A chemoenzymatic strategy for protein-nanocellulose conjugates

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Roxana Lemnaru

aus Romania

Basel, 2018

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von:

Prof. Dr. Florian Peter Seebeck

Prof. Dr. Dennis Gillingham

Basel, den 20.09.2016

Prof. Dr. Jörg Schibler

Lumina nu vine din lumina, ci din intuneric.

Mircea Eliade

Table of Contents

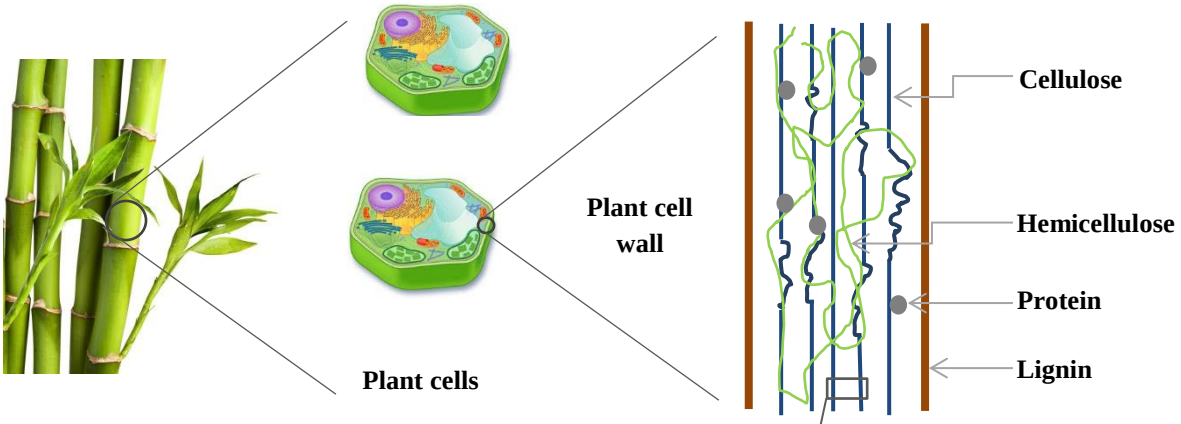
List of abbreviations

AFM – Atomic force microscopy
Boc – tert-Butoxycarbonyl
BSA – Bovine serum albumin
bp – Base pair
CBD – Cellulose-binding domain
CESA – Cellulose synthase proteins
CMF – Cellulose microfibrils
CNC – Cellulose crystals
Cys – Cysteine
DNA – Deoxyribonucleic acid
DTT – DL-Dithiothreitol
EDC – 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
Et_3N – Triethylamine
eq – Equivalent(s)
FGly – Formylglycine
His – L-Histidine
HPLC – High-performance liquid chromatography
HRMS – High-resolution mass spectrometry
IgG – Immunoglobulin G
IPTG – Isopropyl B-D-thiogalactopyranoside
MS – Mass spectrometry
NFC – Cellulose nanofibrils
NHS – *N*-Hydroxysuccinimide
Ni-NTA – Ni-Nitrilotriacetic acid resin
NMR – Nuclear magnetic resonance
PCR – Polymerase chain reaction
PEG – Polyethylene glycol
SDS-PAGE – Sodium dodecylsulfate polyacrylamide gel electrophoresis
TBAHS – Tetrabutylammonium hydrogen sulfate
TEM – Transmission electron microscopy
TFA – Trifluoroacetic acid
Tris – 2-Amino-2-(hydroxymethyl)-1,3-propanediol
UDP – Uridine diphosphate

