the ligand. That the ET triggers significant perturbations of the protein scaffold is witnessed by the strong response of the amide I and II bands upon 290 nm excitation (Figure S3).

Carboxymyoglobin (MbCO)

MbCO has a LS ($S=0$) ferrous hexacoordinated planar heme.⁷⁷ Previous ultrafast TA studies using visible and IR probes showed that upon visible excitation, MbCO undergoes prompt CO photo-detachment with unity quantum yield (QY),⁶⁴ and subsequent recombination on a timescale > 50 μs .⁸³ Time-resolved IR studies also reported ps dynamics of the CO ligand, which ends up in a docking site within the protein scaffold.⁸⁴ The latter were established by following the dynamics of the two transient bands at 2120 cm^{-1} and 2130 cm^{-1} due to the dissociated CO having different orientations inside the protein pocket.⁸⁴ Figure 4 compares the transient spectra in the region of the CO stretch frequency at selected time-delays for 315 nm (A) and 290 nm (B) excitation and Figure 5 shows the transients in the 1600-1850 cm^{-1} region. In Figure 4A, it can be seen that the CO stretch band (1942 cm^{-1}) is bleached promptly and remains almost at the same level up to the longest time delays. In addition, two weaker bands,

at 2120 cm^{-1} and 2132 cm^{-1} , grow over a timescale of 100s of ps in line with previous studies.⁸³⁻
⁸⁴ The CO ligand needs ms's to geminately recombine with the heme.⁸⁵ Because of our detection limit of 3 ns, no further evolution of the bleach signal is observed.

Upon 290 nm excitation (Figures 4B and 5B), two main differences show up compared to the 315 nm excitation: (i) the GSB band at 1943 cm^{-1} appears promptly but then grows further on the timescale of several tens of ps; (ii) an ESA band centred at 1915 cm^{-1} appears and grows on a similar timescale, to then decay over 100s of ps; (iii) the 1720 cm^{-1} band attributed to the reduced heme shows up.

Figure S10 shows the kinetic traces of the 1915 cm^{-1} (A), 1943 cm^{-1} (B), 2120 cm^{-1} (C), 2132 cm^{-1} (D) and 1720 cm^{-1} (E) bands upon 290 nm excitation. The extracted time constants and pre-exponential factors using a multiexponential fit function are given in tables 1 and S2. The 1943 cm^{-1} band appears promptly at $t=0$ (due to direct excitation of the heme) but it then undergoes a remarkable additional (~ 3 -fold) increase in amplitude on time scales of 17 ± 8 ps and 106 ± 50 ps. These times correspond to the decay times of Trp14 and Trp7, respectively. The kinetic trace at 1915 cm^{-1} band (A) grows in 17 ± 8 ps and decays in 240 ± 110 ps. The traces in (C) and (D) are quite similar to each other and show a slow growth on a timescale of 223 ± 36 ps. Finally, the trace of the 1720 cm^{-1} band (E) grows on a biexponential timescales of 20 ± 11 ps and 230 ± 70 ps.

The proximity of the 1915 cm^{-1} with the 1943 cm^{-1} bands, and the fact that they both show a growth on the timescale of the *Trp14 decay, leads us to associate the 1915 cm^{-1} band to CO ligands bound to the reduced heme, whose existence is supported by the appearance of the 1720 cm^{-1} band (Figure 5B). The fact that the docked CO bands at 2120 cm^{-1} and 2132 cm^{-1} (C, D) exhibit a rise time similar to the decay of the 1915 cm^{-1} band (240 ± 110 ps), suggests that this timescale is related to the loss of the CO ligand from the reduced heme. This is in line with the kinetic trace of the porphyrin anion band at 1720 cm^{-1} , which shows a bimodal growth on the timescale of 20 ± 11 ps due to the Trp14-to-heme ET, and in 230 ± 70 ps, suggesting that with the loss of CO from the reduced porphyrin, a redistribution of electron density occurs on the macrocycle. Therefore, the scenario for MbCO is much the same as for MbCN even though these two species have different oxidation states. The ET from Trp14 creates a new species with electron density being distributed between the Fe-ligand moiety and the porphyrin. The release of the ligand from the reduced heme is then accompanied by redistribution of the electron density leaving a reduced deoxyMb species. Just as for MbCN, we note that the timescale of

the ligand release (~ 230 ps) is quite close to that of the large amplitude motion of the protein scaffold.^{54, 80-82}

Nitrosylmyoglobin (MbNO)

MbNO has a LS ($S=1/2$) ferrous hexacoordinated planar heme.⁸⁶ As can be seen from figure S6, this system presents the most complex kinetics. Upon photoexcitation, NO ligand dissociation occurs with a QY of $\sim 50\%$.⁶⁴ Its recombination to the porphyrin Fe centre occurs on two typical timescales of 5-30 ps and 130-220 ps, with a weak >1 ns component.^{53, 55-56, 64, 87} The shortest component has been attributed to geminate recombination of NO molecules within the heme pocket, while the ~ 200 ps component was attributed to geminate recombination of NO ligands that migrated farther in the protein, presumably to the Xenon4 pocket.^{54, 81} Finally, the longest component is due to non-geminately recombining NO ligands.^{55, 82, 87} The relaxation pathways and time scales of the undissociated hemes (50%) is considered to be thermal.⁶⁴ Under 290 nm excitation, in addition to the above channels, further reduction by ET from *Trp14 and FRET from Trp7 will add to the depletion of MbNO on the time scale of ~ 20 ps and ~ 110 ps, further increasing the complexity of phototriggered processes. Furthermore, as the *Trp lifetimes (table S1) and the first two NO recombination times are comparable, Infrared probing becomes necessary in this case, as the ligand or heme response do not overlap that of *Trp or of an eventual reduced photoproduct resulting from the ET.

NO-containing ferrous hemoproteins have the parent band centred around 1610 cm^{-1} , while in ferric ones it shifts to 1927 cm^{-1} .⁸⁸⁻⁸⁹ Upon photolysis of ferrous MbNO, the dissociated NO form two species with frequencies near 1855 cm^{-1} ,⁸⁸ close to the gas phase frequency. Figure 6 shows transient spectra at different time delays of MbNO in the region of the NO stretch and of the porphyrin fingerprint, upon 315 nm (A) and 290 nm (B) excitation. The spectral region of the bound NO around 1610 cm^{-1} overlaps that of the amide I and II bands (Figure S7), which are sensitive to the protein conformation, temperature and intraprotein electric fields.⁹⁰⁻⁹² Figure 6A shows a prompt bleach over the entire range from 1570 to 1650 cm^{-1} , with a dominant peak at 1610 cm^{-1} , i.e. the NO stretch band in MbNO. The main band exhibits a red wing, which extends to $\sim 1580\text{ cm}^{-1}$, and a shoulder on the blue side. These results are quite similar to the transient signals reported by Kim *et al*⁵⁶ under 580 nm excitation, who attributed the main peak and its red wing to two distinct conformational states of NO in MbNO with frequencies at 1611 cm^{-1} and 1598 cm^{-1} , respectively. The shoulder on the blue side of the main peak in Figure 6A did not show up in their work but is quite prominent here. This shoulder may reflect a response

of the amide I band (Figure S7A) due to the higher excitation energy (3.936 eV) compared to Kim *et al*'s case (2.138 eV),⁵⁶ but this needs confirmation. The features in Figure 6A are overall reproduced in Figure 6B but some subtle differences show up. Both the red and blue shoulders around the main peak are more extended in frequency. For the red shoulder and especially around 1550 cm⁻¹, this is due to the significant increase of the amide II band under 290 nm excitation (Figure S3). The same holds for the amide I band, which in addition is blue shifted compared to the 315 nm excitation as can be seen from Figure S7. Finally, and strikingly different to deoxyMb, MbCN and MbCO, the 1720 cm⁻¹ band of the reduced heme does not show up under 290 nm excitation.

We note that the strongest signals due to the amide bands occur at the earliest times around 1550 cm⁻¹ and 1650 cm⁻¹ (Figure S7). Compared to the parent NO band (1611 cm⁻¹) at early times (Figure 6B), they represent an amplitude of 10-20%, further decreasing with time. We can therefore consider that the contribution of the amide bands is minor in this region and in particular at the frequency of the NO parent band, onto which we focus our attention hereafter. The kinetic behaviour of the latter is shown in Figure S11. It is clearly different for 315 nm and 290 nm excitation. The prompt rise of the bleach signal at t=0 is due to ultrafast dissociation of NO,^{56, 64} and is followed by a recovery over a few ps to hundreds of ps. After 3 ns, the signal for 315 nm excitation is nearly back to zero while this is not the case for 290 nm excitation, which exhibits a depletion of ~10% of the signal at t=0, close to what was found for MbCN. The fit of the trace for 315 nm excitation shows a recovery on three timescales: 8.1 ± 1.5 ps, 136 ± 13 ps and a minor long-lived component fixed at 1.7 ns (Table 1). These time constants are in agreement with those reported in the literature using different probes.^{56, 64 55, 64, 82, 93-94}

Fitting the kinetic trace for 290 nm excitation is more complex because of the several competing processes coming into play (Figure S6): a) direct excitation of the heme induces dissociation of NO with a yield of ~55 %, ⁶⁴ which recombines on timescales of ~8 ps and ~135 ps as under 315 nm excitation, leading to a recovery of the parent NO bleach band; b) on the other hand ET and FRET from the Trp's will add a further bleaching contribution rising in ~20 ps and ~120 ps; c) the reduced MbNO heme will also decay, but its time scale is unknown. Therefore, the kinetic trace for 290 nm excitation is fitted phenomenologically fixing the above three bleach recovery components (same timescales and relative pre-exponential factors) used to fit the 315 nm trace, and an additional long component of 10 ns to account for the signal not fully recovering within our observation window of 3 ns. Furthermore, additional bleach rise and recovery components were introduced, whose time constants and pre-exponential factors were

free parameters. The fit of the 290 nm kinetic trace reveals an additional component of 17 ± 2 ps in the growth of the depletion of the parent NO band and 350 ± 60 ps in its recovery. The former is close to the decay of Trp14 and suggests that the bleach of the NO parent band is in part caused by ET, but it does not yield a reduced heme as in all other Mbs, as reflected by the absence of the 1720 cm^{-1} band. For the longer component, we believe it is due to the partial recovery of MbNO from the ET process. We again stress that the competing processes under 290 nm excitation of MbNO are too complex for any mechanism to be deduced from the kinetics of a single band and we insist that the fit is purely phenomenological. Further studies would be needed to fully clarify the complex and competing kinetics governing *Trp14-mediated ET in MbNO.

The IR band frequencies and times scales of the various Mb's reported for 290 nm excitation are given in Table 1 along with their assignment, and are compared with those obtained in the optical domain for deoxyMb, metMb and MbCN.⁴⁵⁻⁴⁶ The results of the fits (time constants and pre-exponential factors) of all the kinetic traces herein presented are shown in Table S2.

Discussion

From previous time-resolved fluorescence studies^{24-25, 32-33, 35, 47, 95-98} of Trp in Mb complexes, the decay times (see Table S1), are rather invariant to the oxidation state of the heme Fe ion, the nature of the ligand complexing it and the origin of the Mb (Sperm Whale, Horse Heart, etc.). As already mentioned, Förster theory with, in some cases, molecular dynamics (MD) simulations, were used to interpret the decay times, which were attributed to FRET from the Trp's to the heme porphyrin. The heme has an absorption resonant (Figure S1) with the Trp fluorescence band in the 350 nm region. As a matter of fact, the Trp fluorescence decay times are much longer in apomyoglobin that does not contain a Heme (Table S1). Stevens et al.²⁴⁻²⁵ carried out detailed studies of FRET between Trp and the heme porphyrin in different Mb mutants, in order to develop a novel energy-transfer pair as a molecular ruler in hemoproteins. The calculated Trp lifetimes, assuming FRET to the porphyrin, yielded three components: one near 60 ps, due to Trp14 in the presence of normal hemes, one near 130 ps, due to both Trp14 in the presence of inverted hemes and Trp7 in the presence of normal hemes, and one near 1.8 ns, due to Trp7 in the presence of inverted hemes. Therefore, the calculated values agree well with the experimental ones for Trp7 but deviated by a factor of 2-3 for Trp14. This suggested that another decay process is affecting the relaxation of *Trp14, which was found to be ET to

the heme,⁴⁵⁻⁴⁶ accounting quite well for the above discrepancy between experimental and calculated rates.

The present studies aimed at characterizing the *Trp14-to-heme ET processes in Mbs and their dependence on the diatomic ligand, oxidation and spin states of the heme, using transient IR spectroscopy. In the 1600-1800 cm^{-1} probe region, a band centred appears at 1720 cm^{-1} that we attributed to the reduced heme as: a) it only shows up upon 290 nm photo-excitation of deoxy-Mb, MbCN and MbCO, with a risetime of ~ 20 ps that corresponds to the decay of the Trp14 fluorescence. This parallels the signatures of ET observed in the UV-visible TA of deoxyMb, MbCN and MetMb⁴⁵⁻⁴⁶ (Table 1); b) it does show up in our spectroelectrochemical spectra, nor in those of the literature,^{75-76,99} c) based on our calculations, we attribute it to the reduced heme. However, it may be registering a more complex interaction within the hemoprotein system. Indeed, the main difference between the heme reduction in the spectroelectrochemical studies and in the present case is that a Trp cation is created upon ET.

a) MbCN:

We note the following observations: a) the transient UV-visible studies identified a ferrous Mb^{II}CN, whose rise time is on the time scale of Trp14 decay and whose decay is >700 ps,⁴⁵ b) the 1720 cm^{-1} band grows on a similar time scale; c) the CN stretch frequency (2070 cm^{-1}) lies between the values for Mb^{III}CN (2125 cm^{-1}) and Mb^{II}CN (2057 cm^{-1}).⁵⁷ The lower value for the latter relative to the Mb^{III}CN was attributed to the appearance of additional electron density on the anti-bonding CN orbitals.⁷⁹ Thus the higher value in the present case may hint to somewhat less electron density on these orbitals. This leads us to conclude that while the electron density is predominantly on the Fe ion, some is also present on the porphyrin ring. Alternatively, the 2070 cm^{-1} may reflect the presence of a reduced heme that is generated in an excited electronic state. Surprisingly, as the 2070 cm^{-1} decays in ~ 240 ps and in nsec's, the 1720 cm^{-1} band grows correspondingly. This implies that the 2070 cm^{-1} decay reflects a dissociation of the CN ligands from the reduced heme, accompanied by a possible electron charge redistribution within the latter. The 240 ps dissociation time corresponds to that of protein motion, and is similar to the long timescale for NO recombination to the heme,⁵⁴ also attributed to large amplitude protein motions. Given that a nsec component is also present, we conclude that the ET triggers large amplitude motions that destabilize the reduced heme leading to CN⁻ dissociation.

b) MbCO

Upon 290 nm-excitation, the 1720 cm^{-1} band and an ESA band at 1915 cm^{-1} appear, red-shifted with respect to the parent CO ligand stretch (1943 cm^{-1}). Both grow on the time scale of the ^{14}Trp decay, while the 1915 cm^{-1} band decays in ~ 240 ps. This is somewhat similar to the observation in MbCN, despite the different oxidation states of the Fe ions in these two species. In Fe-carbonyl complexes, the reduction of the Fe atom by two units of charge leads to a ν_{CO} shift of ~ 240 cm^{-1} from $[\text{Fe}^{(0)}(\text{CO})_5]$ ($\nu_{\text{CO}} = 2029$ cm^{-1}) to $[\text{Fe}^{(2-)}(\text{CO})_4]$ (ν_{CO} is 1790 cm^{-1}).⁵⁸ Moreover, previous theoretical investigations⁵⁹ reported smaller ν_{CO} shifts of ~ 120 -140 cm^{-1} between $[\text{Fe}^{(2+)}(\text{CO})_5]^{2+}$ ($\nu_{\text{CO}} = 2151$ cm^{-1} , 2168 cm^{-1}) and $[\text{Fe}^{(0)}(\text{CO})_5]$ ($\nu_{\text{CO}} = 2012$ cm^{-1} , 2029 cm^{-1}). From these results, we can extrapolate a ν_{CO} red shift of ~ 120 -70 cm^{-1} for an $\text{Fe}^{\text{II}} \rightarrow \text{Fe}^{\text{I}}$ reduction, which is about 2 to 4 times larger than the observed ~ 30 cm^{-1} shift. While the latter reflects a more significant back-bonding from the metal, which would imply a stronger bonding of the ligand to the metal atom, the loss of the 1915 cm^{-1} band on a ~ 240 ps time scale points to a loose bond, and in addition to a charge redistribution to the porphyrin after ligand release, as discussed above in the case of MbCN. Indeed, we note that the ~ 240 ps timescale of CN- and CO-lysis, which is similar to the long timescale for NO recombination to the heme,⁵⁴ suggests that the protein scaffold undergoes conformational changes triggered by the ET, which detach the ligands bound to the reduced porphyrin.

c) MbNO

In this system, no extra ESA feature appeared under 290 nm excitation that could be related to the decay of $^*\text{Trp14}$ decay. In particular, the lack of the 1720 cm^{-1} band is remarkable. However, the bleached parent NO band shows different kinetic behaviours at the two pump wavelengths. In particular, for 290 nm excitation, the ~ 20 ps timescale typical of $^*\text{Trp14}$ -heme ET appears. Of all studied ligands, NO has the strongest dependence on the oxidation state of the metal, with an increase of its stretch frequency by ~ 320 cm^{-1} from $\text{Mb}^{\text{II}}\text{NO}$ to $\text{Mb}^{\text{III}}\text{NO}$.¹⁰⁰ Therefore, it is possible that upon $^*\text{Trp14}$ -to-heme ET in $\text{Mb}^{\text{II}}\text{NO}$, the NO stretch frequency may have shifted to lower frequencies out of our probe region. A hint that this is so comes from IR spectroelectrochemical studies of Fe(II)tetraphenylporphyrin, showing that the NO frequency shifts from 1681 cm^{-1} to 1496 cm^{-1} upon reduction.⁹⁹ However, the lack of the 1720 cm^{-1} band rules out heme reduction. We therefore suggest that the electron is most likely localized on the NO ligand itself, generating an $\text{Mb}^{\text{II}}\text{NO}^-$ heme. This is possible because the last antibonding NO π^* orbital contains a lone unpaired electron. This hypothesis is supported by previous cyclic voltammetry experiments showing that the reduction potential of the NO (-

0.63 V vs NHE) is smaller than that of the heme (-0.85 V vs NHE).¹⁰¹ Furthermore, as NO sits closest to the amino-acid (Val68) that is part of the ET pathway,^{46, 49} it acts as an electron trap.

A deeper analysis of the *Trp14 decay can be performed through Förster theory of electronic energy transfer (FRET)^{24-25, 34-35, 47} due to weak dipole-dipole coupling and Marcus theory of ET.^{16, 102} The tryptophan residues are located in an α -helix separated from the heme by the E helix, of which several amino acids: Valine 68 (Val68), Leucine 69 (Leu69), Threonine 70 (Thr70), Glycine (Gly74), Isoleucine (Ile75), and Leu76 lie within the direct path to the heme (Figure S1). As already mentioned, Suess et al.⁴⁹ carried out DFT and TDDFT calculations of FRET and ET in Mb, confirming the experimental results that the ET pathway competes with FRET in the case of Trp14 in Mbs.⁴⁵⁻⁴⁶ They also confirm the hypothesis, made by some of us,⁴⁶ that the ET is mediated by the Leu69 and Val68 residues. Inclusion of these residues is important for the TDDFT calculations. They also compared their calculations to the pathway tunnelling model by Beratan et al.^{50, 103} This model allows for an estimate of the diabatic electron coupling matrix element between the electronic states of the donor (D) and acceptor (A). It supposes that the ET between them is mediated by consecutive interactions between atoms connecting the D-A pair. The possible pathways are determined by a graph search algorithm¹⁰⁴ and their probability (so-called penalty) is determined, such that the coupling matrix element is the product of probabilities through each step. With this approach, it is possible to identify the strongest D-A ET pathways, estimate partial electronic couplings mediated by each pathway, calculate the weight of individual chromophore groups for mediating ET, and ultimately determine the dominant D-A ET pathways. We repeated the calculation of the FRET and ET rates following refs^{24-25, 49, 104-107}, with some differences with ref.⁴⁹ in the sense that we do not construct the location of the Hydrogen atoms from a molecular software but use the original PDB file. Suess et al also showed that the FRET and ET rates change by 20-30% using TDDFT on the S₁ state rather than the S₀ state of Trp.⁴⁹ Here, we took the S₀ state structure from the PDB. Our calculations using the pathways search algorithm¹⁰⁴ fully confirm the conclusions of ref.⁴⁹, providing good agreement for the Trp14 decay in ferric and ferrous Mb's.

Although the different Mb's investigated here have different hemes: a domed ferrous HS for deoxyMb, a planar ferrous LS for MbCO, a planar ferric LS heme for MbCN and a ferrous S=1/2 planar for MbNO, the rate at which they are reduced is identical, in agreement with the invariance of the Trp fluorescence decay times (Table S1). However, they are an essential player along the ET pathway Trp14-Leu69-Val68-heme,⁴⁹ as reflected in the long Trp14

fluorescence lifetime in apomyoglobin (Table S1). An extreme situation with no intermediate amino-acid residues between the sole Trp (Trp59) and the heme is the case of ferric and ferrous Cytochrome c (Cyt c), where the centre-to-centre distance is ~ 9 Å making the Trp-heme pair at almost van der Waals contact. The centre of mass of Trp is almost in the plane of the heme, and the indole plane forms an angle of ~ 70 - 80° with that of the heme. The latter does not favour FRET and one would expect a very efficient ET.¹⁰⁸ The measured Trp59 lifetime is indeed dramatically shortened (compared to Mbs) to ~ 350 fs in ferrous Cyt c and ~ 700 fs in ferric Cyt c.²⁶ Yet ultrafast TA UV-visible studies show no evidence of an ET to the heme,²⁷ or only a weak one, which has been attributed to thermal effects.⁴³ We conclude from this comparison that the entire chain (Trp14-Leu69-Val68-heme) is needed to sustain ET in Mbs, as they provide the electronic couplings between next neighbours, in line with the basic assumptions of the pathway model.⁵²

As already mentioned the decay of the reduced hemes occurs on timescales longer than our temporal window of 3 ns, and it may be due to back ET. In principle, the ET from Trp14 should lead to formation of a protonated Trp, but its detection as well as that of the native Trp is rendered difficult by their overlap with the amide I and II bands that dominate the spectrum. Even though the decay of the reduced hemes is longer than 3 ns, we find that in the cases of MbCN and MbCO, the ligand dissociates from the reduced porphyrin in 200-250 ps. A similar timescale was reported for the slow recombination of NO to the heme,^{54, 82} and it was attributed to large amplitude protein motions. In the present cases, these may be triggered by the ET process, as we clearly see (Figure S3) that the changes in the amide I and II bands are much more dramatic at 290 nm compared to 315 nm excitation.

Conclusions

We presented a time-resolved IR study of various Myoglobins in deuterated solution under physiological conditions (room temperature, pD = 7 solutions). In all cases (independent of the ligand and of the Fe oxidation or spin state) *Trp14 fluorescence quenching is due to both FRET and ET processes occurring at comparable yields.

Thanks to the sensitivity of the IR probe to the marker bands of the porphyrin, the ligand and the protein scaffold, we find that in all cases (except MbNO) the porphyrin macrocycle is the main acceptor of the electron. However, electron density is also present on the iron ion and, as expected, more so on the ferric than the ferrous ion. This loosens the bond of the ligands to the

Fe ion, which is witnessed by a new absorption band that is red-shifted with respect to parent diatomic ligand band in MbCN and MbCO. Thus, the ligand dissociates from the heme on time scales of 200-250 ps, which reflect large scale conformational changes of the scaffold. This is accompanied by a redistribution of charge to the porphyrin macrocycle. The loss of the reducing electron occurs on timescales > 1 ns. In the case of MbNO, from the kinetics of the parent NO band bleach, we conclude that the electron is localised on the NO itself, in line with the reduction potential of NO that is smaller than that of the heme and the fact that it sits closest to the Val68 amino-acid. Finally, the invariance of the ET rate across the different Mbs is due to the large driving force due to photoexcited Trp and from the comparison with apomyoglobin and cytochrome c, we conclude that the complete sequence is needed in order to allow the ET.

Given the ability to site-directly mutate its amino-acid residues,²³ the present studies show that Myoglobin is an ideal model system to determine the role of amino-acid residues in directing and enhancing ET processes in proteins and how this is balanced by competing energy transfer processes. Furthermore, the present work shows the power (and complementarity) of transient IR spectroscopy in unravelling new and insightful details of the electron transfer process that cannot be obtained by transient UV-visible⁴⁵⁻⁴⁶ or transient X-ray spectroscopy.¹⁰⁹

Finally, and as mentioned in the introduction, Trp-based FRET is widely used to determine distances in biomolecules and cells, however it can be unambiguous only if the decay of the Trp donor fluorescence is reflected by the rise of the acceptor population. This is seldom the case and the present findings highlight the importance of carefully assessing the process, as possible parallel ET processes can affect the results.

Acknowledgments:

We thank the NCCR:MUST (Molecular Ultrafast Science and Technology) of the Swiss NSF for its support. We also thank the University of Nottingham and EPSRC for funding. YZ thanks the China Postdoctoral Council for a grant within its International Postdoctoral Exchange Fellowship Program.

Table 1: Assignment and rise and decay times of bands appearing as a result of the Trp-to-Heme electron transfer in the various Myoglobins studies in this work (infrared bands) and in previous ultrafast UV-visible studies.

System	Infrared band			UV-visible bands ⁴⁵⁻⁴⁶			Assignments
	cm ⁻¹	τ_{rise} (ps)	τ_{decay} (ps)	nm	τ_{rise} (ps)	τ_{decay} (ns)	

metMb				530-600	20 ± 2	>0.7	deoxyMb
DeoxyMb	1720*	20 ± 2	>3 ns	500-600	20	>1	Fe^{II} -porphyrin [•]
MbCN	2068*	26 ± 4	243 ± 15 3.0 ± 0.1 ns	300-340 570, 530	19 ± 1	>2.5	CN@ Fe^{II} -porphyrin [•]
	1720*	26 ± 4 243 ± 15	3.0 ± 0.1 ns				Fe^{II} -porphyrin [•]
MbCO	1720*	20 ± 11 228 ± 70					Fe^{II} -porphyrin [•]
	1915*	17 ± 8	240 ± 110				CO@ Fe^{II} -porphyrin [•]
	1943§	17 ± 8 106 ± 47					CO in MbCO
	2120* 2132*	223 ± 36 223 ± 36					Docked CO Docked CO
MbNO	1611§	17 ± 2	350 ± 60				NO in MbNO

* positive absorption band; § bleach band.

Figure captions:

Figure 1: Transient spectra in the finger print region for Deoxy-Mb, upon 315 nm excitation (A) and upon 290 nm excitation (B) at selected time delays.

Figure 2: Transient spectra of MbCN in the CN stretching region, upon 315 nm excitation (A) and upon 290 nm excitation (B) at selected time delays.

Figure 3: Transient spectra in the fingerprint region of the porphyrin for MbCN upon 315 nm excitation (A) and 290 nm excitation (B).

Figure 4: Transient spectra at selected time delays of MbCO upon 315 nm (A) and 290 nm excitation (B). The ground state bleach of the CO band is centred at 1942 cm^{-1} , while the two bands of the free CO excited state absorption are centered at 2120 cm^{-1} and 2132 cm^{-1} .

Figure 5: Transient spectra in the fingerprint region for MbCO upon 315 nm excitation (A) and 290 nm excitation (B).

Figure 6: Comparison of selected transient spectra of MbNO upon 315 nm (A) and 290 nm (B) excitation