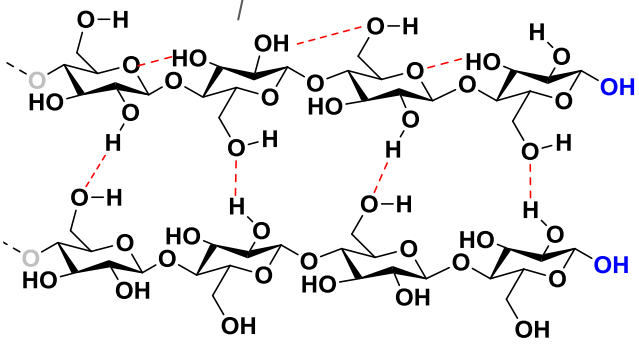
Chapter 1 - Novel composites of nanocellulose

1. 1 Function of molecular protein-nanocomposites in nature

Nature's hardest materials - nacre[1], bone[2], chitin[3] and wood[4], are made of hard/soft hybrid composites.[5] It is this fine interweaving of structural motifs that allows natural materials to achieve such high performances at the macroscale level.[5] For instance, in enamel, proteins comprise 1% of the total enamel weight, however removal of these proteins from the interfaces of the rods results in a 40% decrease of their toughness.[6] In the exoskeleton of insects, strong chitin nanofibrils are embedded in calcium carbonates and/or protein matrices[3], and in wood, strong fibres such as cellulose microfibrils are embedded in a matrix of soft hemicelluloses, protein and lignin[4] as shown in **Figure 1.1**.



a) **b)**



Hydrogen bonding within and between (1, 4)-*β*-D-glucan chains

c)

Figure 1.1 Schematic representation of the cellulose microfibrils in plants. **a)** Zoom into the plant wall. **b)** Zoom into the plant cell wall showing cellulose microfibrils (blue) placed into a matrix of hemicellulose (green), lignin (brown) and proteins (grey). **c)** Chemical structure of the (1,4)-*β*-D-glucan cellulose chains. The intra- and inter-molecular hydrogen bonding network in cellulose chains is highlighted in red. The cellulose chain has non-reducing ends (acetal functional groups, highlighted in grey) and reducing-ends (hemiacetal functional group, highlighted in blue) at either end.

Cellulose is a polymer of D-glucose in which each unit is linked to the next by a covalent bond with *β* stereochemistry at the anomeric centre.[7] The intra-molecular hydrogen bonding between the hydroxyl groups and oxygen atoms, stabilises the linear structure of the long cellulose chains.[7] Even in the first steps of its biosynthesis, cellulose never occurs as a single chain, but rather exists as composites of many chains called microfibrils, which associate very strongly *via* intra- and inter- chain hydrogen bonding.[8] Because of their chemical inertness, cellulose microfibrils which are embedded in a polysaccharide matrix of hemicellulose and pectin, protect and provide plants with rigidity.[8,9] However, the plant cell wall is not just an inert amalgam of polymers. During plant cell growth, it is the precise interaction of cellulose microfibrils with a class of proteins, known as expansins, that makes the wall susceptible to extension.[10,11] It is hypothesised that under the mechanical stress arising from cell enlargement, this matrix of interlinked polysaccharides is disrupted by expansins, as shown in the model in **Figure 1.2**. Cellulose microfibrils are believed to be adhered by strings of hemicelluloses *via* hydrogen bonds. During the turgor driven wall extension, expansins act as a primary wall loosening agents by breaking the hydrogen bond network which cross-links the hemicelluloses strings with cellulose microfibrils.

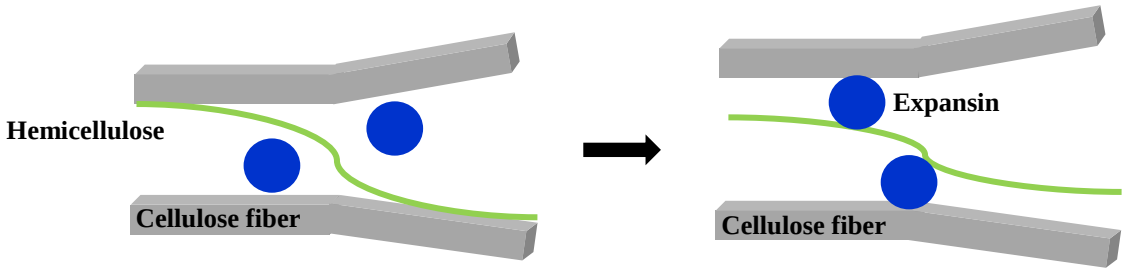


Figure 1.2[12] Model describing the action of expansins. By disrupting the hydrogen bonds between hemicelluloses and cellulose, expansins irreversibly induce cell wall extension without hydrolysing cell wall polymers.

1. 2 Biosynthesis of cellulose

Cellulose is the most abundant organic polymer on Earth, with an annual production estimated to be 7.5×10^{10} tons.[13] However, it was 20 years ago when the plant cellulose synthase genes were identified.[14] Following the plant model in *Arabidopsis thaliana*[14], cellulose microfibrils are synthesised by large membrane complexes, known as the cellulose synthase proteins (CESA). CESA are assembled into hexameric rosette units which are localised in the plasma membrane (**Figure 1.3**, left).[10] It is believed that the synthase proteins are assembled into the rosette subunits by disulfide bonds between their cysteines residues at the *N*-teminus of CESA. However, little is known about the protein composition of the rosette and the interaction between the cellulose synthase proteins within the assembly.

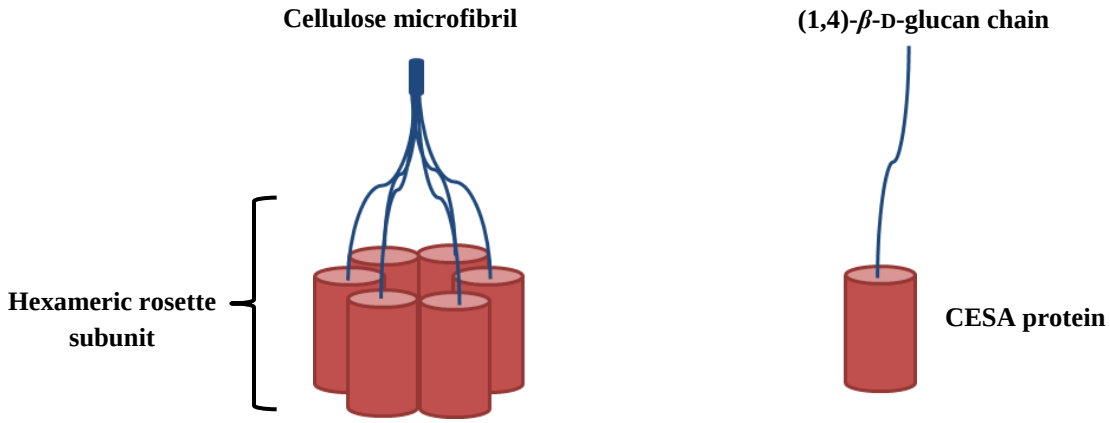


Figure 1.3[10] Model of the cellulose synthesising-machinery of the plant cell. **Left.** Representation of the synthesis of cellulose microfibrils by a hexameric assembly of CESA proteins arranged in rossete subunits. **Right.** Representation of the CESA protein which can synthesise a single (1,4)-*β*-D-glucan chain. The nascent glycan chains are released by the plasma membrane, like a spider extruding its thread.

As the catalytic domains of CESA are conserved among the cellulose-synthesising organisms (i.e. plants, bacteria, algae and tunicates), bacterial cellulose synthase proteins can be used as simpler models for understanding the biosynthesis of cellulose in plants. Bacteria *Acetobacter, Sarcina ventriculi* and *Agrobacterium* produce and secrete crystalline cellulose, which then serves as a building block for the formation of protective envelopes, called pellicles.[15] In bacteria, cellulose is produced and secreted by a protein complex consisting of at least three subunits (BcsA, BcsB and BcsC).[16] While BcsA contains a conserved family-2 glycosyltransferase (GT) domain, membrane-

anchored BcsB subunit plays a role in the translocation of the cellulose chain. The BcsC subunit is believed to transport the polymer across the outer membrane. Inverting glycosyltransferases catalyse glycosidic bonds formation by transferring the sugar unit from UDP-glucose (i.e. glucose containing a nucleoside phosphate leaving group) to the C_4 atom of the growing polysaccharide chain by an S_N2-like mechanism (**Figure 1.4**).[18]