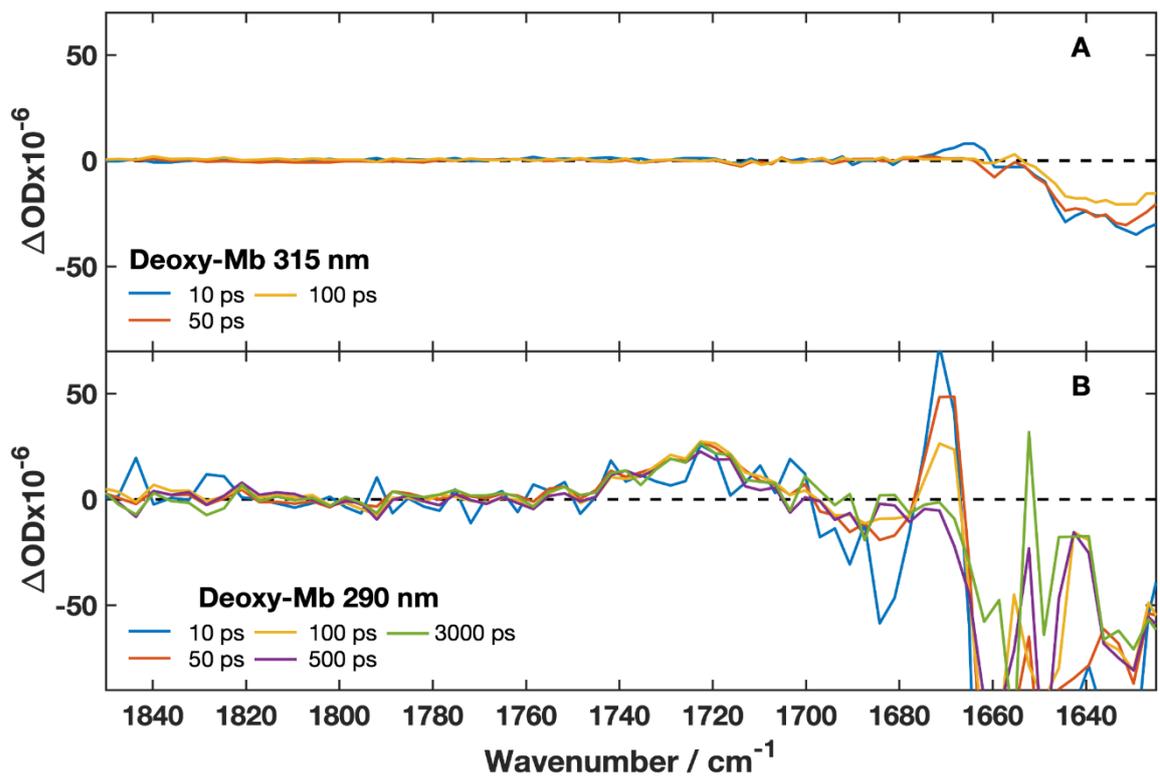


Figure 1

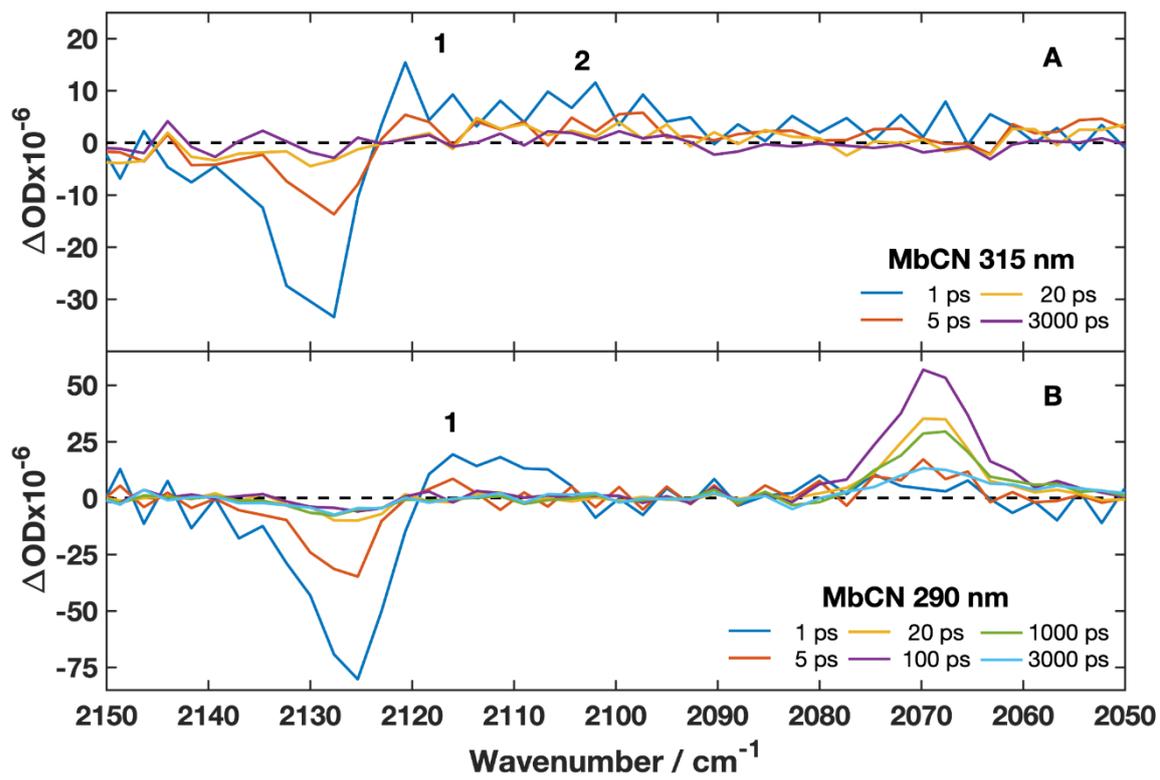


Figure 2

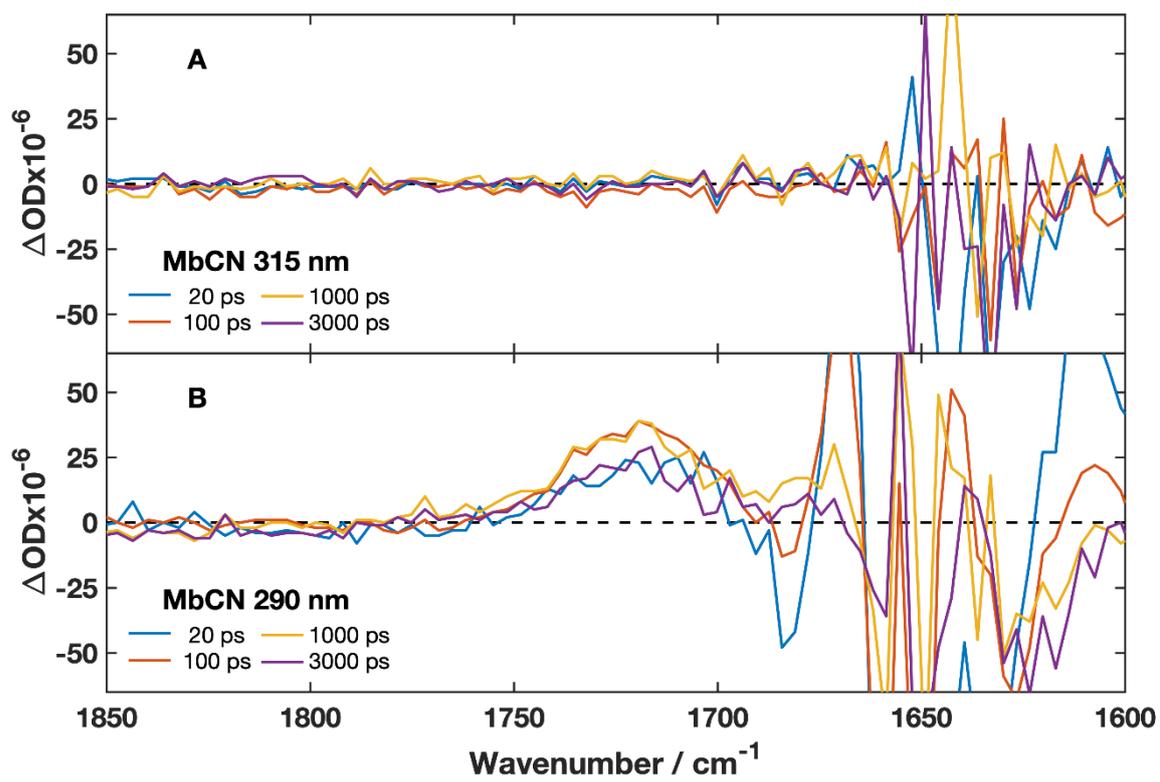


Figure 3

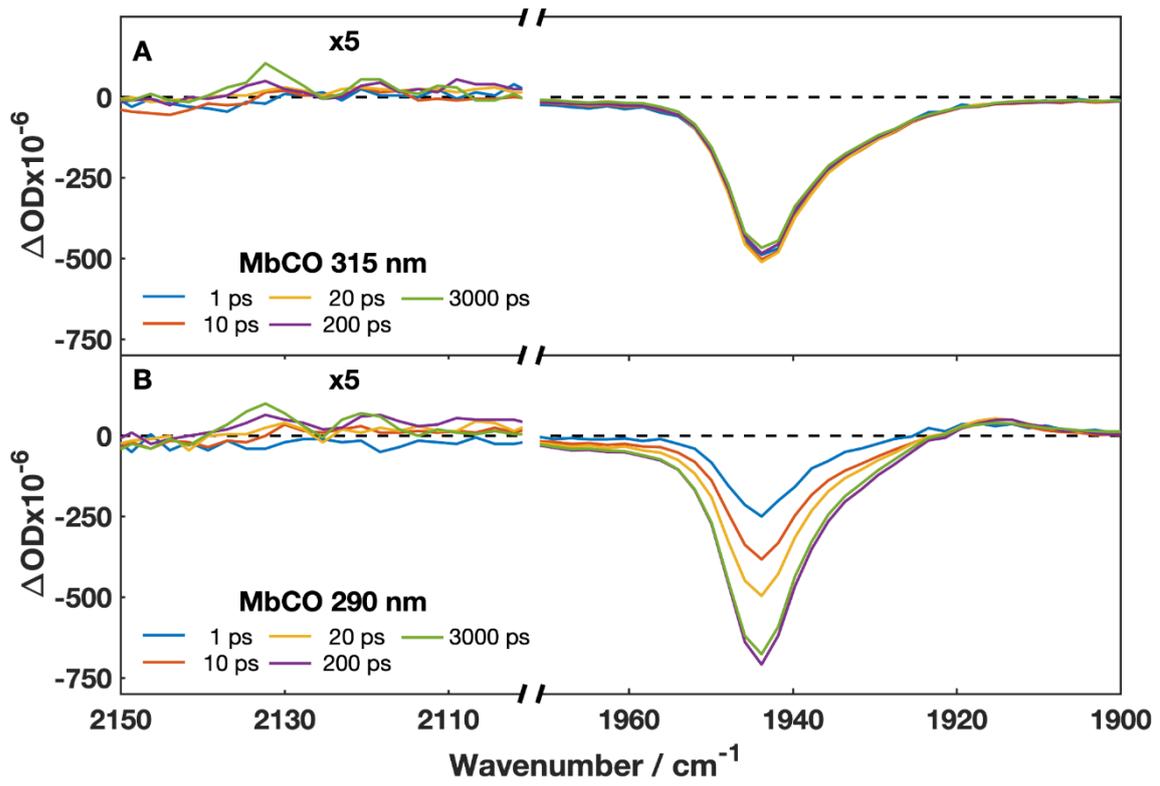


Figure 4

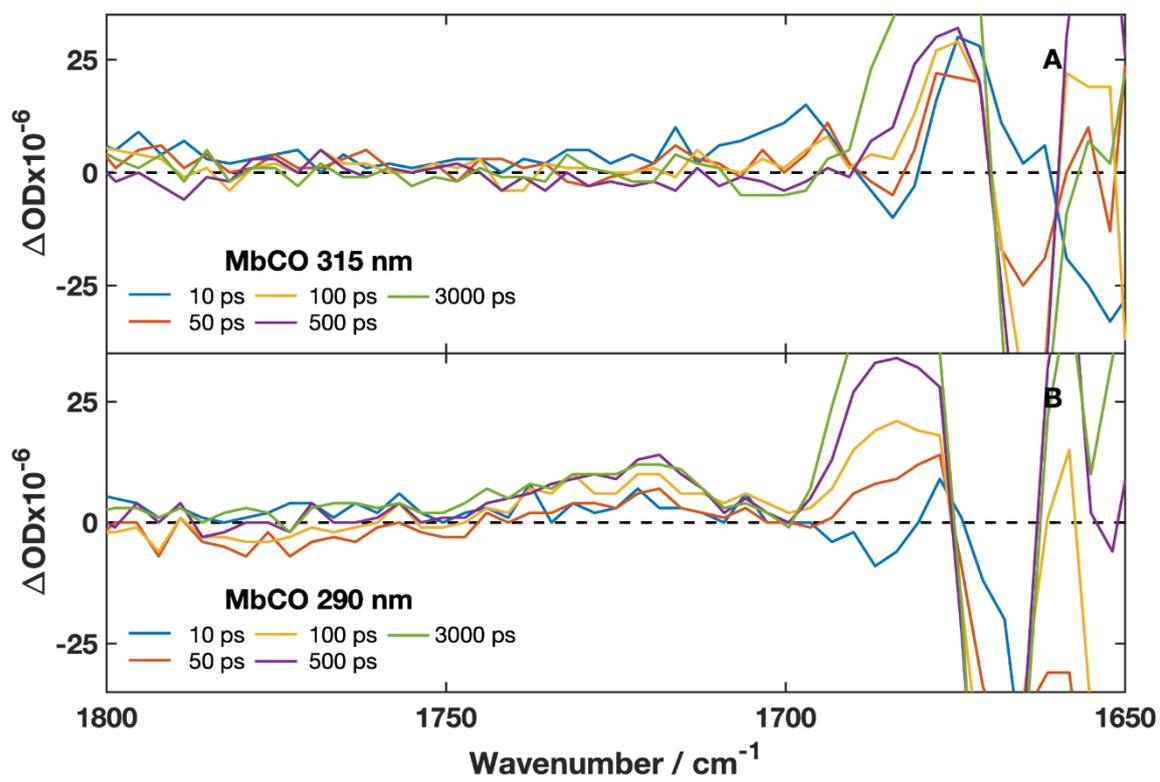


Figure 5.

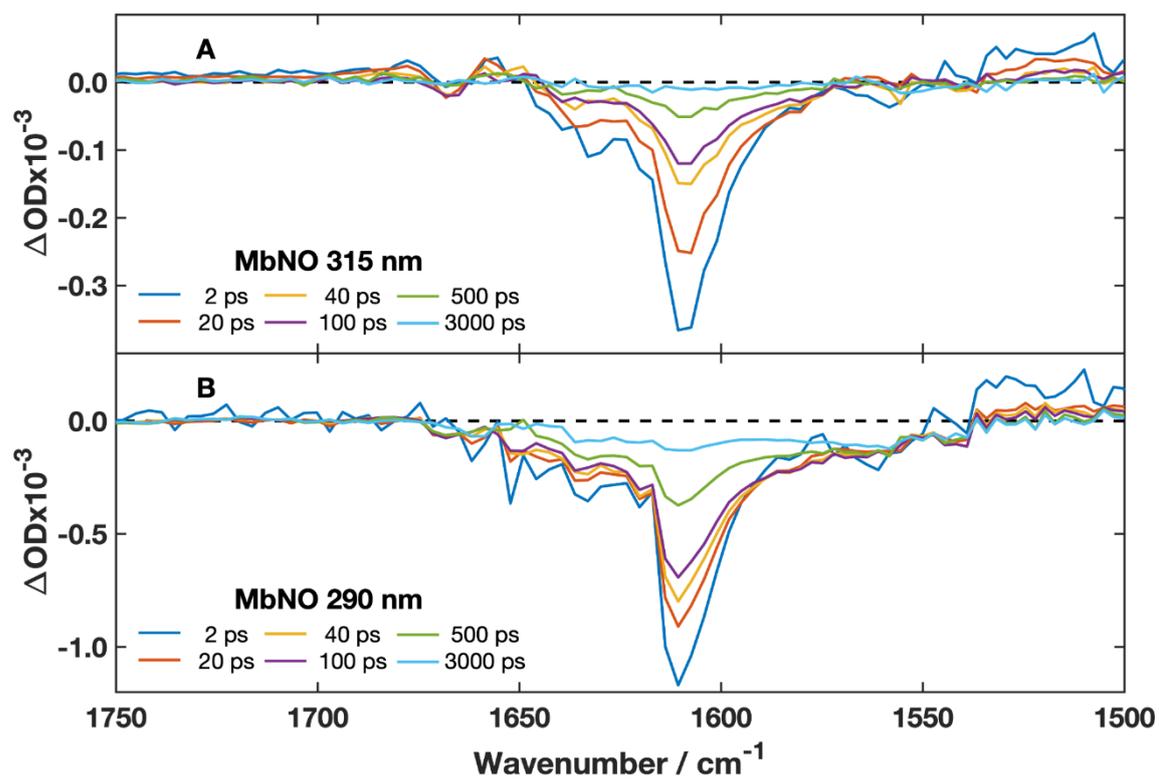


Figure 6.

Bibliography

1. Gray, H. B.; Winkler, J. R., Electron transfer in proteins. *Annu Rev Biochem* **1996**, *65*, 537-561.
2. Gray, H. B.; Winkler, J. R., Long-range electron transfer. *P Natl Acad Sci USA* **2005**, *102* (10), 3534-3539.
3. Cordes, M.; Giese, B., Electron transfer in peptides and proteins. *Chem Soc Rev* **2009**, *38* (4), 892-901.
4. Farver, O.; Pecht, I., Electron transfer in blue copper proteins. *Coordin Chem Rev* **2011**, *255* (7-8), 757-773.
5. Blumberger, J., Recent advances in the theory and molecular simulation of biological electron transfer reactions. *Chem Rev* **2015**, *115* (20), 11191-11238.
6. Giovani, B.; Byrdin, M.; Ahmad, M.; Brettel, K., Light-induced electron transfer in a cryptochrome blue-light photoreceptor. *Nat Struct Biol* **2003**, *10* (6), 489-490.
7. Byrdin, M.; Sartor, V.; Eker, A. P. M.; Vos, M. H.; Aubert, C.; Brettel, K.; Mathis, P., Intraprotein electron transfer and proton dynamics during photoactivation of DNA photolyase from E-coli: review and new insights from an "inverse" deuterium isotope effect. *Bba-Bioenergetics* **2004**, *1655* (1-3), 64-70.
8. Moser, C. C.; Page, C. C.; Dutton, P. L., Darwin at the molecular scale: selection and variance in electron tunnelling proteins including cytochrome c oxidase. *Philos T R Soc B* **2006**, *361* (1472), 1295-1305.
9. Chergui, M., Ultrafast photoinduced energy and charge transfer. *Faraday discussions* **2019**, *216*, 9-37.
10. Webb, M. A.; Kwong, C. M.; Loppnow, G. R., Excited-state charge-transfer dynamics of azurin, a blue copper protein, from resonance Raman intensities. *The Journal of Physical Chemistry B* **1997**, *101* (25), 5062-5069.
11. Fraga, E.; Loppnow, G., Proteins as solvents: blue copper proteins as a molecular ruler for solvent effects on resonance Raman intensities. *The Journal of Physical Chemistry B* **1998**, *102* (39), 7659-7665.
12. Brunori, M.; Giuffre, A.; Sarti, P.; Stubauer, G.; Wilson, M. T., Nitric oxide and cellular respiration. *Cell Mol Life Sci* **1999**, *56* (7-8), 549-557.
13. Rochaix, J. D., Regulation of photosynthetic electron transport. *Bba-Bioenergetics* **2011**, *1807* (3), 375-383.
14. Belevich, I.; Verkhovskiy, M. I.; Wikstrom, M., Proton-coupled electron transfer drives the proton pump of cytochrome c oxidase. *Nature* **2006**, *440* (7085), 829-832.
15. Therien, M. J.; Bowler, B. E.; Selman, M. A.; Gray, H. B.; Chang, I. J.; Winkler, J. R., Long-Range Electron-Transfer in Heme-Proteins - Porphyrin Ruthenium Electronic Couplings in 3 Ru(His)Cytochromes-C. *Adv Chem Ser* **1991**, (228), 191-199.
16. Siddarth, P.; Marcus, R. A., Electron-Transfer Reactions in Proteins - Electronic Coupling in Myoglobin. *Journal of Physical Chemistry* **1993**, *97* (23), 6111-6114.
17. Lancaster, K. M.; Farver, O.; Wherland, S.; Crane, E. J.; Richards, J. H.; Pecht, I.; Gray, H. B., Electron Transfer Reactivity of Type Zero Pseudomonas aeruginosa Azurin. *Journal of the American Chemical Society* **2011**, *133* (13), 4865-4873.
18. Middleton, C. T.; de La Harpe, K.; Su, C.; Law, Y. K.; Crespo-Hernandez, C. E.; Kohler, B., DNA Excited-State Dynamics: From Single Bases to the Double Helix. *Annu Rev Phys Chem* **2009**, *60*, 217-239.
19. Trifonov, A.; Buchvarov, I.; Wagenknecht, H. A.; Fiebig, T., Real-time observation of hydrogen bond-assisted electron transfer to a DNA base. *Chem Phys Lett* **2005**, *409* (4-6), 277-280.
20. Schuler, B.; Soranno, A.; Hofmann, H.; Nettels, D., Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins. *Annual review of biophysics* **2016**, *45*, 207-231.

21. Truong, K.; Ikura, M., The use of FRET imaging microscopy to detect protein-protein interactions and protein conformational changes in vivo. *Current Opinion in Structural Biology* **2001**, *11* (5), 573-578.
22. Lakowicz, J. R., *Principles of fluorescence spectroscopy*. 3rd ed.; Springer: New York, 2006; p xxvi, 954 p.
23. Qiu, W.; Li, T.; Zhang, L.; Yang, Y.; Kao, Y.-T.; Wang, L.; Zhong, D., Ultrafast quenching of tryptophan fluorescence in proteins: Interresidue and intrahelical electron transfer. *Chem Phys* **2008**, *350* (1-3), 154-164.
24. Stevens, J. A.; Link, J. J.; Kao, Y. T.; Zang, C.; Wang, L. J.; Zhong, D. P., Ultrafast Dynamics of Resonance Energy Transfer in Myoglobin: Probing Local Conformation Fluctuations. *J Phys Chem B* **2010**, *114* (3), 1498-1505.
25. Stevens, J. A.; Link, J. J.; Zang, C.; Wang, L. J.; Zhong, D. P., Ultrafast Dynamics of Nonequilibrium Resonance Energy Transfer and Probing Globular Protein Flexibility of Myoglobin. *Journal of Physical Chemistry A* **2012**, *116* (11), 2610-2619.
26. Bram, O.; Consani, C.; Cannizzo, A.; Chergui, M., Femtosecond UV Studies of the Electronic Relaxation Processes in Cytochrome c. *J Phys Chem B* **2011**, *115* (46), 13723-13730.
27. Consani, C.; Bram, O.; van Mourik, F.; Cannizzo, A.; Chergui, M., Energy transfer and relaxation mechanisms in Cytochrome c. *Chem Phys* **2012**, *396*, 108-115.
28. Roy, M.; Periasamy, N., Resolution of spectral and lifetime heterogeneity of tryptophan fluorescence in bacteriorhodopsin. *Photochem Photobiol* **1995**, *61* (3), 292-297.
29. Nomura, M.; Yamashita, M.; Kobayashi, S.; Sato, T.; Tokunaga, F., Picosecond UV Fluorescence Spectroscopy of Bacteriorhodopsin. *The Review of Laser Engineering* **1985**, *13* (11), 871-881.
30. Antipin, S.; Fel'dman, T.; Gostev, F.; Smitienko, O.; Sarkisov, O.; Ostrovsky, M. In *Femtosecond Dynamics of Intramolecular Energy Transition in Visual Pigment Rhodopsin*, Doklady Biochemistry and Biophysics, Springer: 2004; pp 125-127.
31. Chen, J. J.; Flaugh, S. L.; Callis, P. R.; King, J., Mechanism of the highly efficient quenching of tryptophan fluorescence in human gamma D-crystallin. *Biochemistry-U.S.* **2006**, *45* (38), 11552-11563.
32. Hochstrasser, R. M.; Negus, D. K., Picosecond Fluorescence Decay of Tryptophans in Myoglobin. *P Natl Acad Sci-Biol* **1984**, *81* (14), 4399-4403.
33. Willis, K. J.; Szabo, A. G.; Zuker, M.; Ridgeway, J. M.; Alpert, B., Fluorescence Decay Kinetics of the Tryptophyl Residues of Myoglobin - Effect of Heme Ligation and Evidence for Discrete Lifetime Components. *Biochemistry-U.S.* **1990**, *29* (22), 5270-5275.
34. Lakowicz, J. R., *Principles of Fluorescence Spectroscopy*, (1999). Kluwer Academic/Plenum Publishers, New York: 2004.
35. Gryczynski, Z.; Lubkowski, J.; Bucci, E., Intrinsic fluorescence of hemoglobins and myoglobins. *Fluorescence Spectroscopy* **1997**, *278*, 538-569.
36. Gryczynski, Z.; Olson, J. S.; Scott, E. S.; Bucci, E., Time resolved emissions of single tryptophan myoglobins reveal long range heme protein interactions. *Faseb J* **1998**, *12* (8), A1425-A1425.
37. Sahoo, H.; Roccatano, D.; Hennig, A.; Nau, W. M., A 10-angstrom spectroscopic ruler applied to short polyprolines. *Journal of the American Chemical Society* **2007**, *129* (31), 9762-9772.
38. Schuler, B.; Lipman, E. A.; Steinbach, P. J.; Kumke, M.; Eaton, W. A., Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. *P Natl Acad Sci USA* **2005**, *102* (8), 2754-2759.
39. Shih, C.; Museth, A. K.; Abrahamsson, M.; Blanco-Rodriguez, A. M.; Di Bilio, A. J.; Sudhamsu, J.; Crane, B. R.; Ronayne, K. L.; Towrie, M.; Vlček, A., Tryptophan-accelerated electron flow through proteins. *Science* **2008**, *320* (5884), 1760-1762.
40. Takematsu, K.; Pospíšil, P.; Pižl, M.; Towrie, M.; Heyda, J.; Záliš, S.; Kaiser, J. T.; Winkler, J. R.; Gray, H. B.; Vlček, A., Hole Hopping Across a Protein-Protein Interface. *The Journal of Physical Chemistry B* **2019**.
41. Takematsu, K.; Williamson, H. R.; Nikolovski, P.; Kaiser, J. T.; Sheng, Y.; Pospíšil, P.; Towrie, M.; Heyda, J.; Hollas, D.; Záliš, S.; Gray, H. B.; Vlček, A.; Winkler, J. R., Two Tryptophans Are Better Than One in Accelerating Electron Flow through a Protein. *ACS Central Science* **2019**, *5* (1), 192-200.

42. Gu, Y.; Li, P.; Sage, J. T.; Champion, P. M., Photoreduction of heme proteins: spectroscopic studies and cross-section measurements. *Journal of the American Chemical Society* **1993**, *115* (12), 4993-5004.
43. Löwenich, D.; Kleinermanns, K.; Karunakaran, V.; Kovalenko, S. A., Transient and stationary spectroscopy of cytochrome c: Ultrafast internal conversion controls photoreduction. *Photochem Photobiol* **2008**, *84* (1), 193-201.
44. Sun, Y.; Benabbas, A.; Zeng, W.; Kleingardner, J. G.; Bren, K. L.; Champion, P. M., Investigations of heme distortion, low-frequency vibrational excitations, and electron transfer in cytochrome c. *Proceedings of the National Academy of Sciences* **2014**, *111* (18), 6570-6575.
45. Consani, C.; Aubock, G.; van Mourik, F.; Chergui, M., Ultrafast Tryptophan-to-Haem Electron Transfer in Myoglobins revealed by UV 2D spectroscopy. *Science* **2013**, *339*, 1586-1589.
46. Monni, R.; Al Haddad, A.; van Mourik, F.; Auböck, G.; Chergui, M., Tryptophan-to-heme electron transfer in ferrous myoglobins. *Proceedings of the National Academy of Sciences* **2015**, *112* (18), 5602-5606.
47. Gryczynski, Z.; Bucci, E., Time resolved emissions in the picosecond range of single tryptophan recombinant myoglobins reveal the presence of long range heme protein interactions. *Biophys Chem* **1998**, *74* (3), 187-196.
48. Gryczynski, Z.; Lubkowski, J.; Bucci, E., Intrinsic fluorescence of hemoglobins and myoglobins. In *Methods in enzymology*, Elsevier: 1997; Vol. 278, pp 538-569.
49. Suess, C. J.; Hirst, J. D.; Besley, N. A., Quantum Chemical Calculations of Tryptophan -> Heme Electron and Excitation Energy Transfer Rates in Myoglobin. *J Comput Chem* **2017**, *38* (17), 1495-1502.
50. Beratan, D.; Betts, J.; Onuchic, J., Protein electron transfer rates set by the bridging secondary and tertiary structure. *Science* **1991**, *252* (5010), 1285-1288.
51. Beratan, D. N.; Onuchic, J. N.; Hopfield, J., Electron tunneling through covalent and noncovalent pathways in proteins. *The Journal of chemical physics* **1987**, *86* (8), 4488-4498.
52. Onuchic, J. N.; Beratan, D. N.; Winkler, J. R.; Gray, H. B., Pathway analysis of protein electron-transfer reactions. *Annu Rev Bioph Biom* **1992**, *21* (1), 349-377.
53. Petrich, J. W.; Lambry, J. C.; Kuczera, K.; Karplus, M.; Poyart, C.; Martin, J. L., Ligand-Binding and Protein Relaxation in Heme-Proteins - a Room-Temperature Analysis of No Geminate Recombination. *Biochemistry-US* **1991**, *30* (16), 3975-3987.
54. Ionascu, D.; Gruia, F.; Ye, X.; Yu, A. C.; Rosca, F.; Beck, C.; Demidov, A.; Olson, J. S.; Champion, P. M., Temperature-dependent studies of NO recombination to heme and heme proteins. *Journal of the American Chemical Society* **2005**, *127* (48), 16921-16934.
55. Kruglik, S. G.; Yoo, B. K.; Franzen, S.; Vos, M. H.; Martin, J. L.; Negreterie, M., Picosecond primary structural transition of the heme is retarded after nitric oxide binding to heme proteins. *P Natl Acad Sci USA* **2010**, *107* (31), 13678-13683.
56. Kim, S.; Jin, G.; Lim, M., Dynamics of geminate recombination of NO with myoglobin in aqueous solution probed by femtosecond mid-IR spectroscopy. *J Phys Chem B* **2004**, *108* (52), 20366-20375.
57. Reddy, K.; Yonetani, T.; Tsuneshige, A.; Chance, B.; Kushkuley, B.; Stavrov, S.; Vanderkooi, J., Infrared spectroscopy of the cyanide complex of iron (II) myoglobin and comparison with complexes of microperoxidase and hemoglobin. *Biochemistry-US* **1996**, *35* (17), 5562-5570.
58. Sathyanarayana, D. N., *Vibrational Spectroscopy: theory and applications*. New Age International Publisher: 2005.
59. Tsalavoutis, J. T.; Sigalas, M. P., Density Functional Investigation and Bonding Analysis of Pentacoordinated Iron Complexes with Mixed Cyano and Carbonyl Ligands. *J Comput Chem* **2010**, *31* (10), 1969-1978.
60. Barth, A., Infrared spectroscopy of proteins. *Bba-Bioenergetics* **2007**, *1767* (9), 1073-1101.
61. Phillips, G. N.; Teodoro, M. L.; Li, T.; Smith, B.; Olson, J. S., Bound CO is a molecular probe of electrostatic potential in the distal pocket of myoglobin. ACS Publications: 1999.
62. Park, E. S.; Boxer, S. G., Origins of the sensitivity of molecular vibrations to electric fields: Carbonyl and nitrosyl stretches in model compounds and proteins. *The Journal of Physical Chemistry B* **2002**, *106* (22), 5800-5806.

63. Greetham, G. M.; Donaldson, P. M.; Nation, C.; Sazanovich, I. V.; Clark, I. P.; Shaw, D. J.; Parker, A. W.; Towrie, M., A 100 kHz Time-Resolved Multiple-Probe Femtosecond to Second Infrared Absorption Spectrometer. *Appl Spectrosc* **2016**, *70* (4), 645-653.
64. Ye, X.; Demidov, A.; Champion, P. M., Measurements of the photodissociation quantum yields of MbNO and MbO(2) and the vibrational relaxation of the six-coordinate heme species. *Journal of the American Chemical Society* **2002**, *124* (20), 5914-5924.
65. Møller, J. K.; Skibsted, L. H., Nitric oxide and myoglobins. *Chem Rev* **2002**, *102* (4), 1167-1178.
66. Kholodenko, Y.; Volk, M.; Gooding, E.; Hochstrasser, R. M., Energy dissipation and relaxation processes in deoxy myoglobin after photoexcitation in the Soret region. *Chem Phys* **2000**, *259* (1), 71-87.
67. Ye, X. O.; Demidov, A.; Rosca, F.; Wang, W.; Kumar, A.; Ionascu, D.; Zhu, L. Y.; Barrick, D.; Wharton, D.; Champion, P. M., Investigations of heme protein absorption line shapes, vibrational relaxation, and resonance Raman scattering on ultrafast time scales. *Journal of Physical Chemistry A* **2003**, *107* (40), 8156-8165.
68. Kaiden, K.; Matsui, T.; Tanaka, S., A Study of the Amide-III Band by Ft-Ir Spectrometry of the Secondary Structure of Albumin, Myoglobin, and Gamma-Globulin. *Appl Spectrosc* **1987**, *41* (2), 180-184.
69. Kim, S.; Jin, G.; Lim, M., Structural dynamics of myoglobin probed by femtosecond infrared spectroscopy of the amide band. *B Kor Chem Soc* **2003**, *24* (10), 1470-1474.
70. Freire, P. T.; Barboza, F. M.; Lima, J. A.; Melo, F. E.; Mendes Filho, J., Raman Spectroscopy of Amino Acid Crystals. *Raman Spectroscopy and Applications* **2017**.
71. Takeuchi, H., Raman structural markers of tryptophan and histidine side chains in proteins. *Biopolymers: Original Research on Biomolecules* **2003**, *72* (5), 305-317.
72. Gurudas, U.; Schelvis, J. P. M., Resonance Raman Spectroscopy of the Neutral Radical Trp306 in DNA Photolyase. *Journal of the American Chemical Society* **2004**, *126* (40), 12788-12789.
73. Shafaat, H. S.; Leigh, B. S.; Tauber, M. J.; Kim, J. E., Resonance Raman characterization of a stable tryptophan radical in an azurin mutant. *The Journal of Physical Chemistry B* **2009**, *113* (1), 382-388.
74. Timasheff, S. N.; Rupley, J. A., Infrared titration of lysozyme carboxyls. *Arch Biochem Biophys* **1972**, *150* (1), 318-323.
75. Schlereth, D. D.; Maentele, W., Redox-induced conformational changes in myoglobin and hemoglobin: electrochemistry and ultraviolet-visible and Fourier transform infrared difference spectroscopy at surface-modified gold electrodes in an ultra-thin-layer spectroelectrochemical cell. *Biochemistry-Us* **1992**, *31* (33), 7494-7502.
76. MOSS, D.; NABEDRYK, E.; BRETON, J.; MÄNTELE, W., Redox-linked conformational changes in proteins detected by a combination of infrared spectroscopy and protein electrochemistry: Evaluation of the technique with cytochrome c. *European journal of biochemistry* **1990**, *187* (3), 565-572.
77. Harada, Y.; Taguchi, M.; Miyajima, Y.; Tokushima, T.; Horikawa, Y.; Chainani, A.; Shiro, Y.; Senba, Y.; Ohashi, H.; Fukuyama, H., Ligand energy controls the Heme-Fe valence in aqueous myoglobins. *J Phys Soc Jpn* **2009**, *78* (4), 044802-044802.
78. Helbing, J.; Bonacina, L.; Pietri, R.; Bredenbeck, J.; Hamm, P.; van Mourik, F.; Chaussard, F.; Gonzalez-Gonzalez, A.; Chergui, M.; Ramos-Alvarez, C.; Ruiz, C.; Lopez-Garriga, J., Time-resolved visible and infrared study of the cyano complexes of myoglobin and of hemoglobin I from *Lucina pectinata*. *Biophysical Journal* **2004**, *87* (3), 1881-91.
79. Reddy, K. S.; Yonetani, T.; Tsuneshige, A.; Chance, B.; Kushkuley, B.; Stavrov, S. S.; Vanderkooi, J. M., Infrared spectroscopy of the cyanide complex of Iron(II) myoglobin and comparison with complexes of microperoxidase and hemoglobin. *Biochemistry-Us* **1996**, *35* (17), 5562-5570.
80. Kitagawa, T.; Haruta, N.; Mizutani, Y., Time-resolved resonance Raman study on ultrafast structural relaxation and vibrational cooling of photodissociated carbonmonoxy myoglobin. *Biopolymers* **2002**, *67* (4-5), 207-213.