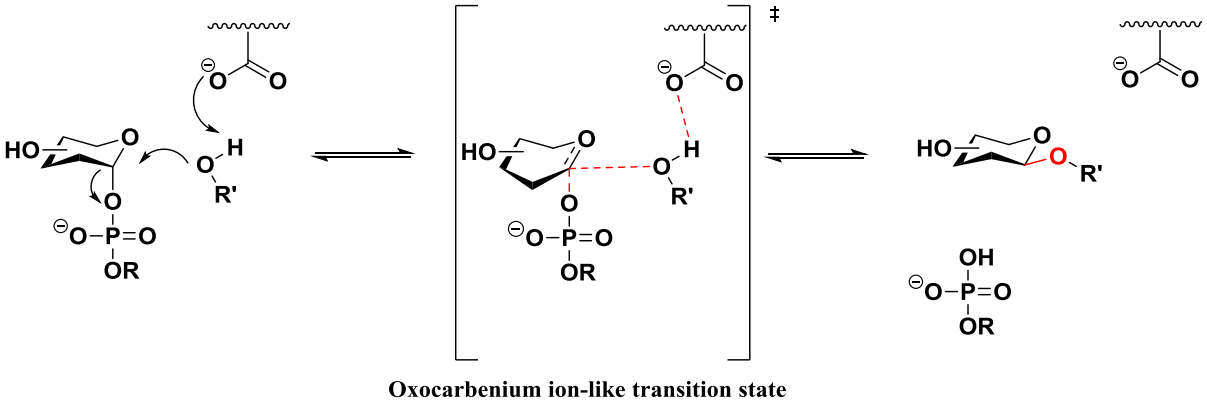


Figure 1.4[19] Mechanistic proposal for the synthesis of (1,4)-*β*-D-glucan cellulose chains *via* an oxocarbenium ion-like transition state, resulting in inversion of the stereochemistry at the anomeric centre. An active site side-chain residue serves as a base catalyst by deprotonating the incoming nucleophile. R is a nucleoside monophosphate and R' is another glucose sugar.

Recently, the first insight into the mechanistic pathways of bacterial cellulose biosynthesis was reported.[17,18] Solved crystal structures of the two subunits of cellulose synthase from *Rhodobacter sphaeroides* describe the architecture of the catalytically active subunit - BcsA and the periplasmic BcsB subunit (**Figure 1.5**). In the active site of glycosyltransferase, the donor glucose (UDP-Glc) binds in a conserved pocket beneath the acceptor molecule. Entry of a new activated glucose moiety in the BcsA binding site, initiates nucleophilic attack by the already formed chain. The newly formed acetal linkage rotates before the glycan chain is translocated into the channel and exported across the membrane.[18]