81. Kim, S.; Lim, M., Protein Conformation-Controlled Rebinding Barrier of NO and Its Binding Trajectories in Myoglobin and Hemoglobin at Room Temperature. *J Phys Chem B* **2012**, *116* (20), 5819-5830.
82. Silatani, M.; Lima, F. A.; Penfold, T. J.; Rittmann, J.; Reinhard, M. E.; Rittmann-Frank, H. M.; Borca, C.; Grolimund, D.; Milne, C. J.; Chergui, M., NO binding kinetics in myoglobin investigated by picosecond Fe K-edge absorption spectroscopy. *Proceedings of the National Academy of Sciences* **2015**, *112* (42), 12922-12927.
83. Jackson, T. A.; Lim, M.; Anfinrud, P. A., Complex Nonexponential Relaxation in Myoglobin after Photodissociation of Mbco - Measurement and Analysis from 2-Ps to 56-Mu-S. *Chem Phys* **1994**, *180* (2-3), 131-140.
84. Lim, M. H.; Jackson, T. A.; Anfinrud, P. A., Ultrafast rotation and trapping of carbon monoxide dissociated from myoglobin. *Nat Struct Biol* **1997**, *4* (3), 209-214.
85. Tian, W. D.; Sage, J. T.; Srajer, V.; Champion, P. M., Relaxation Dynamics of Myoglobin in Solution. *Physical Review Letters* **1992**, *68* (3), 408-411.
86. Morse, R. H.; Chan, S., Electron paramagnetic resonance studies of nitrosyl ferrous heme complexes. Determination of an equilibrium between two conformations. *J Biol Chem* **1980**, *255* (16), 7876-7882.
87. Kholodenko, Y.; Gooding, E. A.; Dou, Y.; Ikeda-Saito, M.; Hochstrasser, R. M., Heme protein dynamics revealed by geminate nitric oxide recombination in mutants of iron and cobalt myoglobin. *Biochemistry-U.S.* **1999**, *38* (18), 5918-5924.
88. Miller, L. M.; Pedraza, A. J.; Chance, M. R., Identification of conformational substates involved in nitric oxide binding to ferric and ferrous myoglobin through difference Fourier transform infrared spectroscopy (FTIR). *Biochemistry-U.S.* **1997**, *36* (40), 12199-12207.
89. Kim, J.; Park, J.; Lee, T.; Lim, M., Dynamics of Geminate Rebinding of NO with Cytochrome c in Aqueous Solution Using Femtosecond Vibrational Spectroscopy. *J Phys Chem B* **2012**, *116* (46), 13663-13671.
90. Dong, A. C.; Huang, P.; Caughey, W. S., Protein Secondary Structures in Water Determined from 2nd Derivative Amide-I Infrared-Spectra. *Biophysical Journal* **1990**, *57* (2), A438-A438.
91. Dousseau, F.; Pezolet, M., Determination of the Secondary Structure-Content of Proteins in Aqueous-Solutions from Their Amide-I and Amide-II Infrared Bands - Comparison between Classical and Partial Least-Squares Methods. *Biochemistry-U.S.* **1990**, *29* (37), 8771-8779.
92. Causgrove, T. P.; Dyer, R. B., Protein Response to Photodissociation of Co from Carbonmonoxymyoglobin Probed by Time-Resolved Infrared-Spectroscopy of the Amide-I Band. *Biochemistry-U.S.* **1993**, *32* (45), 11985-11991.
93. Shreve, A. P.; Franzen, S.; Simpson, M. C.; Dyer, R. B., Dependence of NO recombination dynamics in horse myoglobin on solution glycerol content. *The Journal of Physical Chemistry B* **1999**, *103* (37), 7969-7975.
94. Yoo, B.-K.; Kruglik, S. G.; Lamarre, I.; Martin, J.-L.; Negrerie, M., Absorption band III kinetics probe the picosecond heme iron motion triggered by nitric oxide binding to hemoglobin and myoglobin. *The Journal of Physical Chemistry B* **2012**, *116* (13), 4106-4114.
95. Willis, K. J.; Szabo, A. G.; Krajcarski, D. T., Fluorescence Decay Kinetics of the Tryptophyl Residues of Myoglobin Single-Crystals. *Journal of the American Chemical Society* **1991**, *113* (6), 2000-2002.
96. Beechem, J. M.; Brand, L., Time-resolved fluorescence of proteins. *Annu Rev Biochem* **1985**, *54* (1), 43-71.
97. Henry, E. R.; Hochstrasser, R. M., Molecular-Dynamics Simulations of Fluorescence Polarization of Tryptophans in Myoglobin. *P Natl Acad Sci USA* **1987**, *84* (17), 6142-6146.
98. Gryczynski, Z.; Fronticelli, C.; Tenenholz, T.; Bucci, E., Effect of Disordered Hemes on Energy-Transfer Rates between Tryptophans and Heme in Myoglobin. *Biophysical Journal* **1993**, *65* (5), 1951-1958.
99. Choi, I. K.; Liu, Y.; Feng, D.; Paeng, K. J.; Ryan, M. D., Electrochemical and spectroscopic studies of iron porphyrin nitrosyls and their reduction products. *Inorg Chem* **1991**, *30* (8), 1832-1839.

100. Nienhaus, K.; Palladino, P.; Nienhaus, G. U., Structural dynamics of myoglobin: FTIR-TDS study of NO migration and binding. *Biochemistry-Us* **2008**, *47* (3), 935-948.
101. Bayachou, M.; Lin, R.; Cho, W.; Farmer, P. J., Electrochemical reduction of NO by myoglobin in surfactant film: Characterization and reactivity of the nitroxyl (NO-) adduct. *Journal of the American Chemical Society* **1998**, *120* (38), 9888-9893.
102. Marcus, R. A., On the Theory of Oxidation-Reduction Reactions Involving Electron Transfer .1. *J Chem Phys* **1956**, *24* (5), 966-978.
103. Skourtis, S. S.; Beratan, D. N., High and low resolution theories of protein electron transfer. *J Biol Inorg Chem* **1997**, *2* (3), 378-386.
104. Balabin, I. A.; Hu, X.; Beratan, D. N., Exploring biological electron transfer pathway dynamics with the Pathways plugin for VMD. *J Comput Chem* **2012**, *33* (8), 906-910.
105. Gryczynski, Z.; Bucci, E.; Kusba, J., Linear Dichroism Study of Metalloporphyrin Transition Moments in View of Radiationless Interactions with Tryptophan in Hemoproteins. *Photochem Photobiol* **1993**, *58* (4), 492-498.
106. Gryczynski, Z.; Lubkowski, J.; Bucci, E., Heme-protein interactions in horse heart myoglobin at neutral pH and exposed to acid investigated by time-resolved fluorescence in the pico-to nanosecond time range. *J Biol Chem* **1995**, *270* (33), 19232-19237.
107. Skourtis, S. S.; Beratan, D. N., Electron transfer contact maps. *J Phys Chem B* **1997**, *101* (7), 1215-1234.
108. Winkler, J. R., FRETting over the Spectroscopic Ruler. *Science* **2013**, *339* (6127), 1530-1531.
109. Abela, R.; Beaud, P.; Van Bokhoven, J. A.; Chergui, M.; Feurer, T.; Haase, J.; Ingold, G.; Johnson, S. L.; Knopp, G.; Lemke, H., Perspective: Opportunities for ultrafast science at SwissFEL. *Structural Dynamics* **2017**, *4* (6), 061602.

Supplementary information

Tryptophan-to-Haem Electron versus Energy Transfer in Myoglobins

Roberto Monni,¹ Raphael Horvath,² Xue-Zhong Sun,² David Tiemessen,² Yang Zhao,¹
Michael Towrie,³ Michael W. George^{2,4*} and Majed Chergui^{1*}

¹Laboratoire de Spectroscopie Ultrarapide and Lausanne Centre for Ultrafast Science, École Polytechnique
Fédérale de Lausanne, FSB- ISIC, CH-1015 Lausanne, Switzerland

²School of Chemistry, University of Nottingham, Nottingham NG7 2RD, United Kingdom

³Central Laser Facility, CCLRC Rutherford Appleton

Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, United Kingdom

⁴Department of Chemical and Environmental Engineering, University of Nottingham Ningbo China, 199
Taikang East Road, Ningbo, 315100, China

S1. Experimental set-up:

The TRIR experiments were performed at the ULTRA and LIFETIME facilities at the Rutherford Appleton Laboratory, which has been discussed in detail elsewhere.¹⁻² Therefore, only a brief summary of the experimental setup is given here. A Ti:sapphire laser amplifier (Thales Laser) is used to produce 800 nm laser pulses (0.8 mJ, 10 kHz with a pulsewidth of 50 fs). This laser output is split and one portion is used to tune the pump beam to 315 or 290 nm by optical parametric amplification (OPA), the pump energy was then set to ~0.18 μ J/pulse at the sample using a neutral density filter. The second portion is sent through an optical parametric amplifier (Light Conversion) and a difference frequency generator, producing a tuneable mid-IR probe. The diameter of the pump and probe beams were approximately 150 and 80 μ m, respectively and both polarizations were set at the magic angle. After transmission through the sample, the IR probe is dispersed onto two linear 128 element MCT detector arrays (Infrared Associates). By the use of a chopper, the pump-on and pump-off infrared intensities can be measured and difference spectra generated. A small portion of the IR probe beam is taken before the sample and dispersed onto a 64 element MCT detector array (Infrared Associates) to provide a reference spectrum which is subtracted to account for beam intensity

fluctuations, the resultant spectra are also baselined. The experiments were performed with a recirculating flow system consisting of a peristaltic pump, Teflon tubing, and an infrared solution cell (Harrick Corp.) with CaF₂ windows typically at a path length of 50 μm (except in the case of MbCN where a pathlength of 25 μm was used). The cell was rastered in the plane perpendicular to the pump and probe beams and the sample was continually flowed through the cell to avoid degradation of the sample.

Lyophilized horse heart met-Mb and sodium dithionite (Na₂S₂O₄), were purchased from Sigma Aldrich and were used without further purification. The myoglobin (Mb) samples were prepared by dissolving lyophilised horse heart Mb in a deuterated phosphate buffer solution (D₂O, pD 7) at a concentration of ~15 mM while under a nitrogen atmosphere. For Deoxy-Mb, ~20 mg of sodium dithionate was dissolved in the Mb solution resulting in a concentration of ~3 mM for Deoxy-Mb. For MbNO, ~30 mg of sodium dithionite and ~5 mg of sodium nitrite were dissolved in the Mb solution resulting in a concentration of ~3. mM for MbNO. For MbCO, ~0.7 bars of CO gas was added to the flow cell containing the Mb solution, this resulted in a concentration of ~3 mM for MbCO. For MbCN in the CN stretching region, the myoglobin sample was prepared by dissolving Mb in a phosphate buffer solution (H₂O, pH7) at a concentration of ~15 mM and ~15 mg of NaCN were added in order to obtain an MbCN concentration of ~9 mM. For MbCN in the fingerprint region, the sample was prepared as above except in a deuterated buffer solution and at a concentration of ~3 mM.

S2. Spectroelectrochemical measurements:

Sample preparation for Nafion-phosphate film ATR-FTIR spectroscopy

The cell and sample preparation has been described extensively before but it is briefly described below. 2 uL of an ~8% Nafion-phosphate dispersion was deposited onto the silicon ATR prism, an FTIR spectrum was then obtained which acted as a background spectrum. An MbCO sample was prepared as stated above in D₂O. In an anaerobic glove box and a vial flushed with CO, 20 uL of MbCO was mixed with 20 uL of ~8% Nafion-phosphate dispersion. The vial was then left to stand in the dark for 30 minutes and then 2 uL of the mixture was deposited on the silicon ATR prism and left to dry for 5 minutes, resulting in a partially hydrated polymer electrolyte film. A piece of carbon paper was placed on top of the film which provides electrochemical contact across the film of particles. The gas tight ATR cell containing both the counter (Pt wire) and pseudo reference electrode (Ag wire) in an aqueous buffered solution) was then assembled above. FTIR spectra were then obtained at 4cm⁻¹ resolution using a Bruker Vertex 70 spectrometer with an MCT detector at varying potentials.

OTTLE Cell

The setup of an OTTLE cell has been detailed elsewhere. An MbCO sample was prepared as stated above in D₂O. The OTTLE cell was flushed with argon, then the MbCO solution was injected into the OTTLE cell. A cyclic voltammogram (CV) was then performed at a rate of 5 mVs⁻¹. Firstly in the range of -1.1 – 0.7 V and then from -1.5 – 0.7 V to ensure no reduction process was missed.

Figure S12 shows FTIR spectra of MbCO, specifically, the CO band, at different potentials. The CO band is initially located at 1944 cm⁻¹, when the potential is reduced to -400 mV, there is no change in the CO band. At -1000 mV there is an increase in the CO band intensity, however this is probably due to a global overall increase in intensity of the sample with time and is unlikely due to the potential being applied. Upon increasing the potential to +300 mV, a decrease in the CO band at 1944 cm⁻¹ is observed. But there is also an appearance of a peak at 1966 cm⁻¹. This is due to a different orientation of the CO binding. Increasing the potential to +700 mV we see an overall decrease in both of these bands. This all suggests CO loss occurs upon oxidation. The oxidation steps shows the electrochemical setup is working. However, because there is no major shift (or a decrease in intensity) in the CO peak at the reduction potentials, it implies that no reduction of the myoglobin is occurring.

The lack of reduction is further confirmed by the cyclic voltammogram (CV) of MbCO in an OTTLE which is shown in Figure S13. No peak is observed when attempting to reduce the sample. This suggests that MbCO is unable to be reduced using this method. By going all the way to -1.5 V it is believed deuterium evolution starts to occur.

S3. Calculations:

DFT calculations were performed on the deoxy heme structure for both the neutral and anion states, achieving the latter by reducing the overall charge of the complex by one. The initial coordinates of the heme were extracted from a crystal structure of myoglobin (PDB: 1YMB). The location of hydrogen atoms were added using the software IQMol, along with the diatomic ligands. Full geometry optimisations and vibrational frequency calculations were performed with Q-Chem 5.1 using the 6-31G* basis set and B3LYP functional. [cite Q-chem] The optimised structures were validated to be minimum energy states by normal mode analysis. All vibrational frequencies were scaled by a factor of 0.96.

Figure S14 shows the vibrational frequencies and intensities for both neutral and reduced states of the Deoxy heme. There are significant changes between the two hemes, however, many of these changes are either at low frequency or in a region of the spectrum that is obscured experimentally by protein vibrational modes like the amide I and II bands. There is one large

difference that is present at ~ 1750 cm^{-1} which is assigned to the CO stretch of the COOH propionate groups. This is in a region in the experimental FTIR spectrum which is absent of any large absorbance and so is much more observable. The anion oscillator strength of this vibration increases by $\sim 80\%$ when compared with the normal heme, with a very small change in the frequency. Only very minor changes in the intensity and frequency of the other COOH group is observed. In a difference spectrum between the normal and anion heme, the difference in oscillator strength manifests itself as a small positive peak centred at ~ 1750 cm^{-1} . Considering these calculations have been performed without a solvent field and are in vacuum, it is therefore reasonable to identify the 1720 cm^{-1} peak observed in the transient IR spectra as a marker band for the porphyrin radical anion.

The structures of the normal and anion form of deoxy heme are shown in Figure S15. It suggests the main reason for the increase in oscillator strength is the reorientation of the COOH group with respect to the porphyrin.

Table S1 : Fluorescence decay times (in picoseconds) of Trp⁷ and Trp¹⁴ in several SW Mb complexes (except when indicated). References are in brackets.

Myoglobin complex	Trp7	Trp14	Reference
DeoxyMb	105 \pm 0.3	18 \pm 0.6	6
metMb	135	16	7
	112.5 \pm 0.3	21.5 \pm 0.5	6
MbCO	132	26	7
	125.4 \pm 0.4	23.4 \pm 0.7	6
MbCN	113.2 \pm 0.4	28.5 \pm 0.8	6
MbN ₃	109.4 \pm 0.5	27 \pm 0.8	6
MbO ₂	122 \pm 0.5	24.4 \pm 0.9	6
HH apoMb	2840	2060	8

Table S2 : Fit parameters (time constants and pre-exponential factors) of the kinetic traces shown in Figures S4-S7

Sample	Excitation (nm)	IR band (cm ⁻¹)	Lifetimes (ps)				Pre-exponential factors ($\times 10^{-5}$)			
			τ_1	τ_2	τ_3	τ_4	a_1	a_2	a_3	a_4
MbDeoxy	290	1720	20 \pm 2	10000 (fixed)				-4.0 \pm 0.5	4.19 \pm 0.3	
MbCN	315	2125	3.1 \pm 0.4				-1.9 \pm 0.3			
		290	3.1 \pm 0.4				-5.4 \pm 0.3			
	2070	26 \pm 4	243 \pm 15	3000 \pm 147		-6.8 \pm 0.2	2.7 \pm 0.3	5.1 \pm 0.5		
	1720	26 \pm 4	243 \pm 15	3000 \pm 147		-1.9 \pm 0.2	-1.1 \pm 0.3	3.9 \pm 0.5		
MbCO	290	1915	17 \pm 8	238 \pm 112		-3.2 \pm 1.6		2.6 \pm 0.4		
		1943	17 \pm 8	106 \pm 47		30.9 \pm 5.1	7.7 \pm 7.35			
		2120	223 \pm 36				1.2 \pm 0.9			
		2135	223 \pm 36				1.2 \pm 0.9			
		1720	20 \pm 11	228 \pm 70		-0.6 \pm 0.2	-0.9 \pm 0.1			
MbNO	315	1611	8.1 \pm 1.5	136 \pm 13	1700 (fixed)	-	-	-4.2 \pm 1.0		
	290	1611	8.1, 136, 1700 (all fixed)	17 \pm 2	350 \pm 5 6	10000 (fixed)	4.1 \pm 0.1 * (-10.1, -12.9, 4.2, all fixed)	31.4 \pm 2.6	-	- 18.5 \pm 2.6 12.9 \pm 1.5

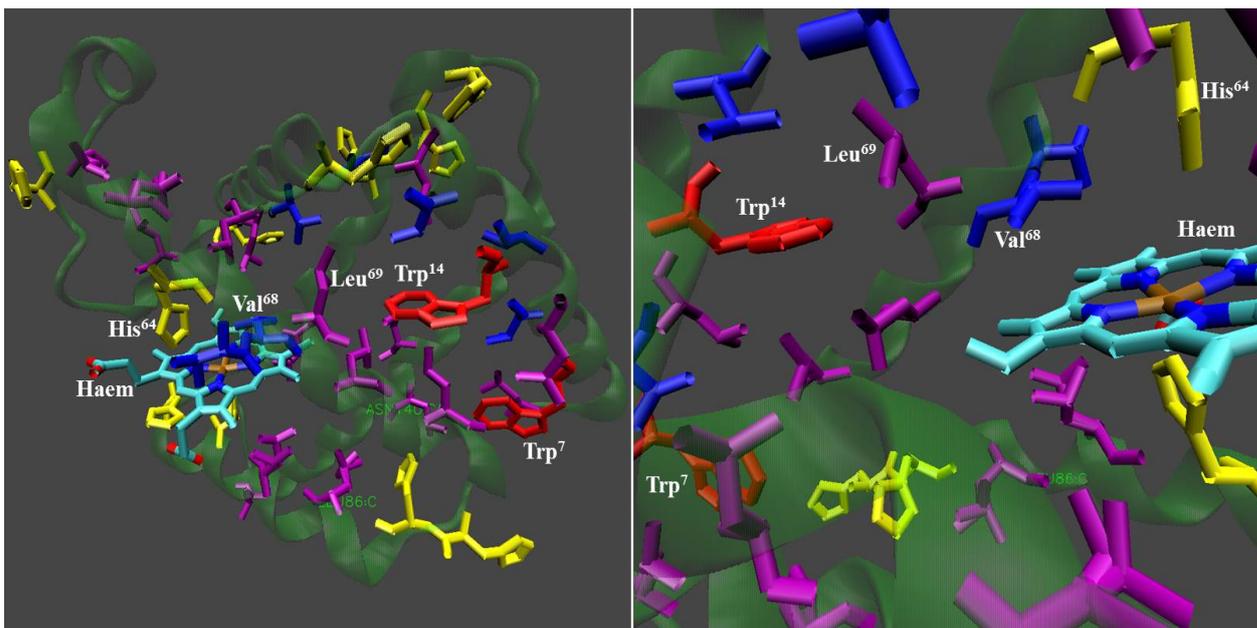


Figure S1: (A) Structure of myoglobin. The backbone is in green, the haem is highlighted by using different colours for each atom (namely C, O, N and Fe) and the most important residues for the electron transfer are labeled (Trp in red, Val in blue, Leu in violet and His in yellow). (B) Zoom of the region most relevant for electron transfer showing the residues between Trp and Haem.

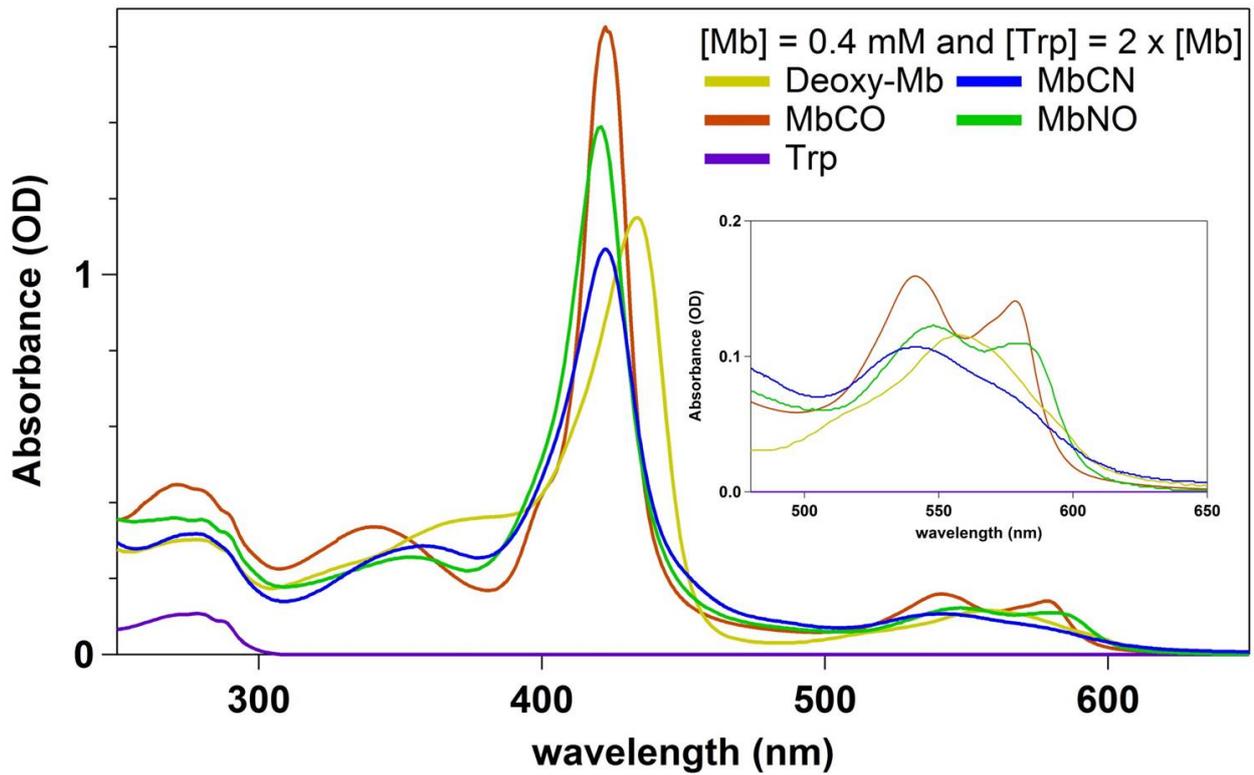


Figure S2 : Static absorption spectra of several myoglobins dissolved in aqueous solution phosphate buffer with ~ 0.4 mM concentration, together with the static absorption spectrum of Tryptophan in similar conditions but with a double concentration. This highlights the weight of the Trp residues in the absorption spectrum.

deoxyMb

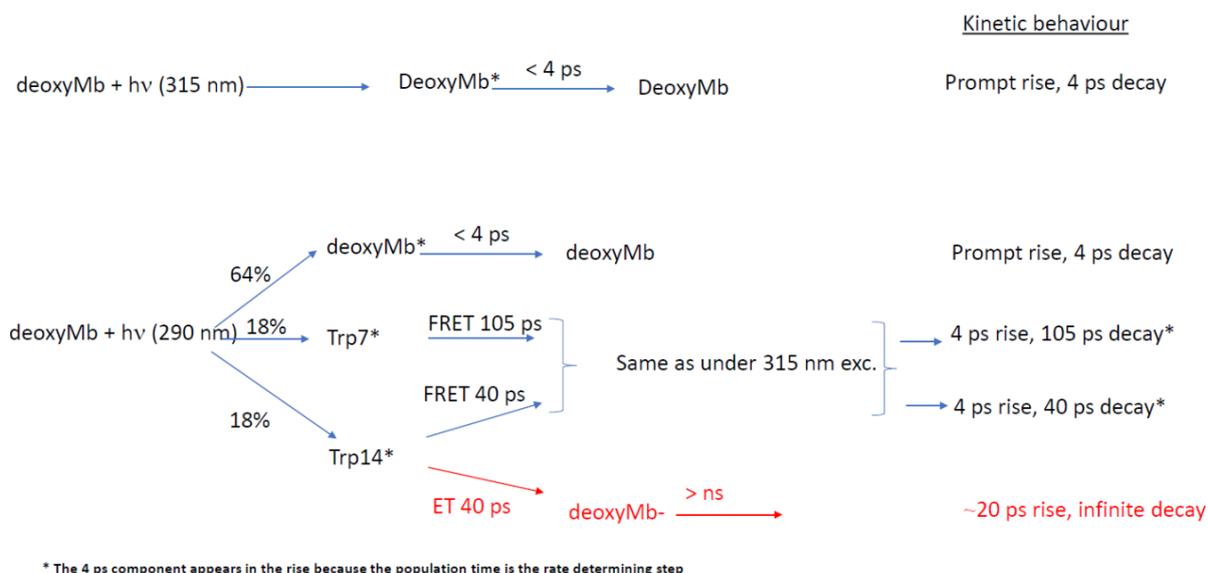


Figure S3: chart showing the processes induced upon 315 nm excitation (top) and 290 nm excitation (bottom) in deoxyMb, as well as the anticipated ones in the latter case.

MbCN and metMb

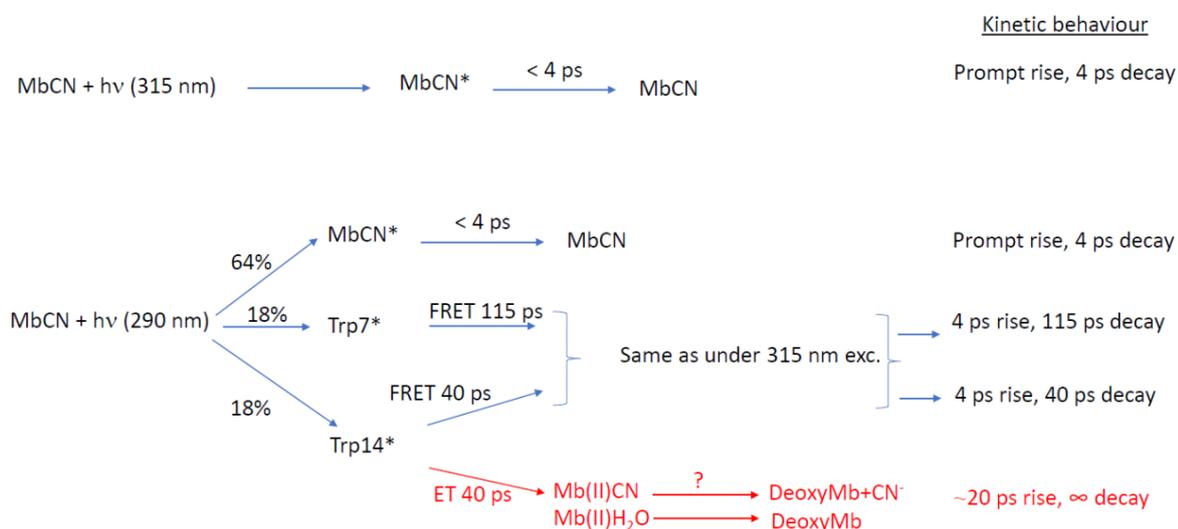


Figure S4: same as figure S3 but for the ferric Mb(III)CN. The results concerning reduction of MetMb(III) are only shown in the bottom line. For both species the outcome of reduction is based on the results of refs {Reddy, 1996 #14813} and {Consani, 2013 #4815}.

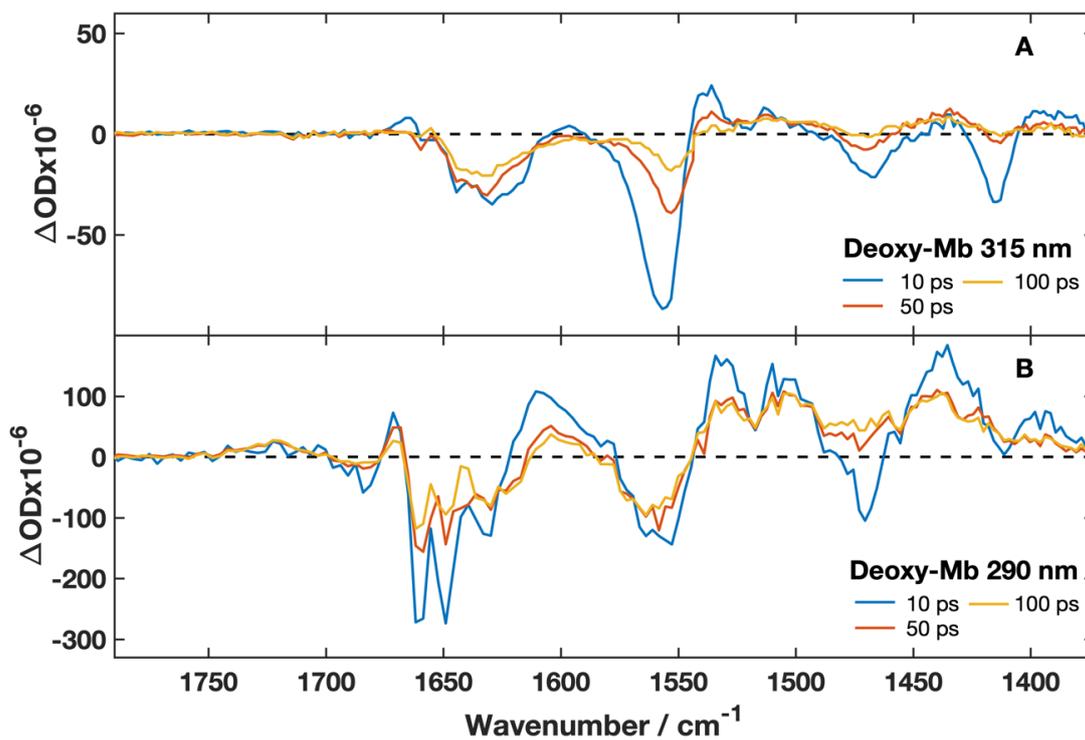


Figure S7: transients at different time delays of deoxyMb under 315 (A) and 290 (B) nm excitation in the 1300-1850 cm^{-1} region.

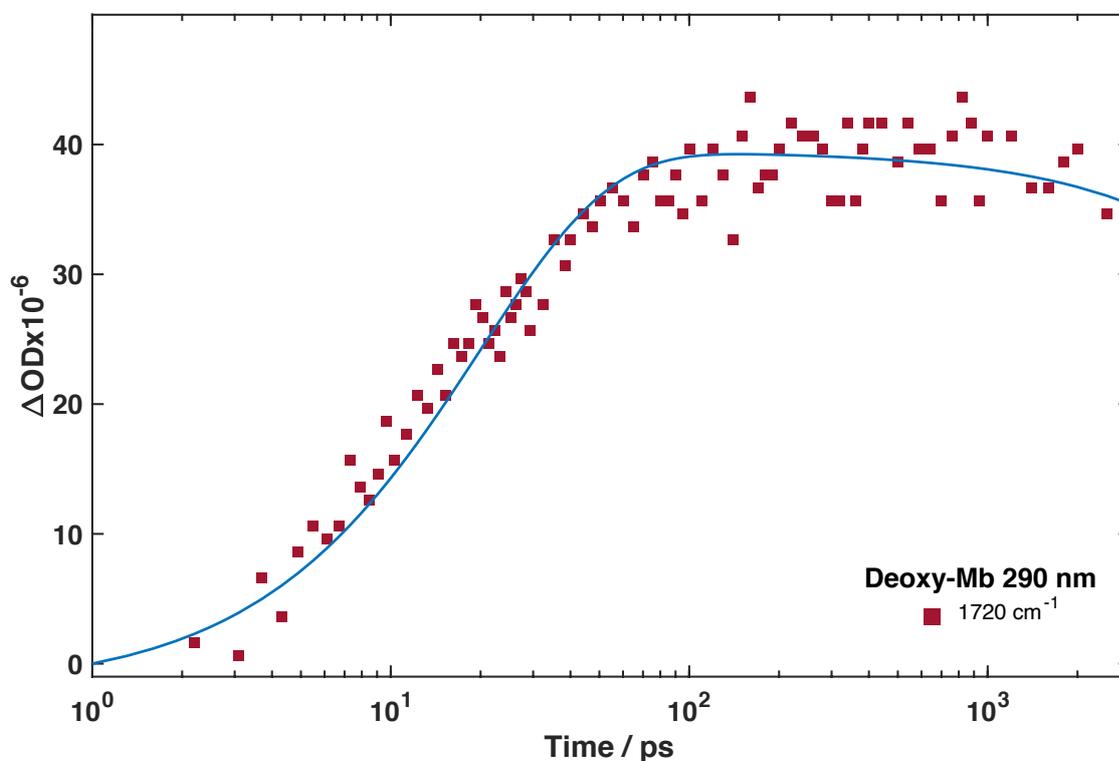


Figure S8: Kinetic trace of the 1720 cm^{-1} band of deoxyMb upon 290 nm excitation. The blue line shows the fit delivering a rise time of 20 ± 2 ps and a decay of >3 ns.

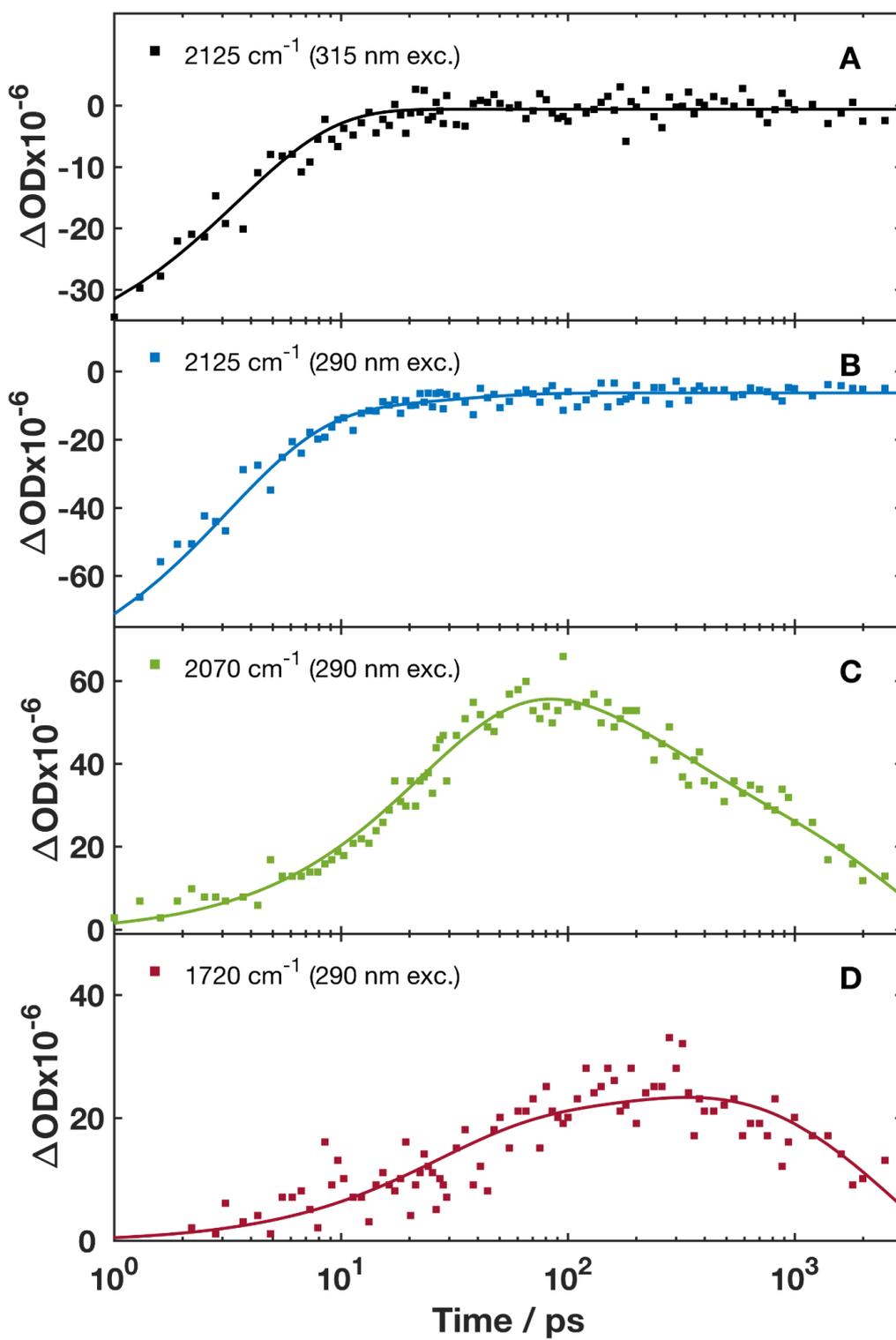


Figure S9: Kinetic traces of MbCN at: (A) 2125 cm^{-1} upon 315 nm excitation and (B-D) at 2125 cm^{-1} , 2070 cm^{-1} and 1720 cm^{-1} upon 290 nm excitation.

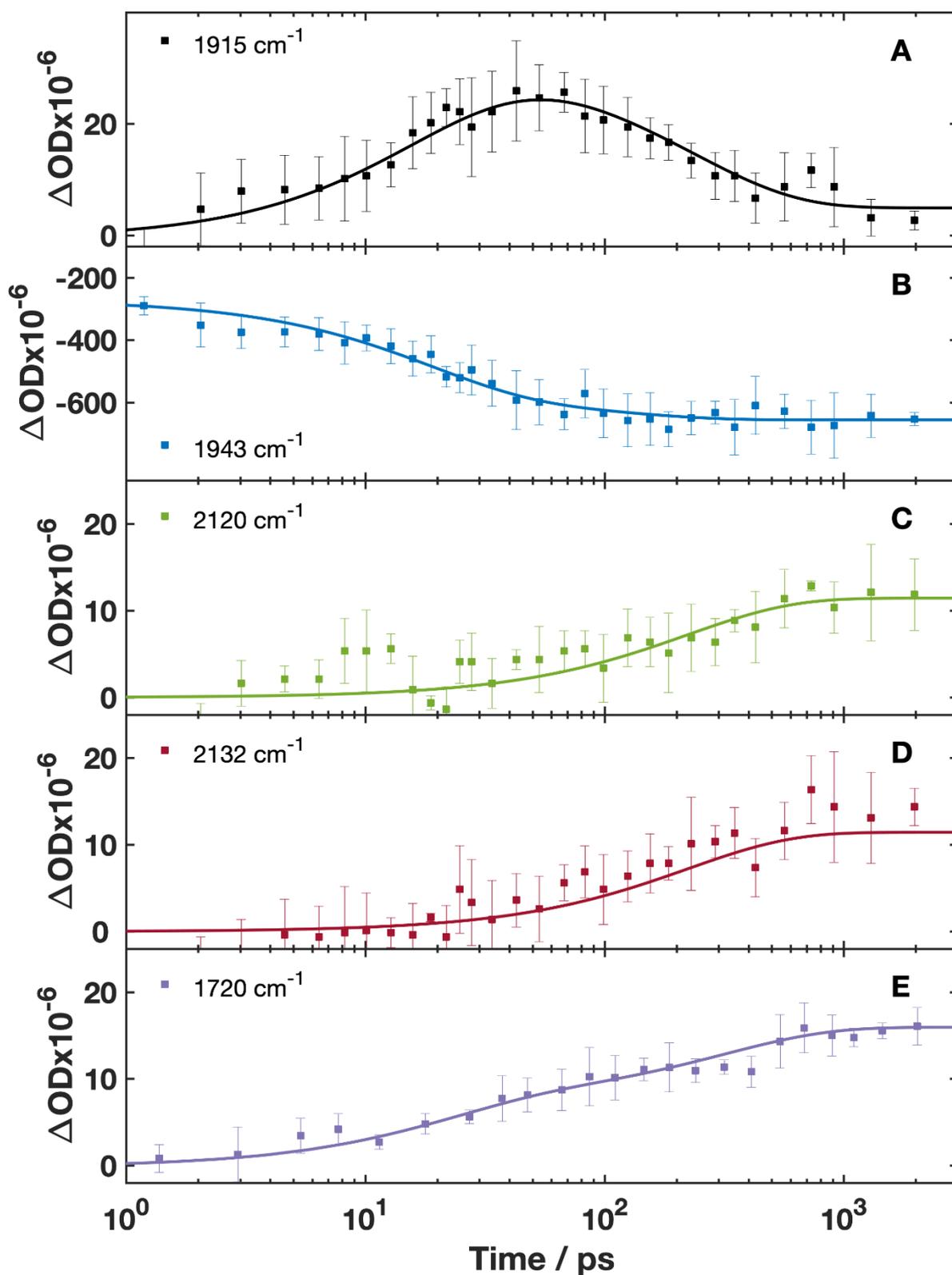


Figure S10: Kinetic traces at 1915 cm^{-1} (A), 1943 cm^{-1} (B), 2120 cm^{-1} (C), 2132 cm^{-1} (D) and 1720 cm^{-1} (E) of MbCO upon 290 nm excitation. Every three delays were averaged together in order to reduce the noise of the kinetic trace. The error bars are the standard deviation of the points.

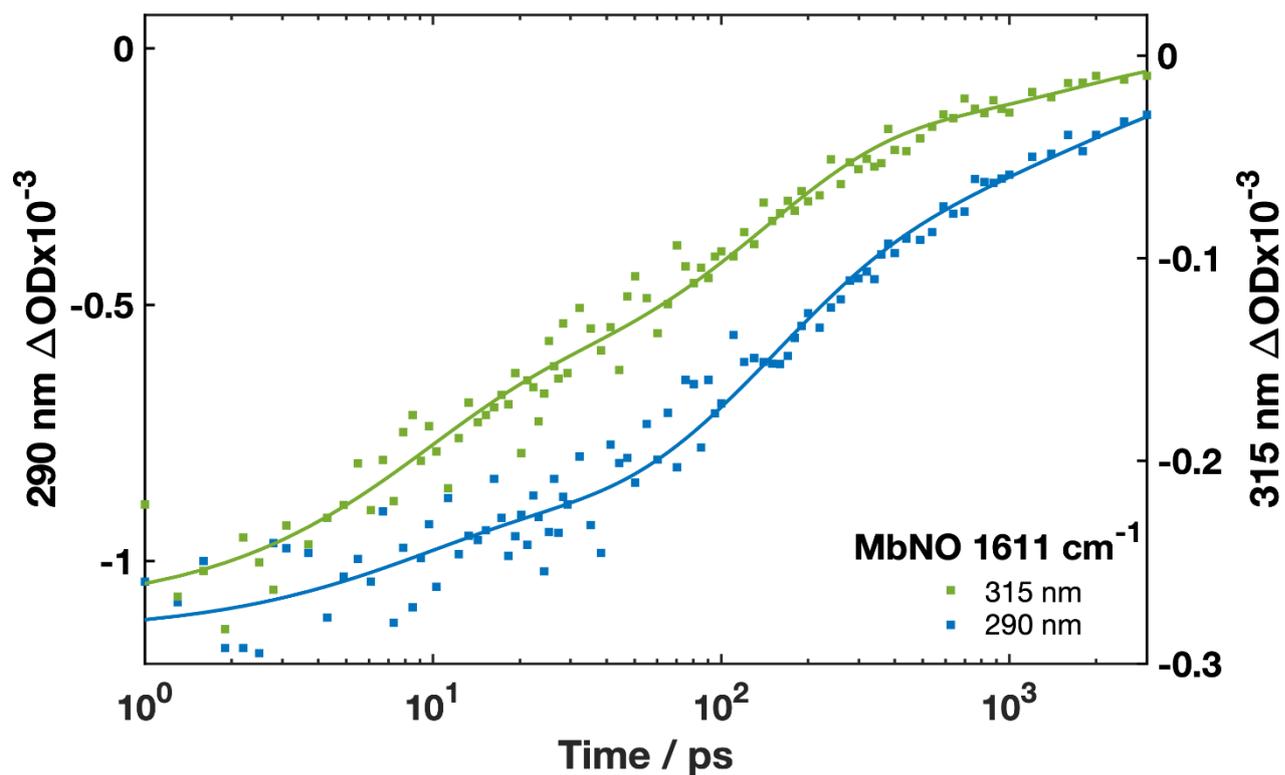


Figure S11: Kinetic traces of the 1611 cm^{-1} bleach band of NO in MbNO upon 315 nm (A) and 290 nm (B) excitation.

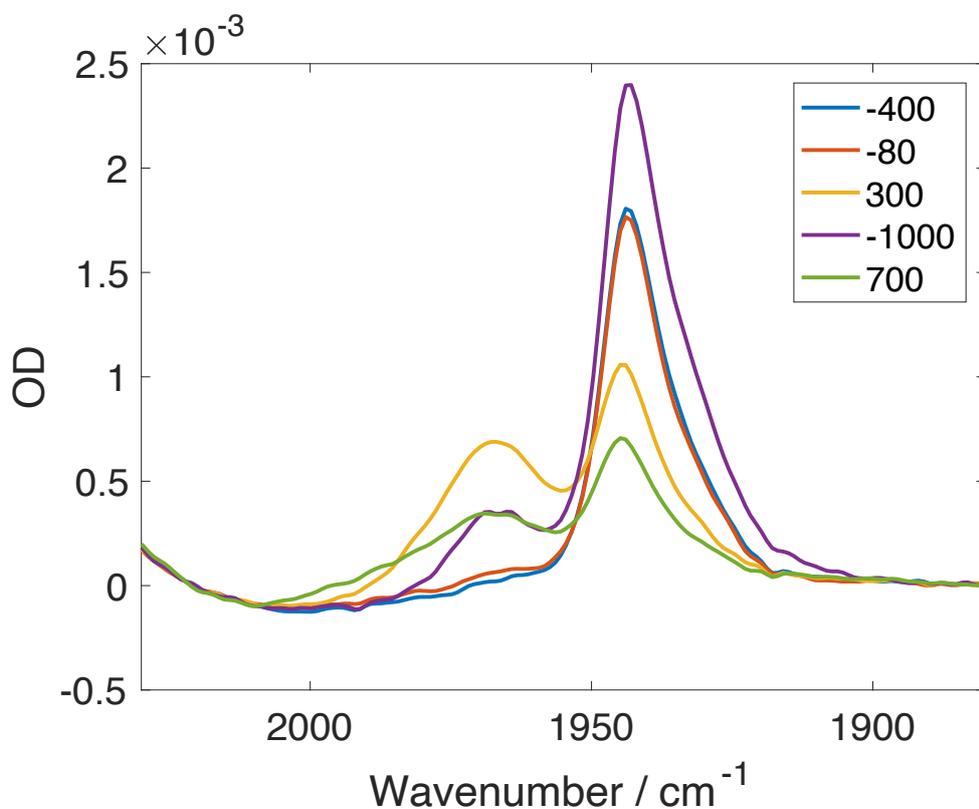


Figure S12: ATR-FTIR spectra of MbCO looking specifically at the CO band. Each spectrum is obtained at a different potential (mV).

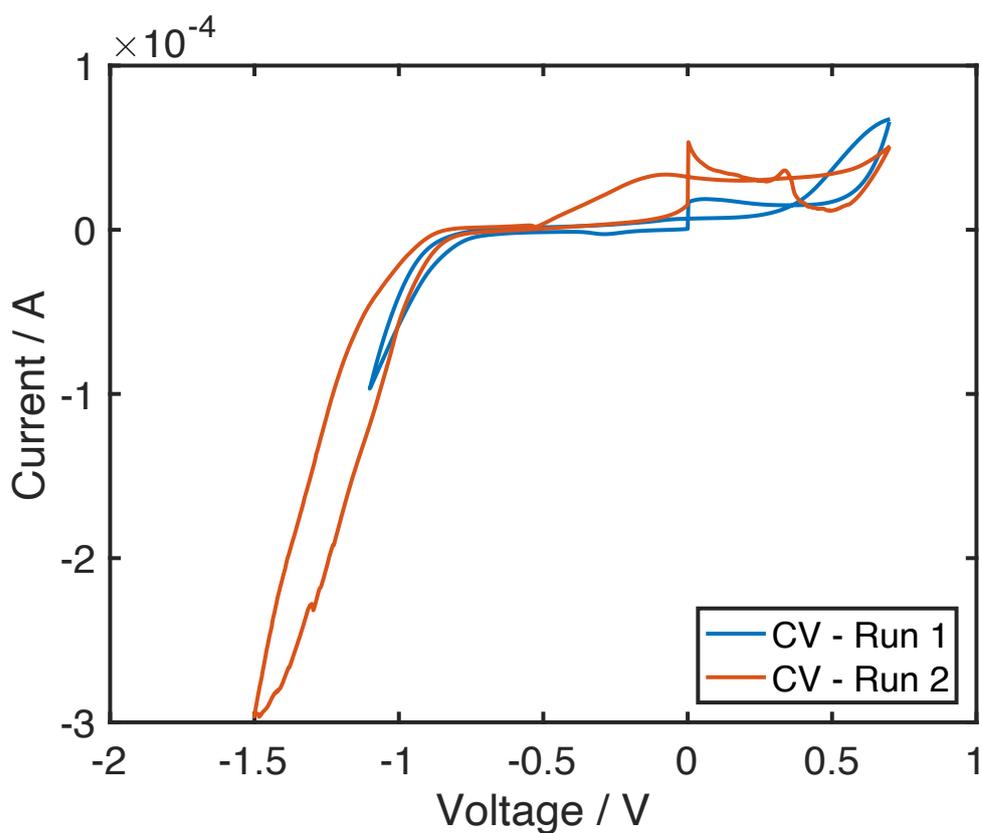


Figure S13: CV of MbCO sample in an OTTLE cell. No evidence of MbCO reduction is observed.

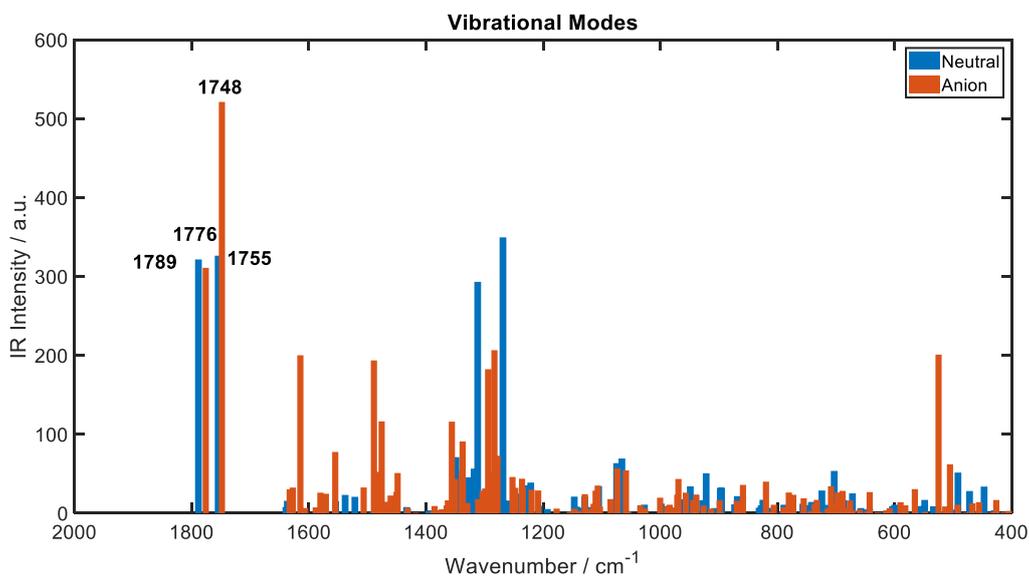


Figure S14: DFT calculated vibrational frequencies for deoxy heme in the neutral and reduced forms. Peak labels highlight the COOH vibrational mode intensity changes due to the heme being reduced.



Figure S15: Deoxy heme calculated structures for the neutral (left) and reduced (right) hemes.

Bibliography

1. Greetham, G. M.; Burgos, P.; Cao, Q.; Clark, I. P.; Codd, P. S.; Farrow, R. C.; George, M. W.; Kogimtzis, M.; Matousek, P.; Parker, A. W.; Pollard, M. R.; Robinson, D. A.; Xin, Z. J.; Towrie, M., ULTRA: A Unique Instrument for Time-Resolved Spectroscopy. *Applied Spectroscopy* **2010**, *64* (12), 1311-9.
2. Greetham, G. M.; Donaldson, P. M.; Nation, C.; Sazanovich, I. V.; Clark, I. P.; Shaw, D. J.; Parker, A. W.; Towrie, M., A 100 kHz time-resolved multiple-probe femtosecond to second infrared absorption spectrometer. *Applied spectroscopy* **2016**, *70* (4), 645-653.
3. Healy, A. J.; Reeve, H. A.; Vincent, K. A. Development of an Infrared Spectroscopic Approach for Studying Metalloenzyme Active Site Chemistry under Direct Electrochemical Control. *Faraday Discuss.* **2010**, *148* (0), 345–357.
4. Gamage, R. S. K. A.; Umapathy, S.; McQuillan, A. J. OTTLE Cell Study of the UV-Visible and FTIR Spectroelectrochemistry of the Radical Anion and Dianion of 1,4-Benzoquinone in DMSO Solutions. *J. Electroanal. Chem. Interfacial Electrochem.* **1990**, *284* (1), 229–235.
5. Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert *et al.*, M. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.* **2015**, *113* (2), 184–215
6. Willis, K. J.; Szabo, A. G.; Zuker, M.; Ridgeway, J. M.; Alpert, B., Fluorescence Decay Kinetics of the Tryptophyl Residues of Myoglobin - Effect of Heme Ligation and Evidence for Discrete Lifetime Components. *Biochemistry* **1990**, *29* (22), 5270-5275.
7. Hochstrasser, R. M.; Negus, D. K., Picosecond Fluorescence Decay of Tryptophans in Myoglobin. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **1984**, *81* (14), 4399-4403.
8. Glandières, J. M.; Twist, C.; Haouz, A.; Zentz, C.; Alpert, B., Resolved fluorescence of the two tryptophan residues in horse apomyoglobin. *Photochemistry and photobiology* **2000**, *71* (4), 382-386.

Appendix C

**Schengen-pathway controls spatially
separated and chemically distinct lignin
deposition in the endodermis**

Two chemically distinct root lignin barriers control solute and water balance

Guilhem Reyt¹, Priya Ramakrishna ^{1,10}, Isai Salas-González ², Satoshi Fujita ^{3,11}, Ashley Love⁴, David Tiemessen⁴, Catherine Lapierre⁵, Kris Morreel^{6,7}, Monica Calvo-Polanco ^{8,12}, Paulina Flis ¹, Niko Geldner ³, Yann Boursiac ⁸, Wout Boerjan ^{6,7}, Michael W. George^{4,9}, Gabriel Castrillo¹ & David E. Salt ¹ 

Lignin is a complex polymer deposited in the cell wall of specialised plant cells, where it provides essential cellular functions. Plants coordinate timing, location, abundance and composition of lignin deposition in response to endogenous and exogenous cues. In roots, a fine band of lignin, the Casparian strip encircles endodermal cells. This forms an extracellular barrier to solutes and water and plays a critical role in maintaining nutrient homeostasis. A signalling pathway senses the integrity of this diffusion barrier and can induce over-lignification to compensate for barrier defects. Here, we report that activation of this endodermal sensing mechanism triggers a transcriptional reprogramming strongly inducing the phenylpropanoid pathway and immune signaling. This leads to deposition of compensatory lignin that is chemically distinct from Casparian strip lignin. We also report that a complete loss of endodermal lignification drastically impacts mineral nutrients homeostasis and plant growth.

¹Future Food Beacon of Excellence & School of Biosciences, University of Nottingham, Sutton Bonington, UK. ²Curriculum in Bioinformatics and Computational Biology, Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ³Department of Plant Molecular Biology, Biophore, University of Lausanne, Lausanne, Switzerland. ⁴School of Chemistry, University of Nottingham, Nottingham, UK. ⁵Institut Jean-Pierre Bourgin, INRAE, AgroParisTech, Université Paris-Saclay, Versailles, France. ⁶Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium. ⁷Center for Plant Systems Biology, VIB, Ghent, Belgium. ⁸Biochimie & Physiologie Moléculaire des Plantes, University of Montpellier, CNRS, INRAE, SupAgro, Montpellier, France. ⁹Department of Chemical and Environmental Engineering, The University of Nottingham Ningbo China, Ningbo, China. ¹⁰Present address: Department of Botany and Plant Biology, University of Geneva, Geneva, Switzerland. ¹¹Present address: National Institute of Genetics, Mishima, Shizuoka, Japan. ¹²Present address: Excellence Unit AGRIENVIRONMENT, CIALE, University of Salamanca, Salamanca, Spain. email: david.salt@nottingham.ac.uk

Lignin is a phenolic polymer and is one of the main components of secondary-thickened cell wall (CW) in vascular plants. Its chemical properties give strength, stiffness and hydrophobicity to the CW. Lignin provides mechanical support, modulates the transport of water and solutes through the vascular systems, and provides protection against pathogens^{1,2}. Lignin polymerization occurs through oxidative coupling of monolignols and other aromatic monomers^{3,4}. The monolignols that is *p*-coumaryl, coniferyl and sinapyl alcohols are synthesized from the amino acid phenylalanine through the phenylpropanoid pathway. They are then polymerized into lignin to form the *p*-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) subunits of the lignin polymer. Lignin composition and abundance are highly variable among and within plants species, tissues and cell types and can be modulated by environmental cues¹.

In roots, large amounts of lignin are deposited in xylem vessels, an important component of the vascular system^{5,6}. Lignin is also deposited in the endodermal cells surrounding the vascular tissues, for Casparian strip (CS) formation⁷. Both the vascular system and the CS play a critical role in water and mineral nutrient uptake from the soil, and their transport toward the shoot^{8–10}. In *Arabidopsis thaliana*, the composition of lignin monomers in the CS and xylem is similar with a strong predominance of the G-unit (>90%)⁷. However, the machinery required for CS lignification appears to be distinct from that needed for xylem lignification^{6,11}.

The deposition of the CS in the endodermal CW prevents the apoplastic diffusion of solutes between the outer and inner tissues of the root, forcing solutes to pass through the symplast of endodermal cells⁸. CS lignin encircles each endodermal cell, forming a bridge between them. This precise lignin deposition is defined by the presence of the transmembrane Casparian strip membrane domain proteins (CASPs)¹², peroxidases^{13,14} and the dirigent-like protein ESB1¹⁵. The expression of this lignin polymerization machinery is tightly controlled by the transcription factor MYB36^{16,17}. A surveillance mechanism for CS integrity, called the Schengen-pathway, boosts CS deposition and is necessary for CS fusion and sealing of the extracellular space (apoplast)¹⁸. This pathway involves vasculature-derived peptides CASPARIAN STRIP INTEGRITY FACTORS 1 and 2 (CIF1 and 2)^{19,20} and their perception by the leucine-rich repeat receptor-like kinase SCHENGEN3 (SGN3, also called GSO1). Their interaction triggers a cascade of signalling events mediated by kinases, that involves SGN1, and the activation of the NADPH oxidase RBOHF (SGN4) leading to the ROS production necessary for lignin polymerization^{13,18,21}. These kinase signalling events occur on the cortex-facing side of the CS and mediate the transition from a discontinuous CS with islands of lignin into a continuous CS with its characteristic ring of lignin sealing the apoplast¹⁸. Once the CS is sealed, CIF peptide diffusion is blocked and the Schengen-pathway becomes inactive. In mutants with an impaired CS, such as *esb1* and *myb36*, the Schengen-pathway is constitutively activated due to a constant leak of the CIF(s) peptides through the CS^{12,15,17,18,22}. This induces in endodermal CW compensatory lignification in the cell corners and suberisation of the cell surface. However, the role of this compensatory lignin, and the mechanism controlling its deposition are not fully understood.

Here, we demonstrate that constitutive activation of the Schengen-pathway induces the deposition of compensatory lignin in the corners of endodermal cells that is chemically distinct from CS lignin. We characterized this lignin and found commonalities with stress- and pathogen-response lignin, which has a high content of the H subunit. Furthermore, we demonstrate that this cell-corner lignification is preceded by a transcriptional reprogramming of endodermal cells, causing a strong induction of the

phenylpropanoid pathway, and a significant inactivation of aquaporin expression. Our findings also establish that the activation of the Schengen-pathway, in order to compensate for a defective CS, is of critical importance for plants to maintain their mineral nutrients homeostasis and water balance.

Results and discussion

Two pathways of endodermal lignification. In order to disentangle the role of MYB36 and SGN3 in controlling endodermal lignification, we generated the double mutant *sgn3-3 myb36-2*. We analyzed the endodermal accumulation of lignin in *sgn3-3 myb36-2*, and the corresponding single mutants *sgn3-3* and *myb36-2* (Fig. 1a–c). In the early stage of endodermal differentiation, six cells after the onset of elongation, we observed deposition of CS lignin in “a string of pearl” manner in WT and *sgn3-3* (Fig. 1a). No endodermal lignification was observed in *myb36-2* or *sgn3-3 myb36-2* at this early developmental stage (Fig. 1a). Later in endodermal development, ten cells after the onset of elongation, we observed a continuous CS ring of lignin, sealing the endodermal cells in WT plants (Fig. 1a–c). As we expected, in *sgn3-3* impaired in the activation of the Schengen-pathway, CS lignification is discontinuous⁹, while *myb36-2* exhibits compensatory lignification in the corners of endodermal cells facing the cortical side of the endodermis, as previously reported^{9,17}. In contrast, no lignification at the CS or cell corners was observed in *sgn3-3 myb36-2* (Fig. 1a–c). These results establish that the cell-corner compensatory lignification observed in *myb36-2* lacking a CS (17) is SGN3 dependent.

To test how these different patterns of endodermal lignification found in WT, *sgn3-3*, *myb36-2* and *sgn3-3 myb36-2* affect the permeability of the root apoplast, we assessed the penetration into the stele of the fluorescent apoplastic tracer propidium iodide (PI)²³ (Fig. 1d). We quantified the percentage of the root length permeable to PI and found it partially increased in both *sgn3-3* and *myb36-2*, in comparison to WT (Fig. 1d). Surprisingly, we observed in *sgn3-3 myb36-2* that the entire length of the root was permeable to PI, indicating an additive effect of *sgn3-3* and *myb36-2* mutations. This result suggests that MYB36 and SGN3 control apoplastic sealing through two independent lignification pathways. The lack of compensatory cell-corner lignification in *sgn3-3 myb36-2* could explain the complete permeability of the roots in the double mutant. This finding supports recent observations assigning a role as an apoplastic barrier to SGN3-dependent cell-corner lignification¹⁴. In addition, constitutive activation of the Schengen-pathway is also known to trigger an enhanced suberisation in certain CS mutants, including *myb36*¹⁷. We confirm this observation in *myb36-2* (Fig. 1e). This enhanced suberisation in *myb36* is also SGN3 dependent since *sgn3-3 myb36-2* shows the same pattern of suberisation as WT plants (Fig. 1e).

Our results indicate that MYB36 and SGN3 control endodermal lignification through two pathways: (a) a pathway involved in CS lignification controlled by both MYB36 and SGN3; and (b) a pathway involved in compensatory lignification of the endodermal cell corners controlled exclusively by SGN3.

Cell-corner lignin is chemically distinct from CS lignin. We investigated the chemical nature and biochemical origins of CS lignin and compensatory cell-corner lignin. For this, we used confocal Raman microscopy on root cross-sections, to spatially resolve the chemistry of these different types of lignin. We triggered endodermal cell-corner lignin deposition by feeding WT plants with CIF2 peptide (+CIF2), the ligand of the SGN3 receptor, in order to activate the Schengen-pathway. We separately imaged regions of interest containing CS lignin in WT

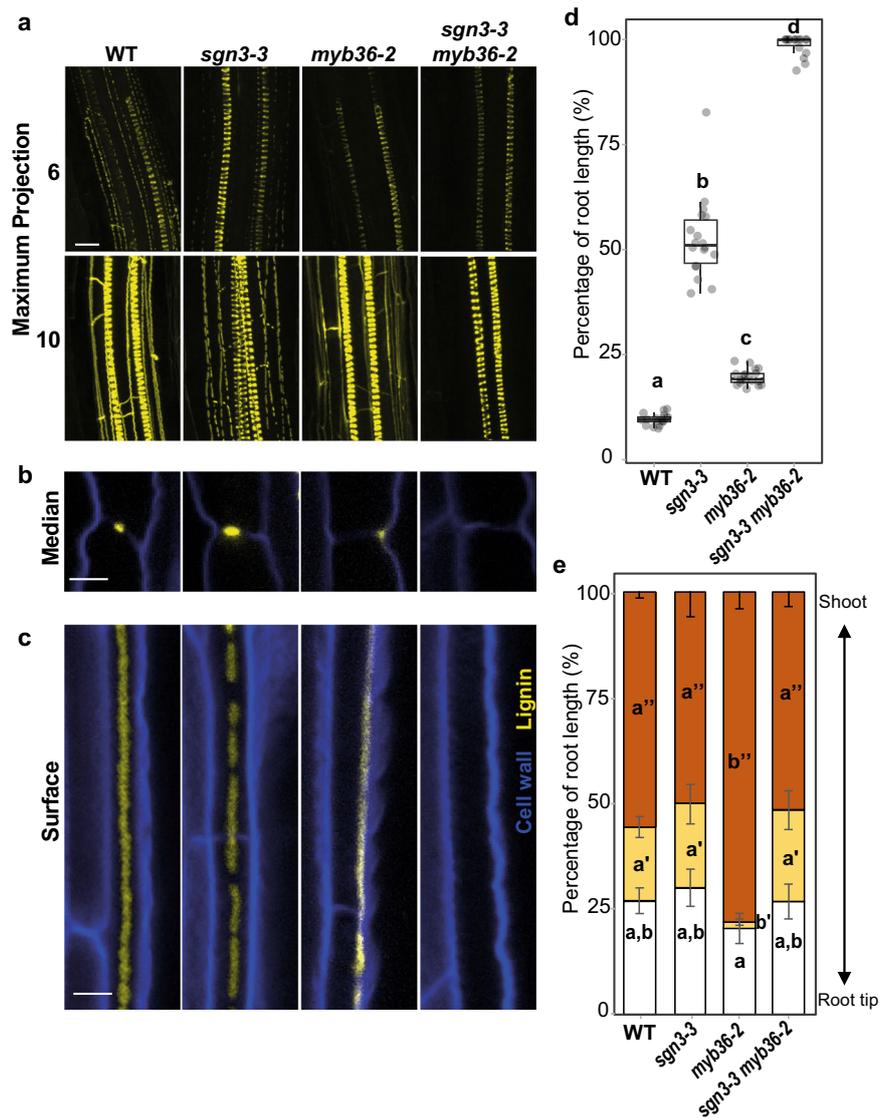


Fig. 1 Disruption of *MYB36* and *SGN3* abolishes endodermal lignification and root apoplastic barrier. **a** Maximum projection of lignin staining at the sixth and tenth endodermal cell after the onset of elongation. Spiral structures in the centre of the root are xylem. Scale bar = 10 μm . Median (**b**) and surface (**c**) view of an endodermal cell at ten cells after the onset of elongation. Scale bar = 5 μm . The roots were cleared and stained with basic fuchsin (yellow) for lignin and with Calcofluor white (blue) for cellulose. The experiment was repeated three times independently with similar results. **d** Boxplot showing the percentage of the root length permeable to propidium iodide. $n = 18$ from two independent experiments. Different letters represent significant differences between genotypes using a two-sided Mann-Whitney test ($p < 0.01$). Centre lines show the medians; box limits indicate the 25th and 75th percentiles. **e** Quantification of suberin staining along the root. The results are expressed in percentage of root length divided in three zones: unsaturated (white), discontinuously saturated (yellow), continuously saturated (orange). $n = 5$ plants for WT, $n = 6$ plants for *sgn3-3*, $n = 7$ plants for *myb36-2* and $n = 7$ plants for *sgn3-3 myb36-2*, error bars: SD, the centre of the error bars represents the mean. Individual letters show significant differences using a two-sided Mann-Whitney test between the same zones ($p < 0.01$). The experiment was repeated two times independently with similar results.

plants, endodermal cell-corner lignin in WT treated with CIF2, and xylem lignin from WT plants, treated or not with CIF2 (Supplementary Fig. 1). Then, we used multivariate curve resolution (MCR) analysis on these Raman images to spatially and spectrally resolve lignin in these different regions. We observed that the CS lignin spectrum is distinct from that of endodermal cell-corner lignin (Fig. 2a, b). For example, peaks known to be assigned to lignin display higher (ex: 1337 cm^{-1} , aliphatic OH bend²⁴) and lower (ex: 1606 cm^{-1} , aromatic ring stretch²⁴) intensity in CS lignin in comparison to endodermal cell-corner lignin. Another striking difference was observed for the peak at $1656\text{--}1659\text{ cm}^{-1}$ assigned to a double bond conjugated to an aromatic ring (e.g.: coniferyl alcohol or coniferaldehyde,²⁴). This

peak is missing in the endodermal cell-corner lignin of WT treated with CIF2 in comparison with CS lignin, suggesting a change in the phenolic composition of the cell-corner lignin. Conversely, the xylem lignin spectrum of plants treated with or without CIF2 was similar, with the most intense peaks showing comparable intensity. This suggests that changes in lignin composition triggered by the constitutive activation of the Shengeng pathway mainly occur in the endodermis, and xylem lignin remains largely unaffected.