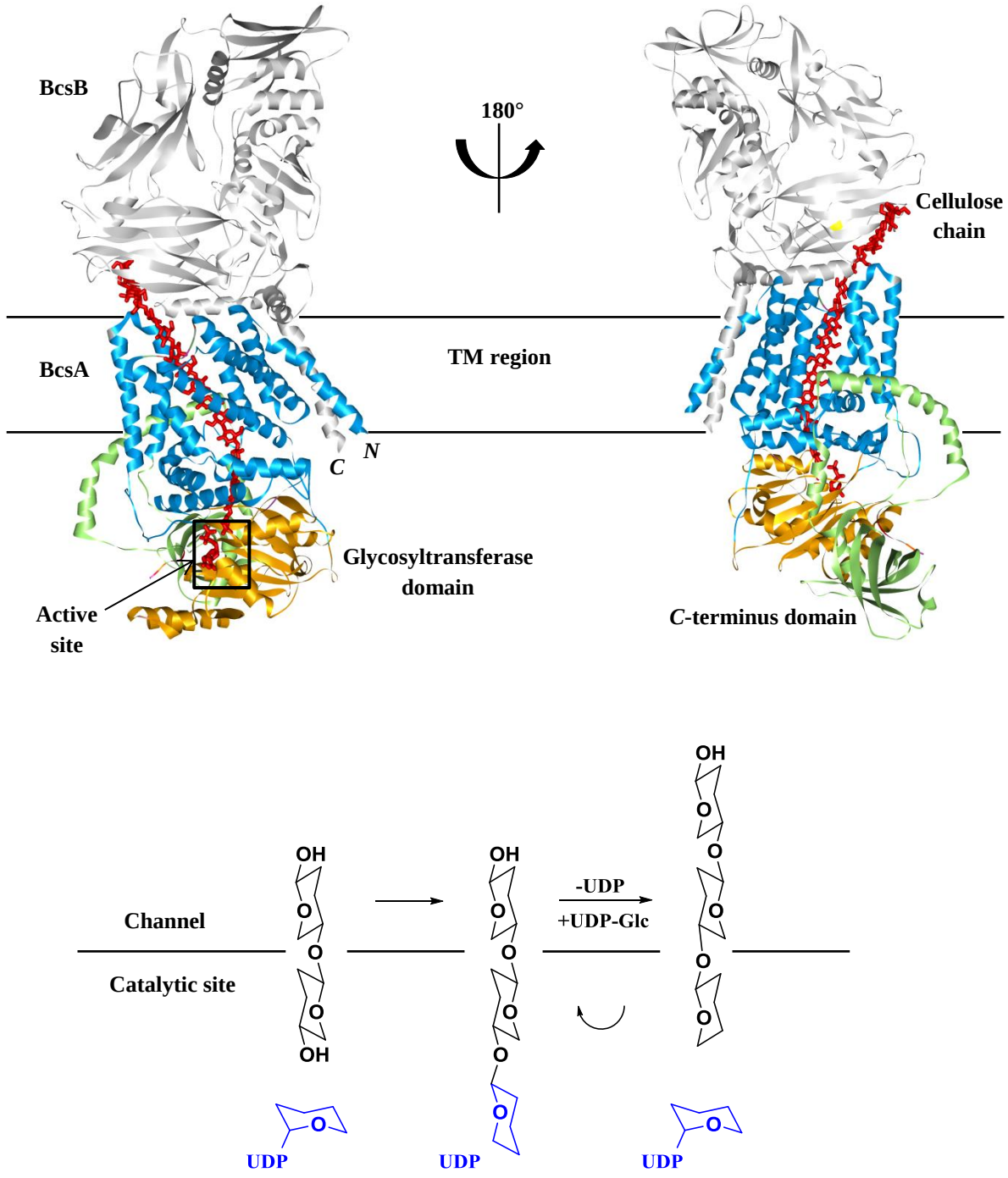


Figure 1.5[17] **Top.** Crystal structure of the BcsA-BcsB complex (pdb: 4HG6)[17] in complex with the translocating cellulose chain shown in red. BcsB is coloured grey, BcsA's transmembrane helices are coloured blue while the glycosyltransferase (GT) module is orange and the *C*-terminus domain is light green. The transmembrane (TM) region delineates the two subunits. **Bottom.** Model for the cellulose biosynthesis and translocation. The binding of substrate, followed by glycosyl transfer, generates a product-bound state which then rotates around the acetal linkage.

1. 3 Discovery and properties of nanocellulose

Cellulose has been used in wood form as a building material and in the form of cotton as a fabric, long before the discovery of its biosynthesis. The importance of cellulose as a material dates back even further to the 3rd century BC with the discovery of the Egyptian papyrus. More recently, at the beginning of the 20th century, brothers Camille and Henry Dreyfus from Basel, discovered a practical way to produce cellulose acetate.[20] Their breakthrough marked the development of manufactured fibres from polymers available in nature. The modified cellulose fibres became popular materials in many industrial fields. For instance, cellulose acetate replaced nitrate film in the movie industry and silk-like artificial fabrics were produced under the name Celanese (**Figure 1.6**).