These conclusions were further confirmed spatially by mapping the intensity of these different lignin spectra on large Raman maps containing xylem and endodermal lignin in WT plants treated or not treated with CIF2 (Fig. 2c). We observed that the

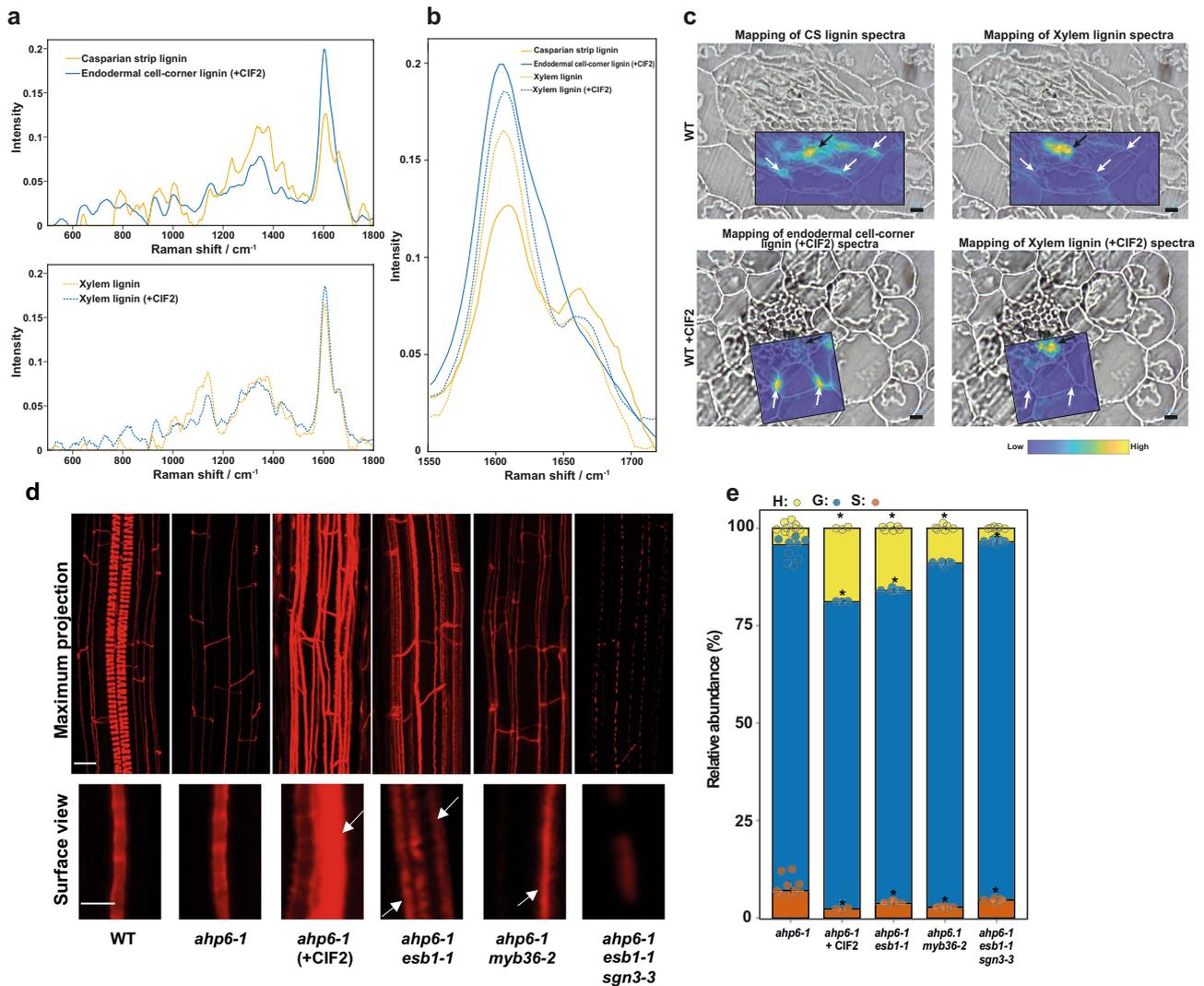


Fig. 2 Activation of the Schengen-pathway triggers the deposition of a distinct “stress” lignin in the endodermis. **a** Raman spectra of lignin of the different regions of interest presented in Supplementary Fig. 1 and determined using a Multivariate Curve Resolution (MCR) analysis. The MCR analysis was performed on small Raman maps from independent plants containing CS lignin of WT ($n = 8$), cell-corner lignin of WT treated with CIF2 (+CIF2; $n = 5$) and for xylem lignin of WT ($n = 2$) and xylem lignin of WT treated with CIF2 ($n = 2$). **b** Close view of Raman spectra presented in (a) in the lignin aromatic region between 1550 and 1700 cm^{-1} . **c** Large Raman maps in root cross-sections of WT and WT treated with CIF2 (+CIF2). The intensity of the different lignin spectra presented in (a) was mapped onto large Raman maps containing xylem and endodermal lignin. Scale bar = 5 μm . **d** Lignin staining with basic fuchsin at a distance of 3 mm from the root tip in WT, *ahp6-1*, *ahp6-1 esb1-1*, *ahp6-1 esb1-1 sgn3-3*, *ahp6-1 myb36-2* and *ahp6-1* treated with CIF2. The plants were grown for 6 days in presence of 10 nM 6-Benzylaminopurine (BA). Upper panel shows a maximum projection of the root (scale bar = 10 μm). Spiral structures in the centre only observed in the WT root are protoxylem. Lower panel shows surface view of endodermal cells. The experiment was repeated two times independently with similar results (scale bar = 5 μm). White arrows indicate ectopic lignification. **e** Relative abundance of the lignin monomers released by thioacidolysis (*p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units) in root tips of *ahp6-1* ($n = 9$), *ahp6-1* treated with CIF2 (+CIF2; $n = 3$), *ahp6-1 esb1-1* ($n = 6$), *ahp6-1 myb36-2* ($n = 6$) and *ahp6-1 esb1-1 sgn3-3* ($n = 6$). Asterisks represent significant differences from the *ahp6-1* control for each individual monomer using a two-sided Mann–Whitney *U* test (p value < 0.05).

CS lignin spectrum localizes to the CS and xylem vessels suggesting a similar lignin composition, as previously shown for monomer composition using thioacidolysis⁷. Additionally, the endodermal cell-corner lignin spectrum localizes almost exclusively to the site of lignification in the corners of the endodermal cells and is essentially absent from the xylem. Furthermore, the xylem lignin spectrum in WT plants treated with CIF2 matches exclusively to the xylem vessel and is not observed at the endodermal cell corners. This strongly supports the conclusion that constitutive activation of the Schengen-pathway triggers the deposition of lignin at endodermal cell corners that has a different chemical composition compared to both CS and xylem lignin.

To confirm these differences between CS and endodermal cell-corner lignin, we adopted an approach to directly measure the subunit composition of endodermal lignin avoiding possible contamination from the highly lignified protoxylem cells⁷. We genetically crossed a collection of CS mutants that represent a different level of lignin accumulation in the endodermis with the *arabidopsis histidine transfer protein 6.1* mutant (*ahp6-1*). This mutant, in the presence of low amounts of the phytohormone cytokinin, shows a strong delay in protoxylem differentiation, without affecting CS formation (Fig. 2d)^{7,25}. Therefore, in the resulting lines the majority of lignin derived from the protoxylem is not present allowing us to analyze primarily

lignin with an endodermal origin. To explore how the chemical composition of the cell-corner lignin differs from CS lignin, we collected root tips (3 mm) of 6-day-old *ahp6-1* and *ahp6-1 esb1-1 sgn3-3* mutants accumulating CS lignin only, and from mutants (*ahp6-1 myb36-2* and *ahp6-1 esb1-1*) with cell-corner lignification and a reduced amount of CS lignin. Additionally, as a control we used *ahp6-1* plants treated with the CIF2 peptide that strongly induces the Schengen-pathway and the deposition of cell-corner lignin (Fig. 2d). We measured the relative content of H, G and S subunits in lignin from all samples using thioacidolysis followed by GC-MS (Fig. 2e). We found that CS lignin monomer composition in our control line *ahp6-1* (H: 5%, G: 87%, S: 8%) was similar to that previously reported⁷. The monomer composition of the defective CS in the mutant *ahp6-1 esb1-1 sgn3-3* is overall similar to WT with a small increase in G and decrease in S subunits. Strikingly, we observed that lignin composition in the lines and treatments that induce the accumulation of cell-corner lignin (*ahp6-1 esb1-1*, *ahp6-1 myb36-2*, *ahp6-1*(+CIF2)) was different from the control and mutant lines that only accumulate lignin in the CS. Lignin extracted from plants accumulating cell-corner lignin showed a higher proportion of H subunits. In the case of *ahp6-1* treated with CIF2, H content was increased to 19% and G content was decreased to 79%.

Such a high content of H subunits in lignin is rarely found in angiosperm. Similar levels of H subunits in lignin mainly occur in compression wood of gymnosperm^{26–29} and in defence-induced lignin and have been termed “stress lignin”^{26,30–33}. We therefore conclude that Schengen-pathway induced endodermal cell-corner lignin is a form of “stress lignin”. Taken together, both chemical analysis of lignin subunits by thioacidolysis and spatially resolved confocal Raman spectroscopy show that lignin deposited in endodermal cell corners upon constitutive activation of the Schengen-pathway is H-rich, and chemically and spatially distinct from both CS and xylem lignin.

Cell-corner lignin composition controls root permeability. We then try to determine if the compositional change of cell-corner lignin has functional consequences. For this, we used *myb36-2* displaying cell-corner lignin only and no CS lignin (Fig. 1b, c). Using a pharmacological approach (Supplementary Fig. 2a), we blocked endogenous monolignol production with an inhibitor of the phenylpropanoid pathway, piperonylic acid (PA)⁷. The PA treatment led to a strong reduction of cell-corner lignin deposition (Supplementary Fig. 2b, c) and a strong increase of root permeability as determined with the apoplastic tracer PI (Supplementary Fig. 2d). We then attempted to chemically complement the PA-induced defects by exogenous application of each of the three canonical monolignols *p*-coumaryl (for H subunit), coniferyl (for G subunit), and sinapyl (for S subunit) alcohols and by combining the two main monolignols *p*-coumaryl and coniferyl alcohols that are incorporated into cell-corner lignin (Fig. 2e). The exogenous application of *p*-coumaryl or coniferyl alcohols can slightly trigger cell-corner lignification but, sinapyl alcohol did not (Supplementary Fig. 2b, c). The combined application of *p*-coumaryl and coniferyl alcohols increased the deposition of cell-corner lignin in comparison with their individual application. None of the three monolignols alone can fully recover the PA-induced defect on root permeability (Supplementary Fig. 2d). The combined application of *p*-coumaryl and coniferyl alcohols fully recovered the effect of PA on root permeability in the *myb36* mutant. This chemical complementation assay indicates that the type of monolignols available controls both cell-corner lignin deposition, and its capacity to seal the endodermal apoplast. The combination of *p*-coumaryl and

coniferyl alcohols is the most effective for increasing lignin deposition and for sealing the apoplast.

Schengen control of the phenylpropanoid pathway. To investigate the biosynthesis of the endodermal H-rich stress lignin, we performed RNA-seq using root tips (5 mm) of WT plants, WT treated with exogenous CIF2, and the mutants *myb36-2* and *esb1-1* that show activation of the Schengen-pathway, and root tips of *sgn3-3*, *esb1-1 sgn3-3*, *sgn3-3 myb36-2* and *sgn3-3* treated with exogenous CIF2 with no Schengen signalling. Clustering analysis of the differentially expressed genes (DEGs) shows that roots displaying cell-corner lignification due to the constitutive activation of the Schengen-pathway (WT treated with exogenous CIF2, *myb36-2* and *esb1-1*) share a similar transcriptional response that is distinct from that observed in the other genotypes (Fig. 3a, Supplementary Fig. 3a and Supplementary Data 1). CIF2 application to *sgn3-3* shows a similar transcriptional response to non-treated WT and *sgn3-3* and does not trigger the transcriptional changes observed during the activation of the Schengen-pathway (WT treated with exogenous CIF2, *myb36-2* and *esb1-1*). This is in line with previous published transcriptomic analysis using plant treated with CIF2¹⁸, and the idea that SGN3 is the only receptor for CIF2 in roots. We observed that genes in cluster C1 are upregulated by the activation of the Schengen-pathway. This cluster is enriched in genes involved in the phenylpropanoid pathway (Supplementary Fig. 3b) and contains a large set of peroxidases and laccases (Supplementary Fig. 3c). We hypothesized that the activation of this pathway would provide the phenolic substrates that are subsequently polymerized by laccases and peroxidases for the enhanced lignification and suberisation induced by the Schengen-pathway. We observed strong activation of expression of genes encoding all the key enzymes of the phenylpropanoid pathway required for monolignol biosynthesis, with the exception of *C3'H*, *C3H*, *HCT* and *F5H* (Fig. 3b and Supplementary Fig. 3d). H-rich lignin is known to be accumulated when expression of *C3'H* or *HCT* is repressed in *A. thaliana* and poplar^{35–40}. This activation of all the main enzymes of the phenylpropanoid pathway, apart from *C3'H* and *HCT*, observed after triggering the Schengen-pathway, could explain the high level of H-units incorporation into endodermal cell-corner lignin (Fig. 3b and Supplementary Fig. 3d). Similarly, the roots of the cellulose synthase isomer mutant *ectopic lignification1 (eli1)* accumulate H-rich lignin and display strong gene activation for most of the phenylpropanoid pathway, with the exception of *C3'H*³⁴. Interestingly, ectopic lignification in *eli1* is also under the control of another receptor-like kinase, *THE1* (THESEUS), also involved in CW integrity sensing^{35,36}.

We then tried to identify transcriptional regulators with a role in the Schengen-pathway controlled regulation of phenylpropanoid synthesis. We performed a gene expression correlation analysis between the phenylpropanoid pathway genes and their transcriptional regulators^{3,37} (Supplementary Fig. 3d). We found that the expression of the transcription factor *MYB15* highly correlates with the expression of most of the genes required for monolignol biosynthesis, with the notable exception of *C3'H* (Supplementary Fig. 3d). Upregulation of *MYB15* in response to CIF2 has been previously shown¹⁸. This transcription factor is known to bind to the promoter of *PAL1*, *C4H*, *HCT*, *CCoAOMT1* and *COMT* but does not bind to the promoter of *C3'H* and *F5H*³⁸. Schengen-pathway activation of *MYB15* expression provides a plausible mechanism to explain the induction of the main enzymes of the phenylpropanoid pathway with the exception of *C3'H* and *F5H*. This modulation of gene expression could explain the enhanced incorporation of *p*-coumaryl alcohol into the stress lignin we observed at

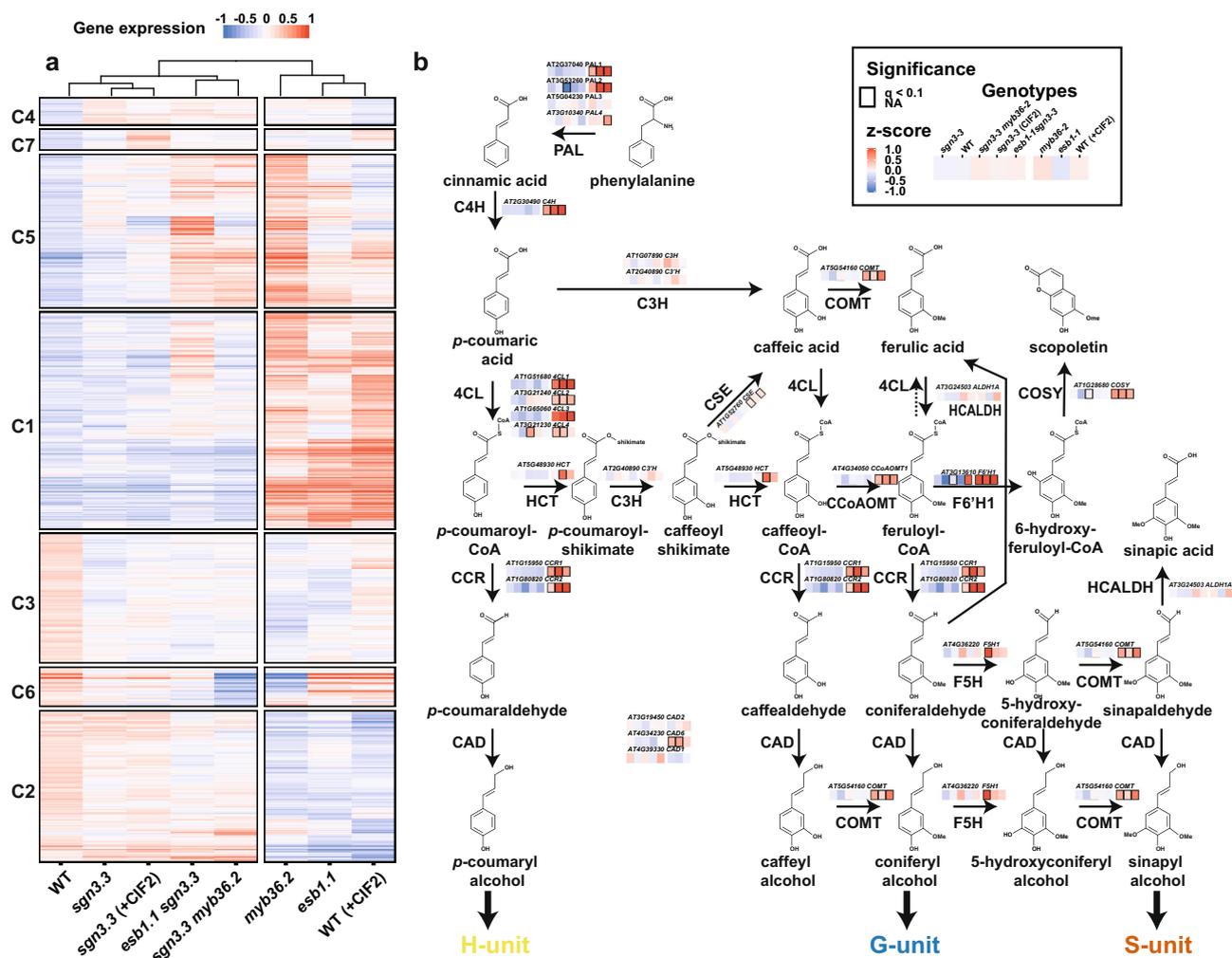


Fig. 3 Modulation of the phenylpropanoid pathway by the Schengen-pathway. **a** Heatmap of the 3564 differentially expressed genes identified in the RNA-seq in root tips of wild-type (WT), *sgn3-3*, *esb1-1*, *myb36-2*, *esb1-1 sgn3-3*, *sgn3-3 myb36-2* plants. Treatment with 100 nM CIF2 was applied as indicated (+CIF2) for WT and *sgn3-3* plants. Clusters (C) are designated with numbers (n = 7). Genes belonging to each cluster are listed in Supplementary Data 1. **b** Phenylpropanoid pathway leading to the lignin monomers and scopoletin biosynthesis (adapted from⁷²). Solid arrows represent enzymatic steps. Gene expression from the genes selected in Supplementary Fig. 3d was mapped on the pathway according to their KEGG enzyme nomenclature. Only the genes with a demonstrated function in lignin biosynthesis as listed in Supplementary Fig. 3d were mapped. PAL PHENYLALANINE AMMONIA-LYASE, C4H CINNAMATE 4-HYDROXYLASE, 4CL 4-COUMARATE:CoA LIGASE, HCT *p*-HYDROXYCINNAMOYL-CoA:QUINATE/SHIKIMATE *p*-HYDROXYCINNAMOYLTRANSFERASE, C'3H *p*-COUMARATE 3'-HYDROXYLASE, C3H COUMARATE 3-HYDROXYLASE, CSE CAFFEOYL SHIKIMATE ESTERASE, CCoAMT CAFFEOYL-CoA O-METHYLTRANSFERASE, CCR CINNAMOYL-CoA REDUCTASE, F5H FERULATE 5-HYDROXYLASE, COMT CAFFEIC ACID O-METHYLTRANSFERASE, CAD CINNAMYL ALCOHOL DEHYDROGENASE, HCALDH HYDROXYCINNAMALDEHYDE DEHYDROGENASE, COSY COUMARIN SYNTHASE, F6'H1 FERULOYL COA ORTHO-HYDROXYLASE 1.

endodermal cell corners. Interestingly, increased incorporation of *p*-coumaryl alcohol into lignin has been found in response to *Pseudomonas syringae* and is partially controlled by the MYB15 transcription factor³⁹. MYB15 is an activator of basal immunity in *A. thaliana* through induction of the synthesis of defence lignin and soluble phenolics³⁸.

To test if constitutive activation of the Schengen-pathway leads to the production of defence-inducible soluble phenolics, we undertook secondary metabolite profiling using ultra high performance liquid chromatography (UHPLC). This analysis was performed using root tips (5 mm) of the *esb1-1* mutant having a defective CS and constitutive activation of the Schengen-pathway, *sgn3-3* and *sgn3-3 esb1-1* mutants having a defective CS and inactivation of the Schengen-pathway and in WT plants. We observed distinct accumulation of soluble secondary metabolites across the different genotypes (Supplementary Fig. 4 and Supplementary Data 2). We

identified 20 out of 52 phenolic compounds that differentially accumulate specifically due to the activation of the Schengen-pathway. We found higher accumulation of the conjugated neolignan G(8-O-4)*p*CA, scopoletin, flavonoid derivatives such as conjugated kaempferol (astragalin and 4'-O-acetylkaempferol-3-O-hexoside), isorhamnetin and acetylhyperoside. Scopoletin biosynthesis is controlled by the enzyme F6'H1 and COSY^{40,41} and the transcription factor MYB15³⁸. We found that the expression of the three genes encoding these proteins is induced by the constitutive activation of the Schengen-pathway (Fig. 3b and Supplementary Fig. 3d). Scopoletin is a modulator of plant-microbe interaction^{45,49-52}. In addition to that, we found a strong induction of genes related to defence (response to chitin/systemic acquired resistance/immune response/hypersensitive response) among the genes induced by the activation of the Schengen-pathway (C1; Fig. 3a and Supplementary Fig. 3b). This is consistent with a previous publication showing

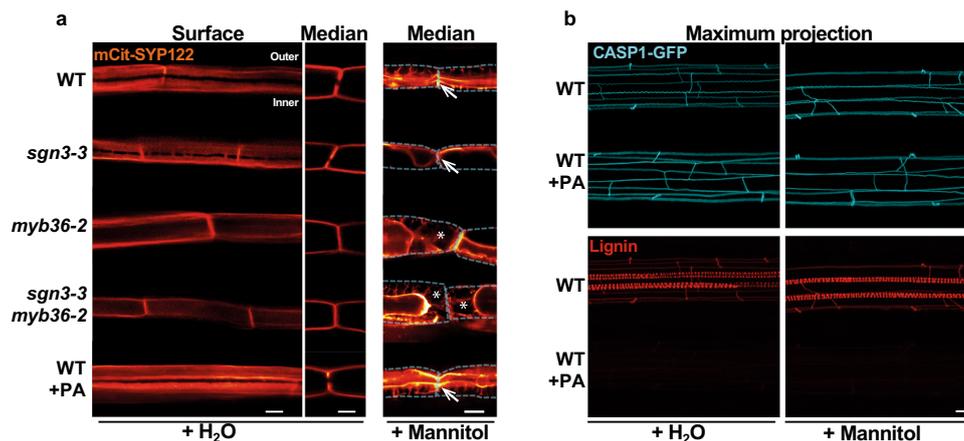


Fig. 4 PM attachment to the CW is MYB36-dependent but does not rely on lignin deposition. **a** Median and surface view of the endodermal plasma membrane using the marker line pELTP::SYP122-mCitrine before plasmolysis (+H₂O) and after plasmolysis (+Mannitol) at 15 cells after the onset of elongation. WT plants were treated or not from germination with 10 μM piperonylic acid (+PA). White asterisks show the exclusion domain at the CSD. The dashed line represents the contours of the cells before plasmolysis. Arrows show the plasma membrane attachment to the cell wall. Blue asterisks show the plasmolysis generated space where no attachment is observed. Scale bar = 5 μm. “inner” designates the stele-facing endodermal surface, “outer”, the cortex-facing surface. The experiment was repeated three times independently with similar results. **b** Maximum projection of CASP1-GFP and lignin staining with basic fuchsin in cleared roots from plants grown with or without 10 μM piperonylic acid and subjected to plasmolysis with Mannitol. Scale bar = 10 μm. The experiment was repeated two times independently with similar results.

similarities between the Schengen-pathway and the microbe-associated molecular patterns signalling pathway¹⁸.

Local activation of genes related to defence. We then tried to determine if this induction of genes related to defence occurs locally or systemically. For this, we used a split-root system, in which the roots of a single plant were physically separated for 3 days. One half of the root system was exposed to the Schengen-pathway activator CIF2, whereas the other side was kept in the same medium without CIF2 (Supplementary Fig. 5a). We observed that cell-corner lignification is deposited only in the root with direct contact with CIF2 (Supplementary Fig. 5b) as observed for the induction of the expression of peroxidase genes (Supplementary Fig. 5c). Similarly, several genes related to defence were found to be induced only locally in the presence of CIF2 (Supplementary Fig. 5d).

The constitutive activation of the Schengen-pathway and subsequent cell-corner lignification described here are triggered locally when the integrity of the endodermal apoplastic barrier is lost. This can also occur during developmental processes such as lateral root emergence and during infection with pathogens. Interestingly, the deposition of endodermal barriers has been associated with increased resistance against a large range of soil-borne pathogens such as *Aphanomyces euteiches*⁴², *Ralstonia solanacearum*⁴³, *Phytophthora sojae*⁴⁴ and nematodes⁴⁵. The resistance to *A. euteiches*, *R. solanacearum* is also accompanied by the production of soluble phenolics^{42,43}.

Cell wall attachment to plasma membrane (PM) relies on CS domain. The apoplast in between two endodermal cells is sealed by the deposition of CS lignin. This sealing is perfected by the anchoring of the CS membrane domain (CSD) to the CW, through an unknown mechanism. Upon plasmolysis, the protoplasts of endodermal cells retract but the CSD remains tightly attached to the CS^{23,46,47}. This attachment appears in a developmental manner during the differentiation of the endodermis. It occurs concomitantly with the recruitment of the CASPs at the CSD, and with CS lignin deposition²³. We then wanted to study whether or not the different types and sites of lignification contribute to the attachment of the PM, to the CW. To visualize the

PM, we introduced an endodermis-specific PM marker (pELTP::mCit-SYP122) in WT, *sgn3-3*, *myb36-2* and *sgn3-3 myb36-2* backgrounds (Fig. 4a). The PM marker is excluded from the CSD in WT as described for other endodermal PM marker lines^{13,23}. This exclusion is still observed in *sgn3-3* but in an interrupted manner similarly to that observed for lignin (Fig. 1a–c). The exclusion domain disappears entirely in *myb36-2* and *sgn3-3 myb36-2*. Additionally, no exclusion zone in the PM is observed in *myb36-2* where cell-corner lignin is deposited.

We then used mannitol-induced plasmolysis to visualize the PM attachment to the CW. Upon plasmolysis, the PM retracts but remains attached to the CS in WT and *sgn3-3*, forming a flattened protoplast (Fig. 4a). However, small portions of the PM are able to detach from the CW in a *sgn3-3* mutant as seen in Supplementary Fig. 6. This is likely to happen where the PM exclusion domain is interrupted in *sgn3-3* (Fig. 4a). In *myb36-2* and *sgn3-3 myb36-2*, the CW attachment to the PM is lost (Fig. 4a and Supplementary Fig. 6). Importantly, retraction of the PM is observed in *myb36-2* at the corner of the endodermal cells on the cortex side where cell-corner lignin is deposited (Fig. 1a). These results establish the requirement of MYB36 for the formation of the CSD excluding the PM marker. Additionally, the presence of CSD, but not cell-corner lignin, is required for PM attachment to the CW.

We then tested if CS lignin is required for the PM attachment to the CS. For this, we used an inhibitor of the phenylpropanoid pathway, PA. Treatment with PA suppresses lignin accumulation in the vasculature and in the CS (Fig. 4b). Absence of lignin did not affect the exclusion of the PM marker at the CSD. This was further confirmed using the CSD marker line pCASP1::CASP1-GFP¹² that showed similar localization independently of the CS lignin presence (Fig. 4b). Additionally, the PM attachment to the CS is still observed when CS lignin deposition is inhibited (Fig. 4a, b and Supplementary Fig. 6). These findings confirm previous reports^{13,18} showing that CS lignin is not required for the formation of the CSD. Importantly, these results indicate that CS lignin does not participate in anchoring the CSD to the CW. Other CW compounds might be involved in that process.

The absence of PM attachment to the site of cell-corner lignification is likely to affect the permeability of the apoplast of

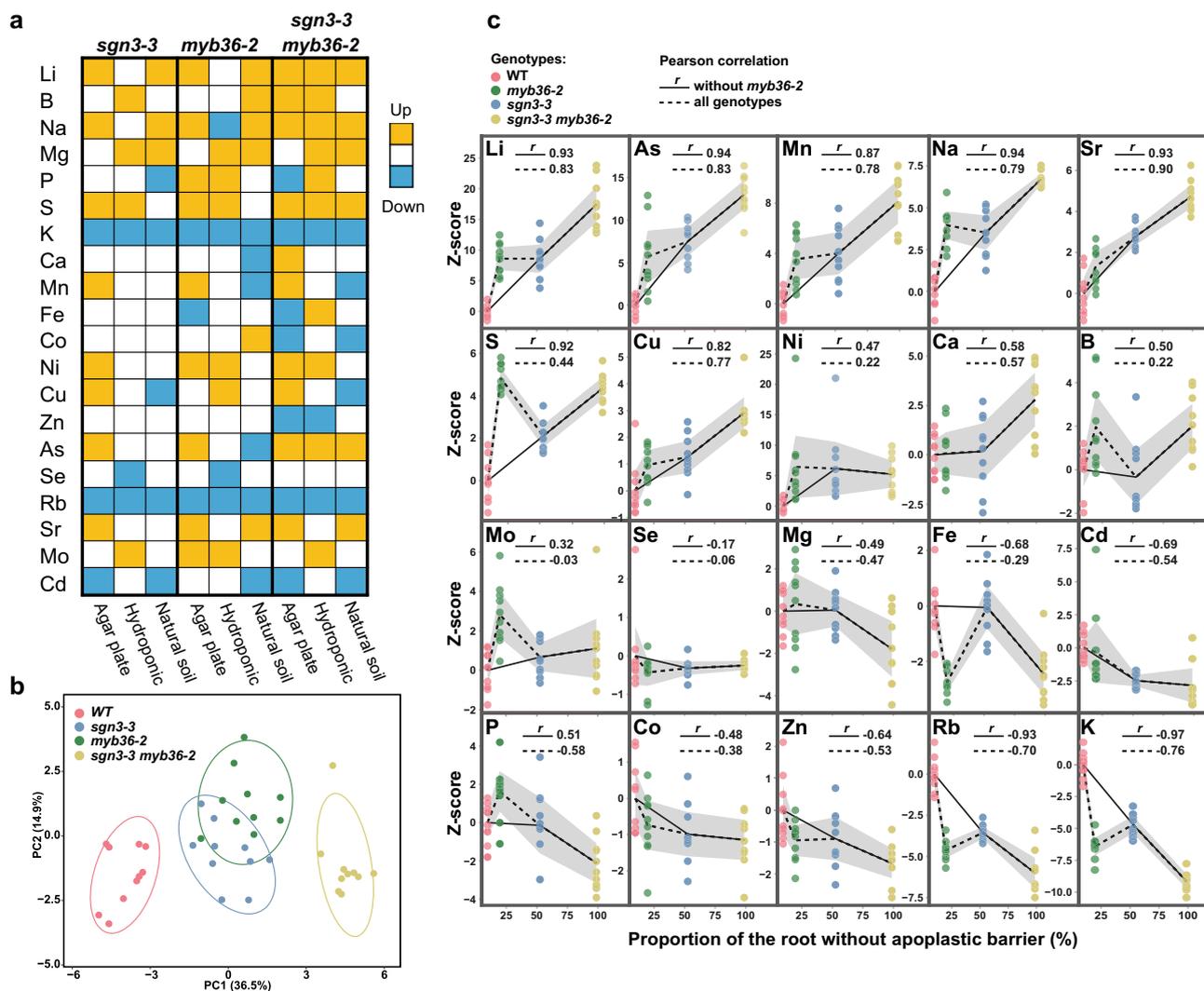


Fig. 5 Absence of endodermal apoplastic barrier triggers major ionic changes. **a** Overview of ions accumulation in shoot of *sgn3-3*, *myb36-2* and *sgn3-3 myb36-2* mutants compared to WT using different growth conditions in agar plates (long day, $n = 10$ individual plants), in hydroponics (short day, $n = 6$ individual plants) and natural soil (short day, $n = 18$ for WT, $n = 18$ for *sgn3-3*, $n = 18$ for *myb36-2* and $n = 13$ for *sgn3-3 myb36-2*). Element concentration was determined by ICP-MS and is available in Supplementary Data 3. Colour code indicates significant changes in accumulation compared with the WT using a two-sided t -test ($p < 0.01$). **b** Principal component analysis (PCA) based on the concentration of 20 elements in shoots of plants grown in agar plates. Ellipses show confidence level at a rate of 90%. $n = 10$ individual plants. **c** Plots presenting the correlation between the z-scores of elements content in shoots of plants grown in agar plates of WT, *myb36-2*, *sgn3-3* and *sgn3-3 myb36-2* against the portion of root length permeable to propidium iodide as determined in Fig. 1d. The black lines show the average and the grey area show the 95% confidence interval ($n = 10$ individual plants).

the endodermal cells. This can consequently affect the transport of water and solutes to the shoot.

Absence of apoplastic barrier triggers major ionic changes.

The mutants described in our previous analyses display differential patterns of lignin deposition and composition, and this consequently affects root apoplastic permeability. This affords a unique opportunity to assess the role of endodermal lignification in controlling nutrient homeostasis in the plant. The *sgn3-3* (delayed CS barrier, no cell-corner lignin), *myb36-2* (no CS lignin, has cell-corner lignin) and *sgn3-3 myb36-2* (no CS or cell-corner lignin) mutants were grown using different growth conditions (agar plate, hydroponic and natural soil) and their leaves were analyzed for their elemental composition (ionome) using inductively coupled plasma-mass spectrometry (ICP-MS) (Fig. 5a and Supplementary Data 3). A principal component analysis (PCA) of the ionome of leaves reveals that all the mutants have different leaf ionomes compared to WT when grown on plates (Fig. 5b),

hydroponically and to a lesser extent in a natural soil (Supplementary Fig. 7a, b). Based on the PC1 axis, *sgn3-3 myb36-2* displayed the most distinct ionic phenotype (Fig. 5b and Supplementary Fig. 7a, b). In line with our previous results (Fig. 1d), this effect indicates an additivity of the two mutations on the leaf ionome. Importantly, this result also supports the idea that cell-corner lignin in *myb36* can act as an apoplastic barrier to mineral nutrients.

We next tested the correlation between the gradient of root apoplastic permeability across WT, *myb36-2*, *sgn3-3* and *sgn3-3 myb36-2* determined in Fig. 1d, with their leaf elemental content (Fig. 5c). We observed that *myb36-2* does not fit into this correlation analysis as well as the other genotypes. This is likely due to activation of the Schengen-pathway leading to deposition of endodermal cell-corner stress lignin, early suberisation, reduced root hydraulic conductivity, activation of ABA signalling in the shoot and stomata closure known to occur in this mutant^{9,48}. Additionally, the *myb36-2* mutation interferes with

overall root development (Supplementary Fig. 7c–e) as previously reported⁴⁹. This is due to the constitutive activation of the Schengen-pathway as *sgn3-3 myb36-2* shows normal root development. Removal of *myb36-2* from the correlation analysis, leaving just lines with an inactive Schengen-pathway, improved the Pearson correlation coefficient for almost all the elements, and we observed a strong correlation ($r \geq 0.5$ or ≤ -0.5) for 15 out of the 20 elements. We observed a strong positive correlation between an increased CS permeability and leaf accumulation of lithium (Li), arsenic (As), manganese (Mn), sodium (Na), strontium (Sr), sulfur (S), copper (Cu), calcium (Ca), boron (B) and a strong negative correlation with iron (Fe), cadmium (Cd), phosphorus (P), zinc (Zn), rubidium (Rb) and potassium (K). This suggests that a functional apoplastic barrier is required to limit the loss of essential elements such as K, Zn, Fe and P. Conversely, a defective apoplastic barrier allows increased leaf accumulation of the essential nutrients Mn, S, Cu, Ca and B. These gradients of higher and lower accumulations of mineral nutrients and trace elements illustrate the bidirectional nature of the CS barrier, by blocking some solutes from entering the vasculature, and by facilitating the accumulation of other solutes in the stele for translocation.

Casparian strips do not control root hydraulic conductivity.

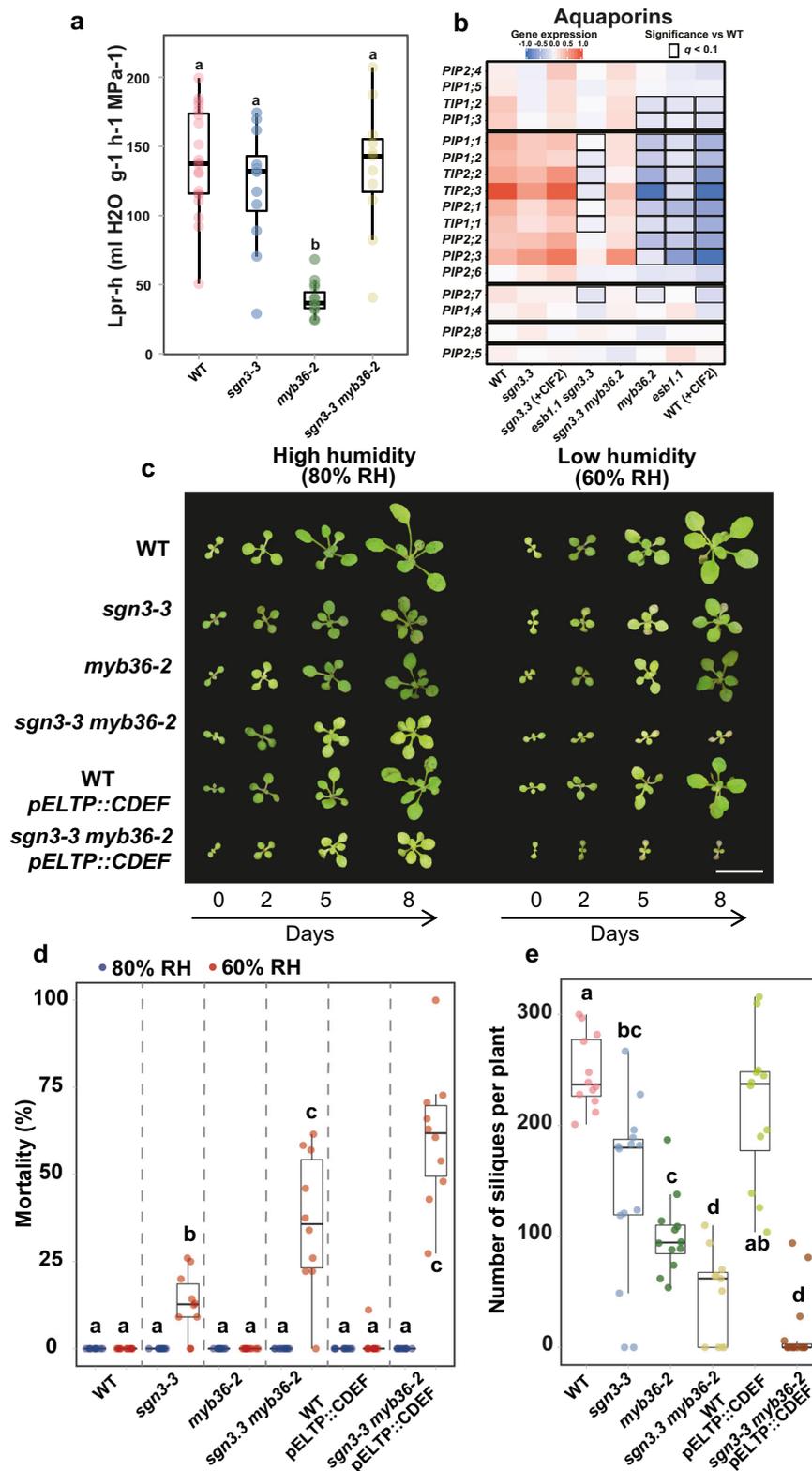
We then measured the capacity of the root to transport water, also called root hydraulic conductivity (Lp_r), in 3-week-old plants grown hydroponically. We observed that the root hydraulic conductivity remains unchanged in *sgn3-3* and *sgn3-3 myb36-2* in comparison with WT (Fig. 6a). In contrast, *myb36-2* showed a strong reduction of root hydraulic conductivity. These results established that the CS-based endodermal apoplastic seal does not control root water transport capacity, as in the absence of any barriers in *sgn3-3 myb36-2* (Fig. 1d) root hydraulic conductivity is the same as WT. This is consistent with water transport occurring mainly via the transcellular pathway, with a major contribution via aquaporins⁵⁰. The reduced hydraulic conductivity observed in *myb36-2* is consistent with that previously observed in *esb1*, which also has an activated Schengen-pathway⁴⁸. The reduced hydraulic conductivity in *esb1* originates mainly from a reduction in aquaporin-mediated water transport as determined using a pharmacological approach⁴⁸. Here, our RNA-seq experiment revealed a GO-term enrichment in cluster C2 (genes repressed by the Schengen-pathway, Fig. 3a) relating to water deprivation (Supplementary Fig. 3b) and importantly, ten aquaporin genes are down regulated by activation of the Schengen-pathway (Fig. 6b). This set of aquaporin genes contains several highly expressed aquaporins in root, including *PIP2,2* known to significantly contribute to root hydraulic conductivity^{51,52}. This provides an explanation for the reduction in root hydraulic conductivity observed in both *myb36* and *esb1*.

Endodermal lignification and plant fitness. Given the significant impacts that CS and Schengen-pathway activation have on mineral nutrient homeostasis (Fig. 5) and water transport (Fig. 6a, b), we further investigated their impact on growth and development. The double mutant *sgn3-3 myb36-2* displayed a severe dwarf phenotype when grown hydroponically or in natural soil but not on agar plates, in comparison with WT and the single mutants (Supplementary Fig. 7c–g). This indicates a critical role of CS for maintaining normal plant growth and development. However, this is conditioned by the growth environment. The high humidity environment and consequently reduced transpiration of plants on agar plates in comparison with the other growth environments could explain these phenotypical differences. Indeed, reduced leaf transpiration is a key part of the

compensation mechanisms mitigating the loss of CS integrity, allowing relatively normal growth as previously reported⁴⁸. We then tested if differences in relative humidity (RH) can affect plant growth in the absence of an endodermal root barrier. For this, we used *sgn3-3* (delayed CS barrier, no cell-corner lignin), *myb36-2* (no CS lignin, has cell-corner lignin) and *sgn3-3 myb36-2* (no CS and no cell-corner lignin). Additionally, we tested if the presence of endodermal suberin can affect plant growth by using lines expressing the Cutinase Destructure Factor (CDEF) under the control of an endodermis-specific promoter (pELTP::CDEF) in a WT and *sgn3-3 myb36-2* background. Lines expressing CDEF show a strong reduction in endodermal suberin deposition (Supplementary Fig. 8a). Seedlings were germinated and grown in soil in a high humidity environment (80% RH) for 7 days, and then transferred to an environment with the same (80% RH) or lower (60% RH) humidity. We measured the leaf surface area as a proxy of plant growth⁵³ at 0, 2, 5 and 8 days after transfer (Fig. 6c and Supplementary Fig. 8b). In both the high and low humidity environment, all mutants with reduced CS functionality (*sgn3-3*, *myb36-2*, *sgn3-3 myb36-2* and *sgn3-3 myb36-2-pELTP::CDEF*) show a reduction of leaf surface in comparison with WT. Importantly, the growth reduction observed in the absence of endodermal lignification (*sgn3-3 myb36-2*) is severe, specifically in low humidity conditions, in comparison with all other genotypes and high humidity conditions. The *sgn3-3 myb36-2* plants with no growth after 9 days started to display necrosis over all the leaf surface and were considered dead as quantified in Fig. 6d. Low humidity triggers a high percentage of mortality in *sgn3-3 myb36-2* and to a lesser extent in *sgn3-3* compared to WT and to the other genotypes in which no mortality is observed when grown in low humidity conditions. Such mortality was not observed when *sgn3-3 myb36-2* was grown in high humidity. This highlights that endodermal lignification is required for maintaining plant growth and survival under low humidity. However, this is not the case for endodermal suberisation, because the removal of suberin through expression of CDEF in WT and *sgn3-3 myb36-2* did not affect mortality or leaf surface area after 8 days at a lower humidity in comparison with their respective backgrounds (Fig. 6c, d and Supplementary Fig. 8b).

The strong growth reduction observed in *sgn3-3 myb36-2* in comparison with WT and the single mutants could be associated with a lack of root selectivity leading to major ionic changes as shown in Fig. 5. Low humidity would generate a higher transpiration stream and consequently more uncontrolled, and potentially detrimental, accumulation of mineral nutrient and trace elements in the leaves, compared with high humidity. Conversely, high humidity would slow transpiration rate, allowing plants to better control nutrient acquisition⁴⁸. To test this, we measured elemental accumulation in leaves of WT, *sgn3-3*, *myb36-2* and *sgn3-3 myb36-2* exposed to 80% RH or 60% RH for 5 days (Supplementary Fig. 8c and Supplementary Data 3). In WT, we observed that 60% RH triggers a decrease in the concentration of Cd, Zn, K, (and chemical analogue Rb), and Ca (and chemical analogue Sr) with a similar trend observed in *myb36-2* for these elements. In contrast, in *sgn3-3 myb36-2*, lower humidity did not trigger such decreases but rather caused an increased accumulation for K (and Rb) and Na (and the chemical analogue Li). This shows that endodermal lignification is essential for the plants to adjust their nutrient balance in response to low humidity. The uncontrolled accumulation of Na could contribute to the growth defect observed in *sgn3-3 myb36-2*.

We then measured the impact of the absence of CS, suberin or of the constitutive activation of the Schengen-pathway on plant fitness. For this, we determined the number of seed-containing siliques per plant as an estimation of fitness (Fig. 6e). A



significant reduction of silique numbers is observed in all the genotypes in comparison to WT, with the exception of *pELTP::CDEF* in the WT background. The *sgn3-3* mutant, with a partial root apoplastic barrier defects, showed a decrease in siliques number in comparison with WT. A similar decrease was observed for *myb36-2* displaying also a partial root apoplastic barrier defect and with cell-corner stress-lignin deposition. Complete disruption of endodermal lignification strongly affects

silique production as observed for *sgn3-3 myb36-2* and *sgn3-3 myb36-2 pELTP::CDEF*. These results clearly establish that the CS is essential for plant fitness. Furthermore, activation of the Schengen-pathway helps protect the plant from the detrimental impact on fitness when the barrier function of the CS is compromised. The *sgn3-3 myb36-2 pELTP::CDEF* line reported here, with its complete lack of endodermal lignin and suberin extracellular barriers, and Schengen-dependent signalling, is a

Fig. 6 Activation of the Schengen-pathway represses water transport and maintains plant growth, survival, and fitness under fluctuating environment.

a Boxplot showing the hydrostatic root hydraulic conductivity ($L_{p,r-h}$) in WT, *sgn3-3*, *myb36-2*, *sgn3-3 myb36-2* grown hydroponically for 19–21 days under environmental controlled conditions. Hydraulic conductivity was measured using pressure chambers ($L_{p,r-h}$). Different letters indicate significant differences between genotypes determined by an ANOVA and Tukey's test as post hoc analyses ($p < 0.01$, $n = 20$ for WT, $n = 12$ for *sgn3-3*, $n = 15$ for *myb36-2* and $n = 11$ for *sgn3-3 myb36-2*). Centre lines show the medians; box limits indicate the 25th and 75th percentiles. The experiment was repeated two times independently with similar results. **b** Heatmap of aquaporins expression across the different genotypes and treatments used in the RNA-seq experiment. **c** Representative pictures of WT, *sgn3-3*, *myb36-2*, *sgn3-3 myb36-2*, WT—*pELTP::CDEF* and *sgn3-3 myb36-2*—*pELTP::CDEF* plants germinated in soil with a high humidity (80%) for 7 days and then transferred in an environment with a lower (60% RH) or with constant humidity (80% RH). Pictures were taken 0, 2, 5 and 8 days after the transfer. Scale bar = 1 cm. The experiment was repeated two times independently with similar results. **d** Boxplots showing the proportion of dead plants after transfer in an environment with constant humidity (80% RH, blue) or with a lower (60% RH, red). The plants displaying no growth after 9 days and showing necrosis in all the leave surface were considered as dead plants. Each point represents the proportion of dead plants in a cultivated pot compared to the total number of plants for one genotype in the same pot. Pots were containing at least eight plants of each genotypes, $n = 10$ pots. Different letters represent significant differences between genotypes using a two-sided Mann-Whitney test ($p < 0.01$). Centre lines show the medians; box limits indicate the 25th and 75th percentiles. **e** Boxplots showing the number of siliques produced per plants. Plants were cultivated in a high humidity environment for 10 days after germination and then transferred to a greenhouse. Each point represents the total number of seeds containing siliques per plant ($n = 12$ for WT, $n = 14$ for *sgn3-3*, $n = 12$ for *myb36-2*, $n = 11$ for *sgn3-3 myb36-2*, $n = 12$ for WT—*pELTP::CDEF*, $n = 15$ for *sgn3-3 myb36-2*—*pELTP::CDEF*). Different letters represent significant differences between genotypes using a two-sided Mann-Whitney test ($p < 0.01$). Centre lines show the medians; box limits indicate the 25th and 75th percentiles.

powerful tool for studying the role of endodermal barriers in a range of processes such as nutrient, hormone and water transport and biotic interaction with soil microorganisms⁵⁴.

The data presented here reveals that the Schengen-pathway is involved in the deposition of two chemically distinct types of lignin. The Schengen-pathway and MYB36 are required for the deposition of CS lignin. Constitutive activation of the Schengen-pathway leads to the deposition of a chemically distinct stress-like type of lignin. This deposition of stress-lignin contributes to sealing the apoplast and maintaining ion homeostasis in the absence of CS integrity. However, no PM attachment to the CW is observed at the site of stress-lignin deposition as seen for the CS, suggesting an inferior seal is formed.

Methods

Plant material. *A. thaliana* accession Columbia-0 (Col-0) was used for this study. The following mutants and transgenic lines were used in this study: *sgn3-3* (SALK_043282)⁹, *myb36-2* (GK-543B11)¹⁷, *pCASPI::CASPI-GFP*¹², *ahp6-125*, *esb1-1*¹⁰, *pELTP::CDEF*⁵⁵, *pELTP::SYP122-mCitrine*.

The corresponding gene numbers are: *SGN3*, At4g20140; *MYB36*, At5g57620; *CASPI*, At2g36100; *AHP6*, At1g80100; *ESB1*, At2g28670; *ELTP*, At2g48140; *CDEF*, At4g30140; *SYP122*, At3g52400.

Generation of transgenic lines. The *pELTP::mCit-SYP122* construct was obtained by recombining three previously generated entry clones for *pELTP*⁵⁶, *mCITRINE* and *SYP122* cDNA⁵⁷ using LR clonease II (Invitrogen). This construct was independently transformed into WT, *sgn3-3*, *myb36-2* or *sgn3-3 myb36-2* using the floral dip method⁵⁸. The construct *pELTP::CDEF*⁵⁵ was independently transformed into WT and *sgn3-3 myb36-2*.

Growth conditions. For agar plates assays, seeds were surface sterilized, sown on plates containing ½ Murashige and Skoog (MS) pH 5.8 with 0.8% agar, stratified for 2 days at 4 °C and grown vertically in growth chamber under long day condition (16 h light 100 μE 22 °C/8 h dark 19 °C) and observed after 6 days. The CIF2 peptide treatment (DY(SO3H)GHSSPKPLVRPPFKLIPN) was applied from germination at a concentration of 100 nM. The CIF2 peptide was synthesized by Cambridge Peptid Ltd.

For ionic analysis, plants were grown using three growth conditions:

- Sterile ½ MS agar plate. Seeds were surface sterilized and sown on plates containing ½ MS with 0.8% agar, stratified for 2 days at 4 °C and grown vertically in growth chamber under long day condition (16 h light 100 μE 22 °C/8 h dark 19 °C). Shoots were collected 2 weeks after germination.
- Hydroponic. Plants were grown for 5 weeks under short day condition (8 h light 100 μE 21 °C/16 h dark 18 °C) at 20 °C with a RH of 65% RH in a media at pH 5.7 containing 250 μM CaCl₂, 1 mM KH₂PO₄, 50 μM KCl, 250 μM K₂SO₄, 1 mM MgSO₄, 100 μM NaFe-EDTA, 2 mM NH₄NO₃, 30 μM H₃BO₃, 5 μM MnSO₄, 1 μM ZnSO₄, 1 μM CuSO₄, 0.7 μM NaMoO₄, 1 μM NiSO₄. Media was changed weekly.
- Natural soil. Plants were grown for 9 weeks in a growth chamber under short day condition (8 h light 100 μE 19 °C/16 h dark 17 °C) at 18 °C with a RH of

70% in a soil collected in the Sutton Bonington campus of the university of Nottingham (GPS coordinate: 52°49'59.7"N 1°14'56.2"W).

Fluorescence microscopy. To determine the functionality of the endodermal apoplastic barriers we used the apoplastic tracer PI (Invitrogen). PI is a fluorescent molecule which diffusion into the tissue layers of the root is blocked only after the dye reaches the differentiated endodermis²³. Six-day-old Col-0 seedlings were incubated in a fresh solution of 10 μg/mL PI (prepared from a stock solution 1 mg/mL) for 10 min in the dark and then rinsed twice with water. Seedlings were carefully placed on a microscope slide with water and covered with a coverslip. Using a fluorescence microscope Leica CTR5000, ×20 magnification, we quantified the number of cells from the onset of elongation until the endodermal cells blocked the PI penetration to the stele.

For lignin staining with basic fuchsin, *CASPI-GFP* visualization and Calcofluor white M2R staining, 6-day-old roots were fixed in 4% paraformaldehyde and then washed twice for 1 min with 1x PBS and transferred in ClearSee solution⁵⁹ (10% xylitol, 15% Sodium deoxycholate, 25% urea) for 24 h. Then, the seedlings were stained overnight in 0.2% Basic Fuchsin in ClearSee for lignin staining. Basic Fuchsin solution was removed and the seedlings were washed three times for 120 min with ClearSee with gentle shaking. Then, the seedlings were stained for 1 h in 0.1% Calcofluor white M2R in ClearSee for CW staining. Calcofluor white M2R solution was removed and the seedlings were washed three times for 30 min with ClearSee with gentle shaking. Roots were carefully placed on a microscope slide with ClearSee and covered with a coverslip. The roots were then observed with confocal microscopes (Zeiss LSM500 and Leica SP8) using their respective software (Zeiss ZEN 2.3 and Leica LAS X). We used an excitation at 594 nm and an emission band path of 600–650 nm for Basic Fuchsin, 405 nm and an emission band path of 425–475 nm for Calcofluor white M2R and 488 nm and an emission band path of 505–530 nm for GFP.

Suberin staining was performed using Fluorol yellow 088^{7,56}. Seedlings were incubated in a freshly prepared solution of Fluorol Yellow 088 (0.01% w/v, in lactic acid) at 70 °C for 30 min, followed by two rinses with water. Then, samples were treated with aniline blue (0.5% w/v, in water) at room temperature for 30 min in darkness. After two rinses with water, several roots were placed on microscope slides with water, covered with a coverslip and the suberin deposition pattern was quantified using an epifluorescence microscope Leica CTR5000, ×20 magnification with a GFP filter. The suberin deposition pattern was quantified as the number of cells in the continuous, patchy and no-suberin zones.

Plasmolysis. Plasmolysis was induced by mounting 6-day-old seedlings in 0.8 M mannitol on microscope slides and directly observed using confocal microscopy (Leica SP8). PA was used to inhibit lignin biosynthesis⁷. PA at a concentration of 10 μM was added to the media from germination. The proportion of the CW length in direct contact with the PM marker *SYP122-mCitrine* after plasmolysis was measured using Fiji after plasmolysis. This measurement was done on a maximum projection of the top endodermal cells as seen on Supplementary Fig. 6. The quantification represents the percentage of CW length in direct contact with the PM marker *SYP122-mCitrine* after plasmolysis. Plasmolysis events were imaged and quantified at 15 cells after the onset of elongation.

For the observation of *CASPI-GFP* and Lignin staining with basic fuchsin, the seedlings were incubated in 0.8 M mannitol for 5 min, and then fixed and cleared as described above.

Thioacidolysis. The plants were grown for 6 days on ½ MS plates supplemented with 10 nM 6-Benzylaminopurine and 0.1 % sucrose. Seeds were sown in three parallel lines per square plates (12 × 12 cm) at high density. Six plates were combined to obtain one replicate. The first 3 mm of root tips as this zone contains no xylem pole were collected in order to obtain 7–15 mg of dry weight. The samples were washed twice with 1 mL methanol, rotated for 30 min on a carousel and centrifuged to eliminate the methanol supernatant. This washing step was repeated once and the final methanol-extracted samples were then dried for 2 days at 40 °C (oven) before thioacidolysis.

The thioacidolyses were carried out in a glass tube with a Teflon-lined screwcap, from about 5 mg sample (weighted at the nearest 0.01 mg) put together with 0.01 mg C21 and 0.01 mg C19 internal standard (50 µL of a 0.2 mg/ml solution) and with 2 ml freshly prepared thioacidolysis reagent. The tightly closed tubes were then heated at 100 °C for 4 h (oil bath), with gentle occasional shaking. After cooling and in each tube, 2 ml of aqueous NaHCO₃ 0.2 M solution were added (to destroy the excess of BF₃) and then 0.1 ml HCl 6 M (to ensure that the pH is acidic before extraction). The reaction medium was extracted with 2 ml methylene chloride (in the tube) and the lower phase was collected (Pasteur pipette) and dried over Na₂SO₄ before evaporation of the solvent (rotoevaporator). The final sample was redissolved in about 2 mL of methylene chloride and 15 µL of this solution were trimethylsilylated (TMS) with 50 µL BSTFA + 5 µL pyridine. The TMS solution was injected (1 µL) into a GC-MS Varian 4000 instrument fitted with an Agilent combiPAL autosampler, a splitless injector (280 °C), and an ion-trap mass spectrometer (electron impact mode, 70 eV), with a source at 220 °C, a transfer line at 280 °C and an *m/z* 50–800 scanning range. The GC column was a Supelco SPB1 column (30 m × 0.25 mm i.d., film thickness 0.25 µm) working in the temperature programme mode from 45 to 160 °C at +30 °C/min and then 160–260 °C at +5 °C/min, with helium as the carrier gas (1 mL/min). The GC-MS determinations of the H, G and S lignin-derived monomers were carried out on ion chromatograms, respectively, reconstructed at *m/z* 239, 269 and 299, as compared to the internal standard hydrocarbon evaluated on the ion chromatogram reconstructed at *m/z* (57 + 71 + 85). Each genotype was analyzed as biological triplicates and each biological triplicate was subjected to two different silylations and GC-MS analyses.

Raman microscopy. Six-day-old seedling was fixed in PBS buffer containing 4% formaldehyde and 1% glutaraldehyde at 4 °C overnight, then washed twice with PBS 30 min. Samples were progressively dehydrated in ethanol (30, 50, 70, 100% ethanol). Samples were aligned and embedded in Leica historesin using the protocol described in⁶⁰. Sections of 5 µm at 4 mm from the root tip were generated using a Leica microtome.

The different samples were embedded in resin and cut at 4 mm from the root tip using a microtome with a thickness of 15 µm. Sections were mounted on superfrost glass slides. The samples were then mapped in a grid over the region of interest. The Raman imaging was performed with a Horiba LabRAM HR spectroscopy equipped with a piezoelectric scan stage (Horiba Scientific, UK) using a 532 nm laser, a ×100 air objective (Nikon, NA = 0.9) and 600 g mm⁻¹ grating. Imaging was performed using the LabSpec 6.4.3 Spectroscopy Suite Software. Maps were collected for the regions of interest by setting equidistant points along the sample to ensure maximum coverage. The main regions covered in the analysis were the endodermal cell–cell junction, the endodermal cell corners towards the endodermal–cortical junction and the xylem poles (Supplementary Fig. 1). The maps were acquired with two accumulations and 30 s integration time. The spectra were acquired in the range 300–3100 cm⁻¹. The spectra were processed using MATLAB and eigenvector software. First, the spectra were trimmed (500–1800 cm⁻¹), smoothed and then baseline corrected using an automatic least squares algorithm. This was followed by a percentile mean subtraction (10–20%) to remove signal from the resin. Finally, Gaussian image smoothing was performed to improve signal to noise of the Raman maps. MCR analysis was performed on the maps containing the specific ROIs and the corresponding lignin spectra were extracted. These lignin spectra were then used as bounds for the MCR analysis of the large maps where the concentration of these spectra is determined, with a high intensity indicating a high concentration of the specific lignin (Fig. 2c).

Monolignol feeding experiment. The plants were grown vertically for 4 days on control condition (½ MS, 0.8% agar) in growth chamber under long day condition (16 h light 100 µE 22 °C/8 h dark 19 °C). Seedlings were then transferred for 3 more days on the same condition on the same media supplemented or not with 10 µM PA (Sigma-Aldrich, P49805), 10 µM PA with 20 µM *p*-coumaryl alcohol (Sigma-Aldrich, PHL82506), 10 µM PA with 20 µM coniferyl alcohol (Sigma-Aldrich, 223735), 10 µM PA with 20 µM sinapyl alcohol (Sigma-Aldrich, 404586) or 10 µM PA with 10 µM *p*-coumaryl alcohol and coniferyl alcohol. Apoplastic permeability was determined using PI and by quantifying the number of cells from the onset of elongation until the endodermal cells blocked the PI penetration to the stele.

For lignin staining, the roots of *myb36-2* were fixed with 4% PFA for 120 min at 20 °C. Seedlings were then washed twice for 1 min with 1x PBS and transferred to the ClearSee solution⁵⁹ (10% xylitol, 15% Sodium deoxycholate, 25% urea) for 24 h. Then, the seedlings were stained overnight in 0.2% Basic Fuchsin in ClearSee for lignin staining. Basic Fuchsin solution was removed and the seedlings were washed three times for 120 min with ClearSee with gentle shaking. Roots were carefully placed on a microscope slide with ClearSee and covered with a coverslip. We

imaged the root at 20 cells after the onset of elongation using a confocal microscope Leica SP8, ×63 objective (NA = 1.2) by performing a z-stack on the top endodermal cells. We used an excitation at 594 nm and an emission band path of 600–650 nm for Basic Fuchsin. For the quantification of pixel intensity of basic fuchsin fluorescence, we performed a maximum intensity projection of the top endodermal cells. Using Fiji, we traced a 1 µm thick segmented line following the cortex-facing CW corner of one endodermal cell. The mean pixel intensity was determined along that line. This represents the pixel intensity for one endodermal cell. Pixel intensity was measured in three to six individual cells per plant. A total of four to six individual plants were measured for each treatment. Pixel intensities were plotted using the SuperPlots tool⁶¹.

RNA-seq. The plants were grown for 6 days on ½ M/S plates. Seeds were sown in three parallel lines per square plates (12 × 12 cm) at high density. The first 5 mm of root tips were collected. One plate was used as a biological replicate. The samples were snap-frozen at harvest and ground into fine powder in a 2 mL centrifuge tube. Total RNA was extracted according to ref.⁶². Samples were homogenized in 400 µL of Z6-buffer containing 8 M guanidine-HCl, 20 mM MES, 20 mM EDTA pH 7.0. After the addition of 400 µL phenol:chloroform:isoamyl alcohol, 25:24:1, samples were vortexed and centrifuged (15,000 × g 10 min, 4 °C) for phase separation. The aqueous phase was transferred to a new 1.5 mL tube and 0.05 volumes of 1 N acetic acid and 0.7 volumes 96% ethanol was added. The RNA was precipitated at -20 °C overnight. Following centrifugation (15,000 × g 10 min, 4 °C), the pellet was washed with 200 µL 3 M sodium acetate at pH 5.2 and 70% ethanol. The RNA was dried and dissolved in 30 µL of ultrapure water and store at -80 °C until use. DNase treatment (DNase I, Amplification Grade, 18068015, Invitrogen) was carried out on the samples to remove genomic DNA. The RNA Concentration and quality were determined using Qubit (Invitrogen; Q10210) and TapeStation (Agilent; G2991A) protocols. Libraries were generated using the Lexogen Quant Seq 3' mRNA Seq (FWD) Library Prep Kit (Lexogen; 015) which employs polyA selection to enrich for mRNA. Library yield was measured by Qubit (Invitrogen; Q10210) and TapeStation (Agilent; G2991A) systems using protocols to determine concentration and library size, these are then pooled together in equimolar concentrations. The concentration of the pool of libraries was confirmed using the Qubit and qPCR and then loaded onto an Illumina NextSeq 500/550 High Output Kit v2.5 (75 Cycles) (Illumina; 20024906), to generate ~5 million 75 bp single-end reads per sample.

BBduk v38.82 was used to identify and discard reads containing the Illumina adaptor sequence⁶³. Then, we mapped the resulting high-quality filtered reads against the TAIR10 Arabidopsis reference genome (Ensembl Plants v48) using STAR 2.7.5c⁶⁴ with the following parameters:

- outFilterType BySJout
- outFilterMultimapNmax 20
- alignSJoverhangMin 8
- alignSJDBoverhangMin 1
- sjdbGTFtagExonParentGene gene_id
- sjdbGTFfile Arabidopsis_thaliana.TAIR10.48.gtf
- outReadsUnmapped Fastx
- quantMode GeneCounts
- outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.1 --alignIntronMin 20
- alignIntronMax 1000000 --alignMatesGapMax 1000000 --outSAMattributes NH HI NM MD
- outSAMtype BAM SortedByCoordinate
- outFilePrefix <outfile>

We used the R package DESeq2 v.1.24.0 to identify DEGs between each genotype, *sgn3-3*, *sgn3-3 myb36-2*, *sgn3-3 (+CIF2)*, *esb1-1*, *esb1-1 sgn3-3* and WT (+CIF2) against WT (Col-0). To do so, we fitted the following generalized linear model:

$$\text{Gene abundance} \sim \text{Rep} + \text{Genotype}$$

A gene was considered statistically differentially expressed if it had a false discovery rate (FDR) adjusted *p* value < 0.05.

For visualization purposes, we created a standardized gene matrix. To do so, we applied a variance stabilizing transformation to the raw count gene matrix followed up by standardizing the expression of each gene along the samples. We used this standardized gene matrix to perform principal coordinate (PC) analysis using the *prcomp* function in R. We displayed the results of the PC analysis using *ggplot2*.

Additionally, we subset the 3564 statistically significant DEGs from the standardized gene matrix. Then, for each DEG we calculated its mean expression across each genotype followed up by hierarchical clustering (R function *hclust* method *ward.D2*) using the euclidean distance for the genotypes and the correlation dissimilarity for the genes. To define the seven clusters of cohesively expressed genes, we cut the gene dendrogram from the hierarchical clustering using the R function *cutree*. We visualized the expression of the 3564 DEGs and the result of the clustering approach using *ggplot2*. We used the *compareCluster* function from the *clusterProfiler* R package to perform gene ontology (GO) analysis for the seven clusters of cohesively expressed DEGs.

We constructed individual heatmaps for the phenylpropanoid pathway, the peroxidases, the laccases and the aquaporin genes by sub setting the corresponding curated gene ids from the standardized gene matrix and procedure described above.

Raw sequence data and read counts are available at the NCBI Gene Expression Omnibus accession number (GEO: [GSE158809](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE158809)). Additionally, the scripts created to analyze the RNA-Seq data can be found at <https://github.com/isaig/schengenlignin> with an assigned DOI (<https://doi.org/10.5281/zenodo.4588023>).

Split-root experiment. Wild-type plants were grown 8 days vertically on ½ M/S square plates (12 × 12 cm) in a growth chamber under long day condition (16 h light 100 µE 22 °C/8 h dark 19 °C). The primary root was then cut below the first two lateral roots and plants were allowed to grow on the same plate for an additional 4 days. Then, plants were transferred on to a round plate (Ø: 10 cm) containing two compartments generated by a dividing partition. The root system of each plant was grouped into two halves and placed on either side of the partition. Two plants per plate were transferred. The compartments were filled with either solid ½ M/S on both compartments, solid ½ M/S supplemented with 100 nM CIF2 in both compartments or solid ½ M/S on one side and solid ½ M/S supplemented with 100 nM CIF2 on the other side. Plants were then grown for an additional 3 days. The roots from plants presenting balanced roots on both sides of the divide were harvested for lignin staining and RNA extraction. For RNA extraction, roots from four to six plates were combined to obtain one replicate and stored in liquid nitrogen. RNA extraction was performed as described in the “RNA-seq” section. 1 µg of total RNA treated with DNase I (Thermo Scientific) was used for reverse transcription (RevertAid First Strand cDNA synthesis kit; Thermo Scientific) with oligo(dT)18. cDNA was diluted twice with water, and 1 µL of each sample was assayed three times by qRT-PCR in a LightCycler 480 (Roche) using LC480-SYBR-Green master I (Roche). Expression levels were calculated relatively to the gene *At4g24550*⁶⁵ using the comparative threshold cycle method. The list of genes related to defence was determined by selecting the genes from the cluster C1 (Fig. 3) displaying an average log2 fold change in *myb36-2*, *esb1-1* and WT (+CIF2) higher than 2.9 and belonging to the following using the Gene Ontology annotations: defence response, defence response to bacterium, defence response to fungus, immune system process, innate immune response, response to chitin, response to fungus and immune response. All primer sets are indicated in Supplementary Data 4.

For lignin staining, roots were fixed in paraformaldehyde and cleared in ClearSee solution⁵⁹ for 24 h. Then, roots were stained overnight in 0.2% Basic Fuchsin in ClearSee for lignin staining. Basic Fuchsin solution was removed and replaced by 0.1% direct yellow 96 (CW) in ClearSee for 1 h. The roots were washed three times for 120 min with ClearSee with gentle shaking. Roots were carefully placed on a microscope slide with ClearSee and covered with a coverslip. We imaged the root at 15 cells after the onset of elongation using a confocal microscope Leica SP8, ×63 objective (NA = 1.2). We used an excitation at 594 nm and an emission band path of 600–650 nm for Basic Fuchsin and 488 nm and an emission band path of 500–540 nm for Direct yellow 96.

Extraction and profiling of metabolites. The plants (WT, *esb1-1*, *sgn3-3* and *esb1-1 sgn3-3*) were grown for 6 days on ½ M/S plates supplemented with 0.1% sucrose. Seeds were sown in three parallel lines per square plates (12 × 12 cm) at high density. The first 5 mm of root tips were collected in order to obtain 10–20 mg of dry weight per replicate. Eight plates were combined to obtain one replicate. Eight replicates per genotypes were harvested. The samples were snap-frozen at harvest and ground into fine powder in a 2 mL centrifuge tube then homogenized in liquid nitrogen and extracted with 1 mL methanol. The methanol extract was then evaporated, and the pellet dissolved in 200 µL water/cyclohexane (1/1, v/v). 10 µL of the aqueous phase was analyzed via reverse phase UHPLC (Acquity UPLC Class 1 systems consisting of a Sample Manager-FTN, a Binary Solvent Manager and a Column Manager, Waters Corporation, Milford, MA) coupled to negative ion ElectroSpray Ionization-Quadrupole-Time-of-Flight Mass Spectrometry (Vion IMS QToF, Waters Corporation) using an Acquity UPLC BEH C18 column (1.7 µm, 2.1 × 150 mm; Waters Corporation). Using a flow rate of 350 µL/min and a column temperature of 40 °C, a linear gradient was run from 99% aqueous formic acid (0.1%, buffer A) to 50% acetonitrile (0.1% formic acid, buffer B) in 30 min, followed by a further increase to 70% and then to 100% buffer B in 5 and 2 min, respectively. Full MS spectra (*m/z* 50–1500) were recorded at a scan rate of 10 Hz. The following ESI parameters were used: capillary voltage 2.5 kV, desolvation temperature 550 °C, source temperature 120 °C, desolvation gas 800 L/h and cone gas 50 L/h. Lock correction was applied. In addition to full MS analysis, a pooled sample was subjected to data-dependent MS/MS analysis (DDA) using the same separation conditions as above. DDA was performed between *m/z* 50 and 1200 at a scan rate of 5 Hz and MS → MS/MS transition collision energy of 6 eV. The collision energy was ramped from 15 to 35 eV and from 35 to 70 eV for the low and high mass precursor ions, respectively.

Integration and alignment of the *m/z* features were performed via Progenesis QI software version 2.1 (Waters Corporation). The raw data were imported in this software using a filter strength of 1. A reference chromatogram was manually chosen for the alignment procedure and additional vectors were added in chromatogram regions that were not well aligned. Peak picking was based on all

runs with a sensitivity set on “automatic” (value = 5). The normalization was set on “external standards” and was based on the dry weight of the samples⁶⁶. In total, 13,091 *m/z* features were integrated and aligned across all chromatograms. Structural annotation was performed using a retention time window of 1 min, and using both precursor ion and MS/MS identity searches. The precursor ion search (10 ppm tolerance) was based on a compound database constructed via instant JChem (ChemAxon, Budapest, Hungary), whereas MS/MS identities were obtained by matching against an in-house mass spectral database (200 ppm fragment tolerance).

Using R vs 3.4.2., *m/z* features representing the same compound were grouped following the algorithm in⁶⁷. Of the 13,091 *m/z* features, 12,326 were combined into 2482 *m/z* feature groups, whereas 765 remained as *m/z* feature singlets (i.e., low abundant features). All statistical analyses were performed in R vs. 3.4.2⁶⁸. Including all *m/z* features and upon applying a prior inverted hyperbolic sine transformation⁶⁹, the data were analyzed via both PCA and one-way analysis of variance (ANOVA; *lm()* function) followed by Tukey Honestly Significant Difference (TukeyHSD() function) post hoc tests. For PCA, the R packages FactoMineR⁷⁰ and factoextra (<https://CRAN.R-project.org/package=factoextra>) were employed: *PCA(scale.unit=T,graph=F)*, *fviz_pca_ind()* and *fviz_pca_biplot()*. Following ANOVA analysis, experiment-wide significant models were revealed via a FDR correction using the *p.adjust(method = “fdr”)* function. Using a FDR-based *Q* value < 0.05, 4244 of the 13,091 *m/z* features were significantly changed in abundance corresponding to 123 *m/z* feature singlets and 1158 of the 2482 compounds. Using a minimum abundance threshold of 500 in at least one of the lines, further analysis was performed on 411 of the 1158 compounds and 11 of the 123 *m/z* feature singlets (411 compounds and 11 singlets representing together 889 *m/z* features).

Root hydraulic conductivity. The procedure was exactly identical to the one described in⁴⁸. Root hydrostatic conductance (*K_r*) was determined in freshly detopped roots using a set of pressure chambers filled with hydroponic culture medium. Excised roots were sealed using dental paste (Coltène/Whaledent s.a.r.l., France) and were subjected to 350 kPa for 10 min to achieve flow stabilization, followed by successive measurements of the flow from the hypocotyl at pressures 320, 160 and 240 kPa. Root hydrostatic conductance (*K_r*) was calculated by the slope of the flow (*J_v*) to pressure relationship. The hydrostatic water conductivity (*L_{p,r}*, ml H₂O g⁻¹ h⁻¹ MPa⁻¹) was calculated by dividing *K_r* by the root dry weight.

Humidity, leaf surface, mortality, ionome and fitness. For the determination of the leaf surface and mortality, the seeds were stratified for 2 days at 4 °C and the plants were grown in Levington M3 compost in a growth chamber under long day condition (16 h light 100 µE 21 °C/8 h dark 19 °C). The plants were grown for 7 days with high RH (80% RH) and then half of the plants were transferred at a lower humidity (60% RH). Leaf surface was determined at 6, 9, 12 and 15 days after germination using the threshold command of the Fiji software. The plants displaying no growth after 9 days and showing necrosis in all the leaf surface were considered as dead plants.

For the determination of the shoot ionome, the plants were cultivated at 80% RH for 10 days and then transferred at 80% RH or 60% RH for 5 additional days. Shoots were harvested for ionomic analysis.

For the determination of the siliques number, the plants were cultivated in a high humidity environment for 10 days after germination and then transferred to a greenhouse. After siliques ripening, only the seeds containing siliques were counted.

Ionomic analysis with ICP-MS. Ionomics analysis of plants grown in soil (or on plate, hydroponically) was performed as described⁷¹. Briefly, samples (shoot) were harvested into Pyrex test tubes (16 × 100 mm) and dried at 88 °C for 20 h. After weighing the appropriate number of samples (these masses were used to calculate the rest of the sample masses; alternatively, all samples were weighed individually—usually for small set of samples), the trace metal grade nitric acid Primar Plus (Fisher Chemicals) spiked with indium internal standard was added to the tubes (1 mL per tube). The samples were then digested in dry block heater (DigiPREP MS, SCP Science; QMX Laboratories, Essex, UK) at 115 °C for 4 h. The digested samples were diluted to 10 mL with 18.2 MΩcm Milli-Q Direct water (Merck Millipore). Elemental analysis was performed with an ICP-MS, PerkinElmer NexION 2000 equipped with Elemental Scientific Inc. autosampler, in the collision mode (He) and Syngistix software. Twenty elements (Li, B, Na, Mg, P, S, K, Ca, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Rb, Sr, Mo and Cd) were monitored. Liquid reference material composed of pooled samples was prepared before the beginning of sample run and was used throughout the whole samples run. It was run after every ninth sample to correct for variation within ICP-MS analysis run⁷¹. The calibration standards (with indium internal standard and blanks) were prepared from single element standards (Inorganic Ventures; Essex Scientific Laboratory Supplies Ltd, Essex, UK) solutions. Sample concentrations were calculated using external calibration method within the instrument software. Further data processing was performed using Microsoft Excel spreadsheet.

Data availability

The data that support the findings of this study are available within the paper and its Supplementary Information or are available from the corresponding author upon reasonable request. The source data underlying Figs. 3 and 6b, Supplementary Fig. 3 and Supplementary data 1 are available at the NCBI Gene Expression Omnibus accession number (GEO: GSE158809). Source data are provided with this paper.

Code availability

All scripts used for the RNA-seq analysis are available on GitHub (<https://github.com/isaigs/schengenlignin>) with an assigned DOI (<https://doi.org/10.5281/zenodo.4588023>).

Received: 12 October 2020; Accepted: 11 March 2021;

Published online: 19 April 2021

References

- Boerjan, W., Ralph, J. & Baucher, M. Lignin biosynthesis. *Annu. Rev. Plant Biol.* **54**, 519–546 (2003).
- Lee, M. H. et al. Lignin-based barrier restricts pathogens to the infection site and confers resistance in plants. *EMBO J.* **38**, e1745–17 (2019).
- Vanhohle, R., De Meester, B., Ralph, J. & Boerjan, W. Lignin biosynthesis and its integration into metabolism. *Curr. Opin. Biotechnol.* **56**, 230–239 (2019).
- Tobimatsu, Y. & Schuetz, M. Lignin polymerization: how do plants manage the chemistry so well? *Curr. Opin. Biotechnol.* **56**, 75–81 (2018).
- Schuetz, M. et al. Laccases direct lignification in the discrete secondary cell wall domains of protoxylem. *Plant Physiol.* **166**, 798–807 (2014).
- Barros, J., Serk, H., Granlund, I. & Pesquet, E. The cell biology of lignification in higher plants. *Ann. Bot.* mcv046 <https://doi.org/10.1093/aob/mcv046> (2015).
- Naseer, S. et al. Casparian strip diffusion barrier in Arabidopsis is made of a lignin polymer without suberin. *Proc. Natl Acad. Sci. USA* **109**, 10101–10106 (2012).
- Geldner, N. The endodermis. *Annu. Rev. Plant Biol.* **64**, 531–558 (2013).
- Pfister, A. et al. A receptor-like kinase mutant with absent endodermal diffusion barrier displays selective nutrient homeostasis defects. *Elife* **3**, e03115 (2014).
- Baxter, I. et al. Biodiversity of mineral nutrient and trace element accumulation in *Arabidopsis thaliana*. *PLoS ONE* **7**, e35121 (2012).
- Barbosa, I. C. R., Rojas-Murcia, N. & Geldner, N. The Casparian strip—one ring to bring cell biology to lignification? *Curr. Opin. Biotechnol.* **56**, 121–129 (2018).
- Roppolo, D. et al. A novel protein family mediates Casparian strip formation in the endodermis. *Nature* **473**, 380–383 (2011).
- Lee, Y., Rubio, M. C., Alassimone, J. & Geldner, N. A Mechanism for Localized Lignin Deposition in the Endodermis. *Cell* **153**, 402–412 (2013).
- Rojas-Murcia, N. et al. High-order mutants reveal an essential requirement for peroxidases but not laccases in Casparian strip lignification. *Proc. Natl Acad. Sci. USA* **117**, 29166–29177 (2020).
- Hosmani, P. S. et al. Dirigent domain-containing protein is part of the machinery required for formation of the lignin-based Casparian strip in the root. *Proc. Natl Acad. Sci. USA* **110**, 14498–14503 (2013).
- Liberman, L. M., Sparks, E. E., Moreno-Risueno, M. A., Petricka, J. J. & Benfey, P. N. MYB36 regulates the transition from proliferation to differentiation in the Arabidopsis root. *Proc. Natl Acad. Sci. USA* 201515576 <https://doi.org/10.1073/pnas.1515576112> (2015).
- Kamiya, T. et al. The MYB36 transcription factor orchestrates Casparian strip formation. *Proc. Natl Acad. Sci. USA* **112**, 10533–10538 (2015).
- Fujita, S. et al. SCHENGEN receptor module drives localized ROS production and lignification in plant roots. *EMBO J.* **39**, e103894 (2020).
- Nakayama, T. et al. A peptide hormone required for Casparian strip diffusion barrier formation in Arabidopsis roots. *Science* **355**, 284–286 (2017).
- Doblas, V. G. et al. Root diffusion barrier control by a vasculature-derived peptide binding to the SGN3 receptor. *Science* **355**, 280–284 (2017).
- Alassimone, J. et al. Polarly localized kinase SGN1 is required for Casparian strip integrity and positioning. *Nat. Plants* **2**, 16113 (2016).
- Li, B. et al. Role of LOTR1 in nutrient transport through organization of spatial distribution of root endodermal barriers. *Curr. Biol.* 1–9 <https://doi.org/10.1016/j.cub.2017.01.030> (2017).
- Alassimone, J., Naseer, S. & Geldner, N. A developmental framework for endodermal differentiation and polarity. *Proc. Natl Acad. Sci. USA* **107**, 5214–5219 (2010).
- Agarwal, U. P., McSweeney, J. D. & Ralph, S. A. FT-Raman investigation of milled-wood lignins: softwood, hardwood, and chemically modified black spruce lignins. *J. Wood Chem. Technol.* **31**, 324–344 (2011).
- Mahonen, A. P. Cytokinin signaling and its inhibitor AHP6 regulate cell fate during vascular development. *Science* **311**, 94–98 (2006).
- Lange, B. M., Lapierre, C. & Sandermann, H. Jr. Elicitor-induced spruce stress lignin (structural similarity to early developmental lignins). *Plant Physiol.* **108**, 1277 (1995).
- Fukushima, K. & Terashima, N. Heterogeneity in formation of lignin. *Wood Sci. Technol.* **25**, 371–381 (1991).
- Westermarck, U. The occurrence of p-hydroxyphenylpropane units in the middle-lamella lignin of spruce (*Picea abies*). *Wood Sci. Technol.* **19**, 223–232 (1985).
- Lapierre, C. in *Forage Cell Wall Structure and Digestibility* 133–166 (John Wiley & Sons, Ltd, 1995). <https://doi.org/10.2134/1993.foragecellwall.c6>.
- Ride, J. P. Lignification in wounded wheat leaves in response to fungi and its possible rôle in resistance. *Physiol. Plant Pathol.* **5**, 125–134 (1975).
- Hammerschmidt, R., Bonnen, A. M., Bergstrom, G. C. & Baker, K. K. Association of epidermal lignification with nonhost resistance of cucurbits to fungi. *Can. J. Bot.* **63**, 2393–2398 (1985).
- Doster, M. A. & Bostock, R. M. Quantification of lignin formation in almond bark in response to wounding and infection by *Phytophthora* species. *Phytopathology* **78**, 473–477 (1988).
- Campbell, M. M. & Ellis, B. E. Fungal elicitor-mediated responses in pine cell cultures: cell wall-bound phenolics*. *Int. J. Plant Biochem.* **31**, 737–742 (1992).
- Rogers, L. A. et al. Comparison of lignin deposition in three ectopic lignification mutants. *N. Phytol.* **168**, 123–140 (2005).
- Hématy, K. et al. A receptor-like kinase mediates the response of Arabidopsis cells to the inhibition of cellulose synthesis. *Curr. Biol.* **17**, 922–931 (2007).
- Cheung, A. Y. & Wu, H.-M. THESEUS 1, FERONIA and relatives: a family of cell wall-sensing receptor kinases? *Curr. Opin. Plant Biol.* **14**, 632–641 (2011).
- Liu, J., Osbourn, A. & Ma, P. MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *Mol. Plant* **8**, 689–708 (2015).
- Chezem, W. R., Memon, A., Li, F. -S., Weng, J. -K. & Clay, N. K. SG2-Type R2R3-MYB transcription factor MYB15 controls defense-induced lignification and basal immunity in Arabidopsis. *Plant Cell Online* **29**, 1907–1926 (2017).
- Kim, S. H. et al. The Arabidopsis R2R3 MYB transcription factor MYB15 is a key regulator of lignin biosynthesis in effector-triggered immunity. *Front. Plant Sci.* **11**, 583153 (2020).
- Kai, K. et al. Scopoletin is biosynthesized via ortho-hydroxylation of feruloyl CoA by a 2-oxoglutarate-dependent dioxygenase in *Arabidopsis thaliana*. *Plant J.* **55**, 989–999 (2008).
- Vanhohle, R. et al. COSY catalyses trans-cis isomerization and lactonization in the biosynthesis of coumarins. *Nat. Plants* **5**, 1066–1075 (2019).
- Djébali, N. et al. Partial resistance of *Medicago truncatula* to *Aphanomyces euteiches* is associated with protection of the root stele and is controlled by a major QTL rich in proteasome-related genes. *MPMI* **22**, 1043–1055 (2009).
- Turner, M. et al. Dissection of bacterial Wilt on *Medicago truncatula* revealed two type III secretion system effectors acting on root infection process and disease development. *Plant Physiol.* **150**, 1713–1722 (2009).
- Thomas, R. et al. Soybean root suberin: anatomical distribution, chemical composition, and relationship to partial resistance to *Phytophthora sojae*. *Plant Physiol.* **144**, 299–311 (2007).
- Holbein, J. et al. Root endodermal barrier system contributes to defence against plant-parasitic cyst and root-knot nematodes. *Plant J.* **100**, 221–236 (2019).
- Kroemer, K. in *Wurzelhaut, Hypodermis und Endodermis der Angiospermenwurzel* 59 (Bibl Bot, 1903).
- Behrisch, R. Zur Kenntnis der Endodermiszelle. *Ber. Dtsch. Bot. Ges.* **44**, 162–164 (1926).
- Wang, P. et al. Surveillance of cell wall diffusion barrier integrity modulates water and solute transport in plants. *Sci. Rep.* **9**, 4227 (2019).
- Fernández-Marcos, M. et al. Control of Arabidopsis lateral root primordium boundaries by MYB36. *New Phytol.* **213**, 105–112 (2016).
- Javot, H. & Maurel, C. The role of aquaporins in root water uptake. *Ann. Bot.* **90**, 301–313 (2002).
- Javot, H. et al. Role of a single aquaporin isoform in root water uptake. *Plant Cell* **15**, 509–522 (2003).
- Alexandersson, E. et al. Whole gene family expression and drought stress regulation of aquaporins. *Plant Mol. Biol.* **59**, 469–484 (2005).
- Olas, J. J., Fichtner, F. & Apelt, F. All roads lead to growth: imaging-based and biochemical methods to measure plant growth. *J. Exp. Bot.* **71**, 11–21 (2019).
- Salas-González, I. et al. Coordination between microbiota and root endodermis supports plant mineral nutrient homeostasis. *Science* **371**, eabd0695 (2021).
- Andersen, T. G. et al. Diffusible repression of cytokinin signalling produces endodermal symmetry and passage cells. *Nature* **555**, 529–533 (2018).
- Barberon, M. et al. Adaptation of root function by nutrient-induced plasticity of endodermal differentiation. *Cell* **164**, 447–459 (2016).
- Vermeer, J. E. M. et al. A spatial accommodation by neighboring cells is required for organ initiation in Arabidopsis. *Science* **343**, 178–183 (2014).
- Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).

59. Ursache, R., Andersen, T. G., Marhavy, P. & Geldner, N. A protocol for combining fluorescent proteins with histological stains for diverse cell wall components. *Plant J.* **93**, 399–412 (2018).
60. Beeckman, T. & Viane, R. Embedding thin plant specimens for oriented sectioning. *Biotech. Histochem.* **75**, 1–4 (1999).
61. Lord, S. J., Velle, K. B., Mullins, R. D. & Fritz-Laylin, L. K. SuperPlots: communicating reproducibility and variability in cell biology. *J. Cell Biol.* **219**, 94–10 (2020).
62. Logemann, J., Schell, J. & Willmitzer, L. Improved method for the isolation of RNA from plant tissues. *Anal. Biochem.* **163**, 16–20 (1987).
63. Bushnell, B., Rood, J. & Singer, E. BBMerge—accurate paired shotgun read merging via overlap. *PLoS ONE* **12**, e0185056 (2017).
64. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
65. Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K. & Scheible, W.-R. Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiol.* **139**, 5–17 (2005).
66. Morreel, K. et al. Genetical metabolomics of flavonoid biosynthesis in Populus: a case study. *Plant J.* **47**, 224–237 (2006).
67. Morreel, K. et al. Systematic structural characterization of metabolites in Arabidopsis via candidate substrate-product pair networks. *Plant Cell* **26**, 929–945 (2014).
68. R Core Team. in *R: A language and environment for statistical computing* (R Foundation for Statistical Computing, R Core Team, 2018).
69. Burbidge, J. B., Magee, L. & Robb, A. L. Alternative transformations to handle extreme values of the dependent variable. *J. Am. Stat. Assoc.* **83**, 123–127 (1988).
70. Lê, S., Josse, J. & Husson, F. FactoMineR: an R package for multivariate analysis. *J. Stat. Softw.* **25**, 1–18 (2008).
71. Danku, J. M. C., Lahner, B., Yakubova, E. & Salt, D. E. Large-scale plant ionomics. *Methods Mol. Biol.* **953**, 255–276 (2013).
72. Vanholme, B., Houari, E.I. & Boerjan, W. Bioactivity: phenylpropanoids' best kept secret. *Curr. Opin. Biotechnol.* **56**, 156–162 (2018).

Acknowledgements

We thank Deep Seq (Next Generation Sequencing Facility of the University of Nottingham, UK), the nmRC (Nanoscale and Microscale Research Centre of the University of Nottingham, UK), the Microscopy and Histology Facility of the University of Aberdeen (UK), the VIB Metabolomics Core (VIB-UGent, Belgium). This work was supported by grants from the UK Biotechnology and Biological Sciences Research Council Grant (Grant Nos. BB/L027739/1 and BB/N023927/1 to D.E.S.), the Coordinating Action in Plant Sciences Promoting sustainable collaboration in plant sciences (Grant Nos. ERACAPS13.089_RootBarriers to D.E.S.), the Engineering and Physical Sciences Research Council (Grant No. EP/R025282/1) and the Future Food Beacon of Excellence at the University of Nottingham (Nottingham Research Fellowship to G.C., and Postdoctoral Research Fellowship to G.R.).

Author contributions

G.R., P.R., G.C. and D.E.S. conceived and designed the experiments, I.S.-G. analyzed RNA-seq data, S.F. generated the endodermal plasma membrane marker, G.R. and P.R. performed Raman imaging, A.L., D.T. and M.W.G. analyzed Raman data, C.L. performed thioacidolysis experiments, K.M. analyzed metabolomic data, M.C.P. and Y.B. performed hydraulic conductivity measurements, P.F. performed ICP-MS measurements, G.R. wrote the paper with multiple rounds of review and editing from D.E.S., G.R., S.F., C.L., P.F., N.G., Y.B., W.B., M.W.G., G.C. and D.E.S. All contributed to final editing and review.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-22550-0>.

Correspondence and requests for materials should be addressed to D.E.S.

Peer review information *Nature Communications* thanks Richard Dixon and the other, anonymous reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

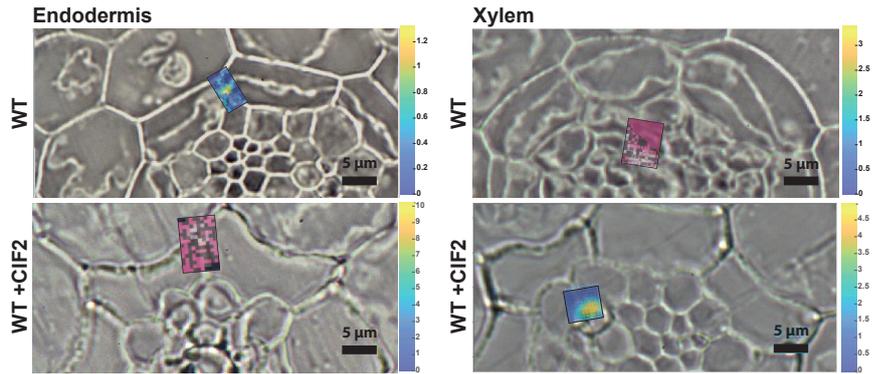
Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



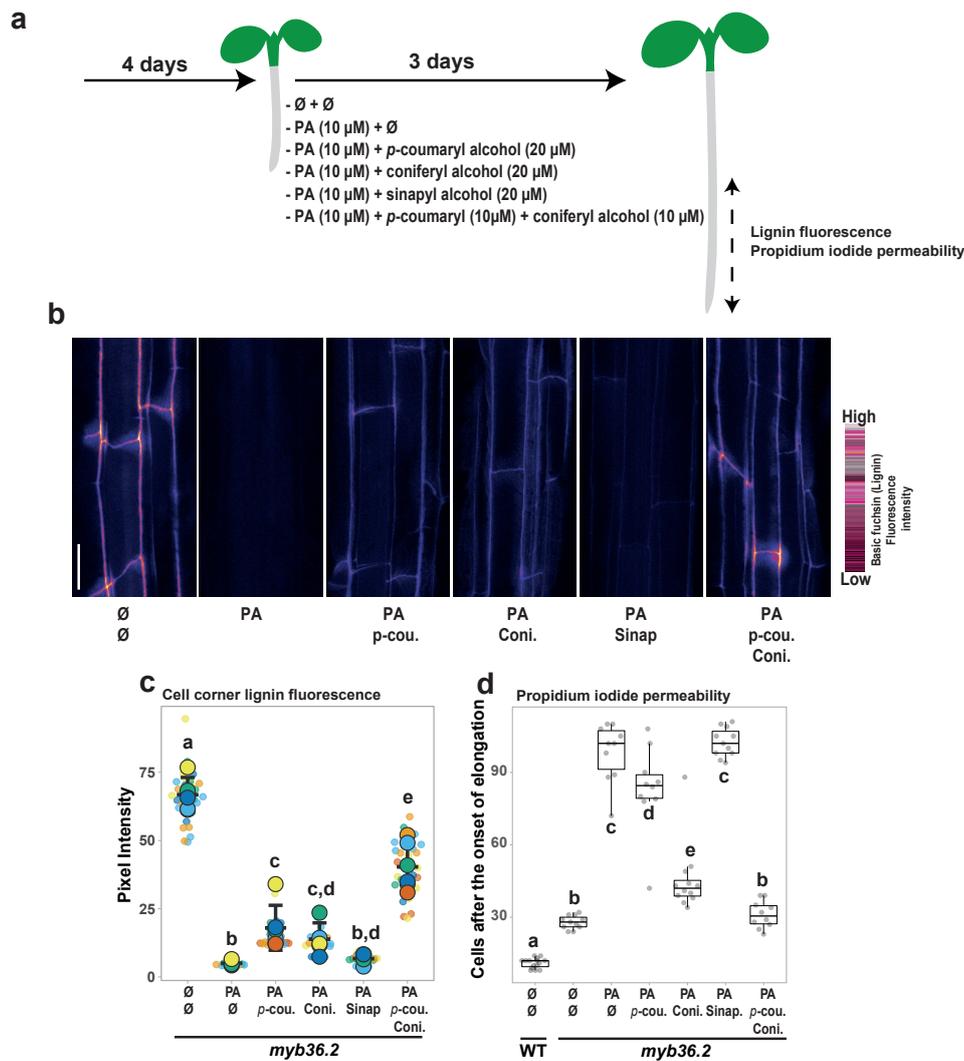
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021



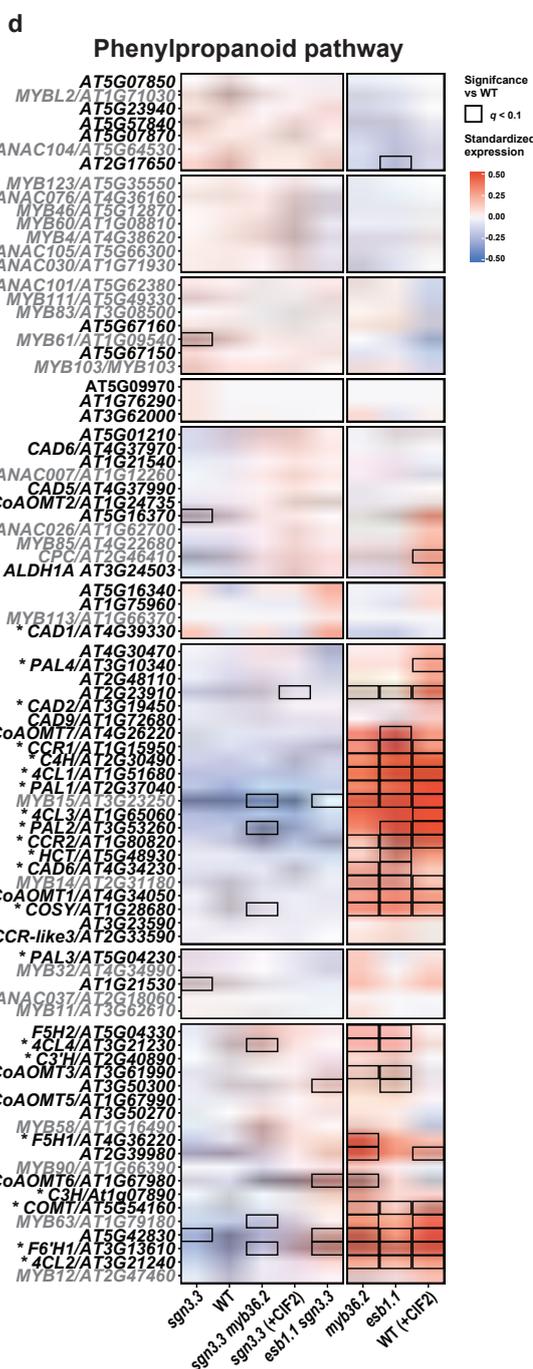
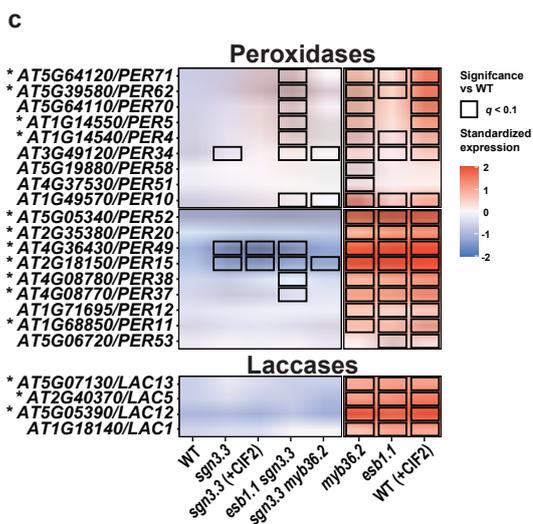
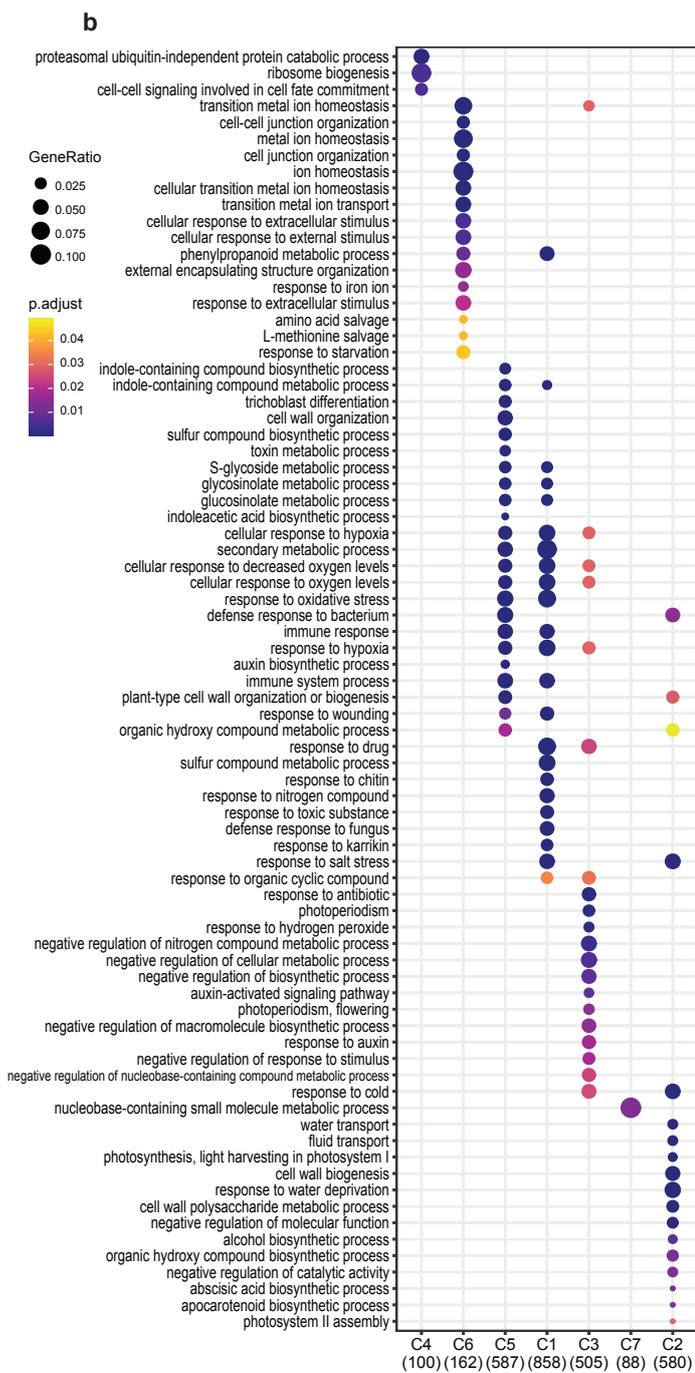
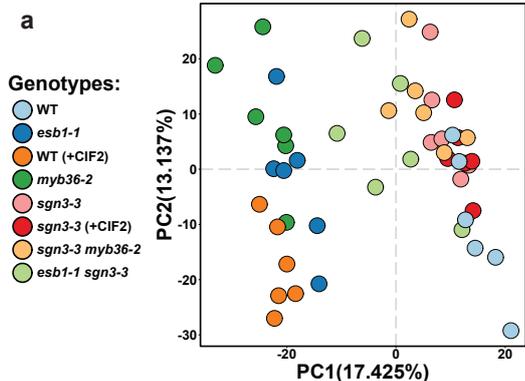
Supplementary Fig. 1. Activation of the Schengen-pathway triggers the deposition of a distinct “stress” lignin in the endodermis.

Examples of small Raman maps for endodermal cells of root cross-sections in WT(\emptyset) and WT(+CIF2) and for xylem of WT(\emptyset) and WT(+CIF2) used for determining the lignin spectra using Multivariate Curve Resolution (MCR) presented in Fig. 2c, d. The colour code represents the intensity of the lignin factor presented in Fig. 2c, d. Similar small Raman maps were obtained from independent plants for CS lignin of WT (n = 8 plants), cell-corner lignin of WT treated with CIF2 (+CIF2; n = 5 plants), for xylem lignin of WT (n = 2 plants) and xylem lignin of WT treated with CIF2 (n = 2 plants).



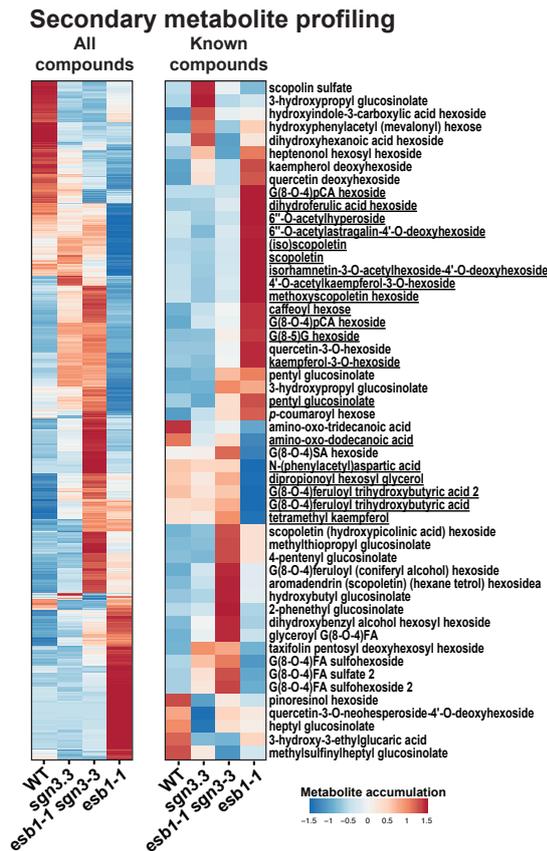
Supplementary Fig. 2. The lignin compositional changes induced by the Schengen pathway contributes to the sealing of the apoplast.

a Scheme of experimental design. The *myb36-2* mutant was germinated for 4 days on control condition and then transferred for three more days on a control media (Ø + Ø), on a media containing 10 µM piperonylic acid only (PA + Ø), or supplemented with 20 µM *p*-coumaryl alcohol (PA + *p*-cou.), with 20 µM coniferyl alcohol (PA + Coni.), with 20 µM sinapyl alcohol (PA + Sinap.) and with the combination of 10 µM *p*-coumaryl alcohol and 10 µM coniferyl alcohol (PA + *p*-cou + Coni.). **b** Maximum intensity projection of the top endodermal cells at ~20 cells after the onset of elongation in roots stained with Basic fuchsin (lignin). Scale bar = 25 µm. Representative pictures are shown. The experiment was repeated two times independently with similar results. **c** Fluorescence intensity of lignin stained using Basic fuchsin in the cortex-facing cell wall corner of endodermal cells at ~20 cells after the onset of elongation in *myb36-2*. Individual data points (small dots) represent the pixel intensity for one endodermal cell. Measurements of pixel intensity were performed in 3 to 6 individual cells per plants and measurements were performed in 5 plants for (Ø + Ø), 4 plants for (PA + Ø), 6 plants for (PA + *p*-cou.), 5 plants for (PA + coni.), 5 plants for (PA + Sinap.) and 6 plants for (PA + *p*-cou + Coni.). Large dots represent the mean of pixel intensity measured in each plant. Black lines represent the mean and standard deviation for each treatment. Letters show significantly different treatments determined by an ANOVA and Tukey's test as post hoc analyses ($p < 0.01$). **d** Boxplot showing the number of cells from the onset of elongation permeable to propidium iodide in wild-type (WT) and *myb36.2* in control condition (Ø + Ø) and after pharmacological treatments described in (a) for *myb36.2*. Measurements were performed in 15 plants for WT (Ø + Ø) WT, 11 plants for *myb36.2* (Ø + Ø), 10 plants for *myb36.2* (PA + Ø), 10 plants for *myb36.2* (PA + *p*-cou.), 12 plants for *myb36.2* (PA + *p*-coni.), 11 plants for *myb36.2* (PA + Sinap.) and 10 plants for *myb36.2* (PA + *p*-cou + Coni.). Center lines show the medians; box limits indicate the 25th and 75th percentiles. Different letters represent significant differences between treatments using a two-sided Mann-Whitney test ($p < 0.05$).



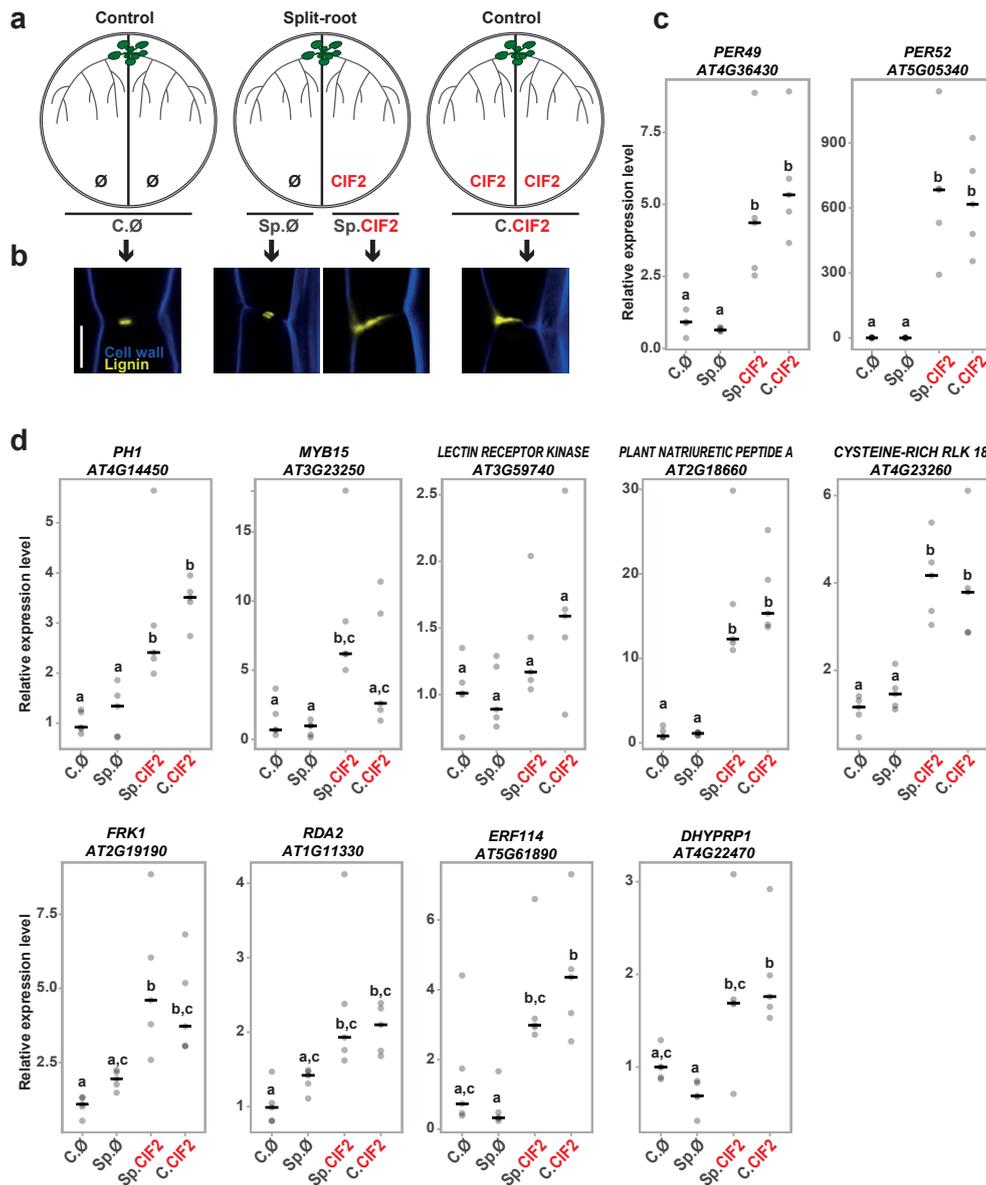
Supplementary Fig. 3. Gene expression profiling in response to the activation of the Schengen-pathway.

a Principal component analysis (PCA) of the differentially expressed genes identified in root tips of wild-type (WT), *sgn3-3*, *esb1-1*, *myb36-2*, *esb1-1 sgn3-3*, *sgn3-3 myb36-2* plants. Treatment with 100 nM CIF2 was applied as indicated (+CIF2) for WT and *sgn3-3* plants (n = 6 biological replicates from two independent experiments). **b** Gene ontology enrichment in the different gene clusters from Fig. 3a. The colour of each point represents the p-value adjusted using the Benjamin-Hochberg procedure, and the size of each point denotes the percentage of total differential expressed genes in the given gene ontology term (Gene Ratio). **c** Heatmap of gene expression for peroxidases and laccases that are upregulated by the constitutive activation of the Schengen pathway as defined by the Cluster 1 (C1) of the Fig. 3a. Asterisks indicate genes previously identified as upregulated in response to CIF2¹. **d** Heatmap of gene expression of genes related to the phenylpropanoid pathway (black)² and their transcriptional regulators (grey)^{3,4}. Genes names are given according to⁵ for genes related to the phenylpropanoid pathway. Asterisks indicate demonstrated function in lignin biosynthesis with an activity demonstrated *in vitro* or *in vivo* for *PAL1-4*⁶, *CAH*⁷, *4CLI-4*^{8,9}, *CCR1* and *2*^{10,11}, *CAD1*, *2* and *6*^{12,13}, *C3'H*¹⁴, *C3H*¹⁵, *COMT* and *CCoAOMT1*¹⁶, *HCT*¹⁷, *CSE*¹⁸, *ALDH1A*¹⁹, *F6'HI*²⁰, *COSY*²¹ and *F5HI*²².



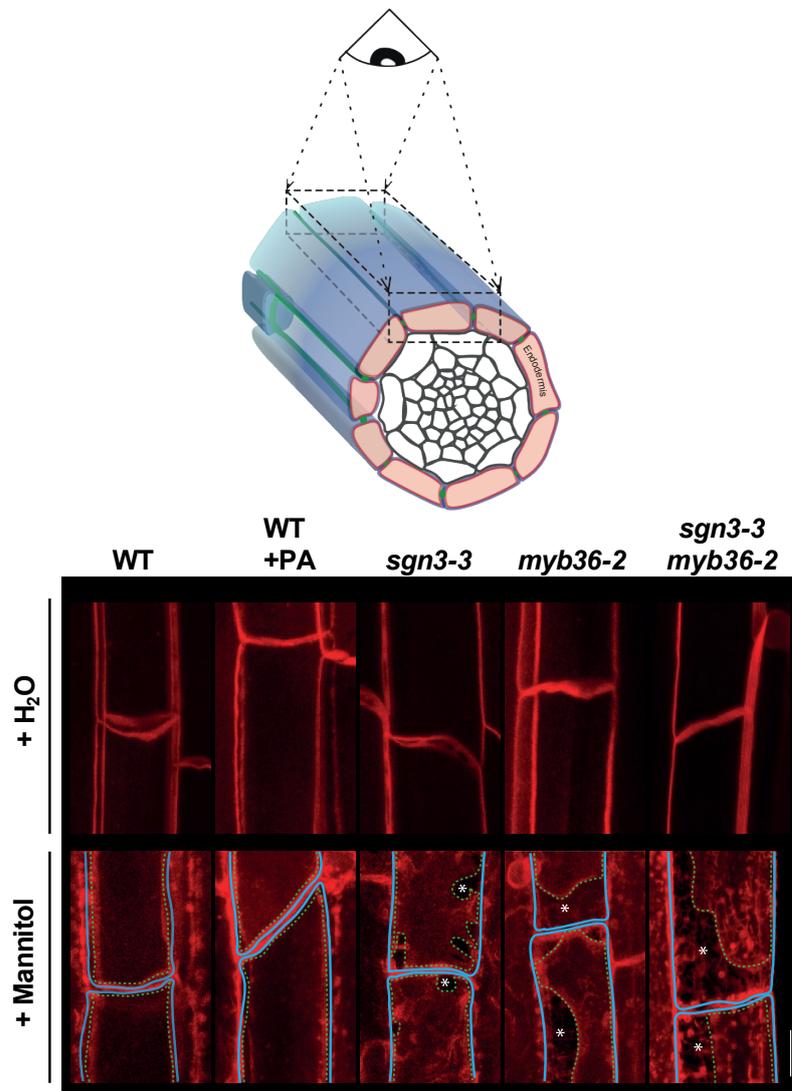
Supplementary Fig. 4. Metabolite profiling in response to the activation of the Schengen-pathway.

Heatmaps of metabolite profiling determined using Ultra High Performance Liquid Chromatography (UHPLC) in 5 mm roots tips of wild-type (WT), *sgn3-3*, *esb1-1* *sgn3-3* and *esb1-1*. The heatmaps show all the compounds (2497, left) and characterised compounds (52, right) that are differentially accumulated (q -value < 0.01, left; q -value < 0.1, right $n = 8$ biological replicates from two independent experiments). Underlined names are for compounds that are only differentially accumulated (q -value < 0.1) in *esb1-1* and not changed in *sgn3-3* and *esb1-1* *sgn3-3* in comparison with WT. Data for the known compounds are presented in Supplementary Data 3.



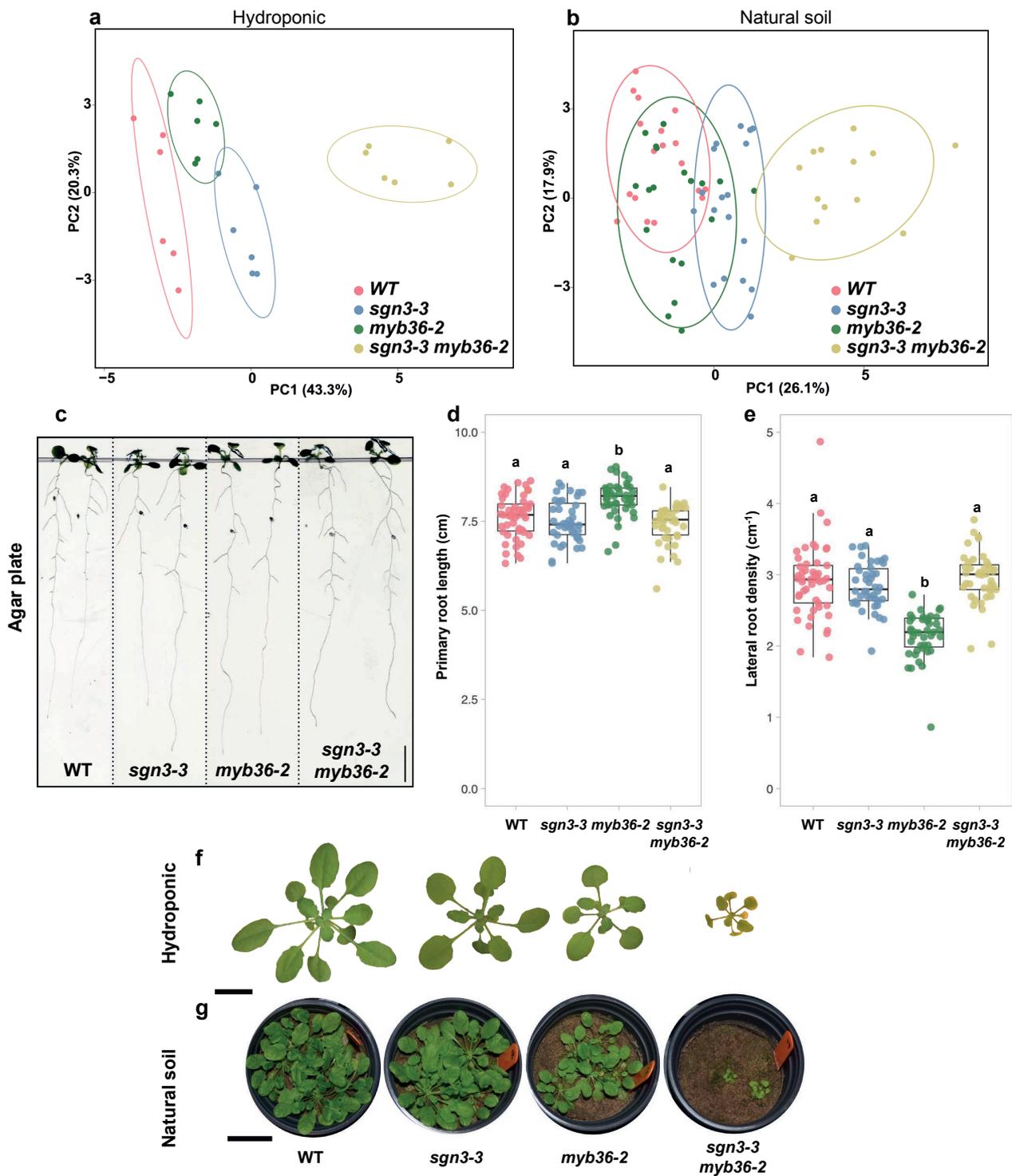
Supplementary Fig. 5. Activation of the Schengen-pathway triggers local lignification and induction of defense-related genes.

a. Scheme of experimental design. Twelve day-old wild-type plants were transferred on round plates containing two compartments. The compartments were filled with control media on both compartments (C.Ø), media supplemented with 100 nM CIF2 on both compartments (C.CIF2) or media only on one side (Sp.Ø) and media supplemented with 100 nM CIF2 (Sp.CIF2) on the other side for the split condition. Plants were grown for 3 more days on these plates and then the roots exposed to each compartment were harvested. **b.** Median view of endodermal cells stained with Direct Yellow 96 (cell wall, blue) and Basic Fuchsin (lignin, yellow). Cells were imaged at 15 cells after the onset of elongation. Scale bar = 10 μ m. Representative pictures are shown. Similar results were observed in at least 5 plants. **c.** Gene expression of *Peroxidase 49* and *52* (*PER49* and *52*) determined by qPCR on each compartment described in **a**. **d.** Gene expression of genes related to defense (*PHI*, *MYB15*, *LECTIN RECEPTOR KINASE*, *PLANT NATRIURETIC PEPTIDE A*, *CYSTEINE-RICH RLK 18*, *FRK1*, *RDA2*, *ERF114* and *DHYPRP1*) determined by qPCR. Different letters in panels **a** et **b** indicate significant differences between treatments determined by an ANOVA and Tukey's test as post hoc analyses ($p < 0.05$, $n = 6$ biological replicates). Horizontal black lines indicate median values.



Supplementary Fig. 6. Plasma membrane attachment to the cell wall.

Maximum projection of the top endodermal cells as shown in the schematic view. The observations were done in lines expressing the plasma membrane marker line pELTP::SYP122mCitrine before plasmolysis (+H₂O) and after plasmolysis (+Mannitol) at 15 cells after the onset of elongation. The dashed line represents the contours of the cells. Asterisks show the plasmolysis generated space where no attachment is observed. Scale bar = 5 μ m. Representative pictures are shown. The experiment was repeated three times independently with similar results.



Supplementary Fig. 7. Absence of endodermal apoplastic barrier triggers major ionic changes in different growth conditions.

a, b Principal component analysis (PCA) based on the concentration of 20 elements in shoots of WT, *sgn3-3*, *myb36.3* and *sgn3-3 myb36-2* plants grown in **(a)** hydroponics (short day, n=6) and **(b)** natural soil (short day, n = 18 for WT, n = 18 for *sgn3-3*, n = 18 for *myb36-2* and n = 13 for *sgn3-3 myb36-2*). Ellipses show confidence level at a rate of 90%. **c** Pictures of 2-week-old wild-type (WT), *sgn3-3*, *myb36-2* and *sgn3-3 myb36-2* plants grown in agar plates. **d, e** Boxplots showing the primary root length (**d**) and lateral roots density (**e**) of 2-week-old WT, *sgn3-3*, *myb36-2* and *sgn3-3 myb36-2* plants grown in agar plates. Letters show significantly different groups determined by an ANOVA and a Tukey's test as post hoc analyses (n = 49 for WT, n = 42 for *sgn3-3*, n = 43 for *myb36-2* and n = 41 for *sgn3-3 myb36-2*), $p < 0.01$). Center lines show the medians; box limits indicate the 25th and 75th percentiles. **f** Pictures of 5-week-old WT, *sgn3-3*, *myb36-2* and *sgn3-3 myb36-2* plants grown in hydroponics. Scale bar = 1 cm. **g** Pictures of 9-week-old WT, *sgn3-3*, *myb36-2* and *sgn3-3 myb36-2* plants grown in natural soil. Scale bar = 3 cm.

Supplementary Fig. 8. Activation of the Schengen-pathway maintains plant growth under fluctuating environment.

a Quantification of suberin staining along the root of 6 days-old plants. The results are expressed in percentage of root length divided in three zones: unsuberised (white), discontinuously suberised (yellow), continuously suberised (orange). $n = 7$ individual plants, error bars: SD, the centre of the error bars represents the mean. Individual letters show significant differences using a two-sided Mann-Whitney test between the same zones ($p < 0.01$). The experiment was repeated two times independently with similar results. **b** Graphs showing leaf surface area of WT, *sgn3-3*, *myb36-2*, *sgn3-3 myb36-2*, WT-*pELTP::CDEF* and *sgn3-3 myb36-2-pELTP::CDEF* plants germinated in soil with a high humidity (80%) for 7 days and then transferred in an environment with constant (80% RH, blue) or with a lower humidity (60% RH, red). Data were collected at 0, 2, 5 and 8 days after the transfer. Each point is the average leaf surface per plant from a single pot ($n = 36$ pots at 0 day for each genotype, $n = 6$ pots at 2, 5 and 8 days for each genotype). Each pot contained at least 6 plants for each genotype. The line shows the average value for each measured time points. Black asterisk indicates a significant difference between high and low humidity for a same genotype at one time point. Blue and red asterisk indicate a significant difference in comparison with WT at the same time point respectively for the high and low humidity environment. The significant differences were determined by an ANOVA and a Tukey's test as post hoc analyses ($p < 0.01$). **c** Graphs showing ions accumulation (Z-score) in shoots of WT, *sgn3-3*, *myb36-2* and *sgn3-3 myb36-2* plants germinated in soil with a high humidity (80%) for 10 days and then transferred in an environment with constant (80% RH, "High" in blue) or with a lower humidity (60% RH, "Low" in red) for 5 more days. Large dots represent the median value for each genotypes and small dots represent individual replicates ($n = 9$ biological replicates). Thick lines indicate a significant difference using a two-sided *t*-test ($p < 0.05$).

SUPPLEMENTARY REFERENCES

1. Fujita, S. *et al.* SCHENGEN receptor module drives localized ROS production and lignification in plant roots. *The EMBO Journal* **39**, e103894 (2020).
2. Mueller, L. A., Zhang, P. & Rhee, S. Y. AraCyc: a biochemical pathway database for Arabidopsis. *PLANT PHYSIOLOGY* **132**, 453–460 (2003).
3. Liu, J., Osbourn, A. & Ma, P. MYB Transcription Factors as Regulators of Phenylpropanoid Metabolism in Plants. *Molecular Plant* **8**, 689–708 (2015).
4. Ohtani, M. & Demura, T. The quest for transcriptional hubs of lignin biosynthesis: beyond the NAC-MYB-gene regulatory network model. *Current Opinion in Biotechnology* **56**, 82–87 (2018).
5. Raes, J., Rohde, A., Christensen, J. H., Van de Peer, Y. & Boerjan, W. Genome-wide characterization of the lignification toolbox in Arabidopsis. *PLANT PHYSIOLOGY* **133**, 1051–1071 (2003).
6. Huang, J. *et al.* Functional Analysis of the Arabidopsis PAL Gene Family in Plant Growth, Development, and Response to Environmental Stress. *Plant Physiol.* **153**, 1526 (2010).
7. Schillmiller, A. L. *et al.* Mutations in the cinnamate 4-hydroxylase gene impact metabolism, growth and development in Arabidopsis. *The Plant Journal* **60**, 771–782 (2009).
8. Costa, M. A. *et al.* Characterization in vitro and in vivo of the putative multigene 4-coumarate:CoA ligase network in Arabidopsis: syringyl lignin and sinapate/sinapyl alcohol derivative formation. *The International Journal of Plant Biochemistry* **66**, 2072–2091 (2005).
9. Li, Y., Kim, J. I., Pysh, L. & Chapple, C. Four Isoforms of Arabidopsis 4-Coumarate:CoA Ligase Have Overlapping yet Distinct Roles in Phenylpropanoid Metabolism. *PLANT PHYSIOLOGY* **169**, 2409–2421 (2015).
10. Lauvergeat, V. *et al.* Two cinnamoyl-CoA reductase (CCR) genes from Arabidopsis thaliana are differentially expressed during development and in response to infection with pathogenic bacteria. *The International Journal of Plant Biochemistry* **57**, 1187–1195 (2001).
11. Baltas, M. *et al.* Kinetic and inhibition studies of cinnamoyl-CoA reductase 1 from Arabidopsis thaliana. *Plant Physiology et Biochemistry* **43**, 746–753 (2005).
12. Sibout, R. *et al.* CINNAMYL ALCOHOL DEHYDROGENASE-C and -D are the primary genes involved in lignin biosynthesis in the floral stem of Arabidopsis. *THE PLANT CELL ONLINE* **17**, 2059–2076 (2005).
13. Eudes, A. *et al.* Evidence for a role of AtCAD 1 in lignification of elongating stems of Arabidopsis thaliana. *Planta* **225**, 23–39 (2006).
14. Franke, R. *et al.* The Arabidopsis REF8 gene encodes the 3-hydroxylase of phenylpropanoid metabolism. *Plant J.* **30**, 33–45 (2002).
15. Barros, J. *et al.* 4-Coumarate 3-hydroxylase in the lignin biosynthesis pathway is a cytosolic ascorbate peroxidase. *Nature Communications* 1–11 (2019). doi:10.1038/s41467-019-10082-7
16. Do, C.-T. *et al.* Both caffeoyl Coenzyme A 3-O-methyltransferase 1 and caffeic acid O-methyltransferase 1 are involved in redundant functions for lignin, flavonoids and sinapoyl malate biosynthesis in Arabidopsis. *Planta* **226**, 1117–1129 (2007).
17. Hoffmann, L., Maury, S., Martz, F., Geoffroy, P. & Legrand, M. Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. *J. Biol. Chem.* **278**, 95–103 (2003).
18. Vanholme, R. *et al.* Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in Arabidopsis. *Science* **341**, 1103–1106 (2013).

19. Nair, R. B., Bastress, K. L., Ruegger, M. O., Denault, J. W. & Chapple, C. The *Arabidopsis thaliana* REDUCED EPIDERMAL FLUORESCENCE1 gene encodes an aldehyde dehydrogenase involved in ferulic acid and sinapic acid biosynthesis. *THE PLANT CELL ONLINE* **16**, 544–554 (2004).
20. Kai, K. *et al.* Scopoletin is biosynthesized via ortho-hydroxylation of feruloyl CoA by a 2-oxoglutarate-dependent dioxygenase in *Arabidopsis thaliana*. *The Plant Journal* **55**, 989–999 (2008).
21. Vanholme, R. *et al.* COSY catalyses trans–cis isomerization and lactonization in the biosynthesis of coumarins. *Nature Plants* 1–12 (2019). doi:10.1038/s41477-019-0510-0
22. Meyer, K., Shirley, A. M., Cusumano, J. C., Bell-Lelong, D. A. & Chapple, C. Lignin monomer composition is determined by the expression of a cytochrome P450-dependent monooxygenase in *Arabidopsis*. *Proceedings of the National Academy of Sciences* **95**, 6619–6623 (1998).