Despite its popularity, the degradability of acetylated cellulose with a low degree of modification has limited its use.[21] One noteworthy example is the progressive deterioration of a piece of art of sculptor Naum Gabo, *Construction in Space: Two cones*. The sculpture which is among the first to be made from fibres, collapsed in 1968 due to the unstable cellulose acetate material.[22] The onset of modern polymer science, with its development of nylon- the first totally synthetic material, resulted in the replacement of natural fibres with plastic. Although displaying superior properties to cellulose acetate, replacement by plastic has led to accumulation of non-biodegradable waste. This waste has built-up over time in the great Pacific garbage patch as a vast collection of plastic fragments.[23]

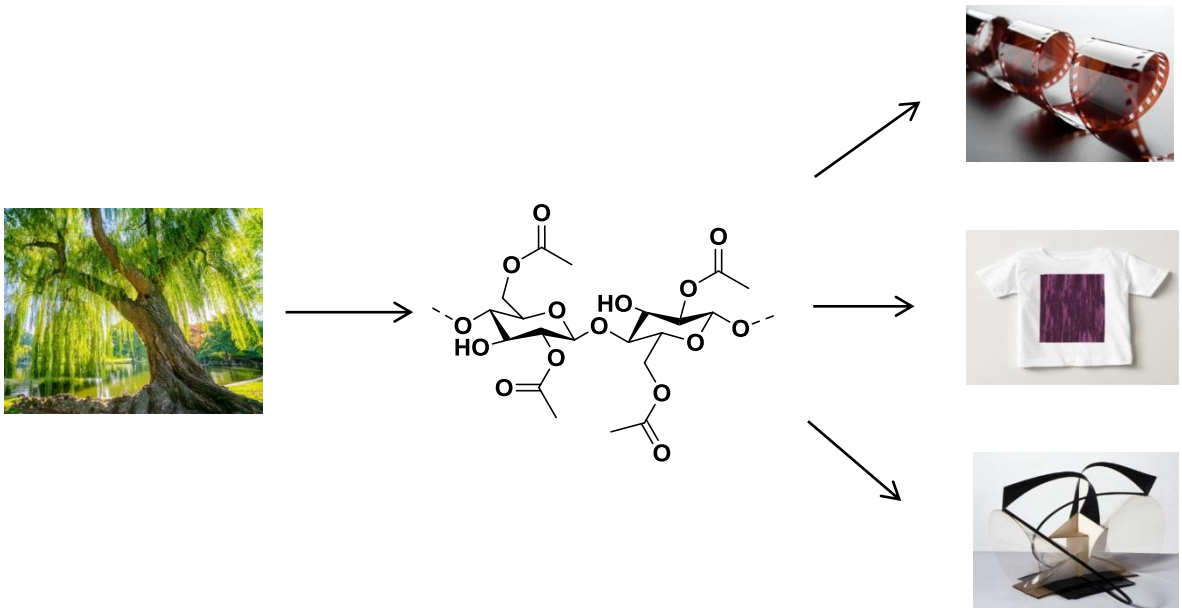


Figure 1.6 Chemical structure of the cellulose acetate chain which was used for manufacturing cellulose acetate film, Celanase fabrics and as a building material in the sculpture *Construction in Space: Two cones* (1936, replica 1969, Tate Gallery) by artist Naum Gabo.

Therefore, the question arises as to whether we can still use natural fibres to make stronger materials that could replace plastic. Almost 70 years ago, Ranby observed for the first time aqueous colloidal solutions of native cellulose fibres after treatment with sulfuric acid (**Figure 1.7**).[24] Further morphological and structural investigations by electron microscopy and X-ray diffraction measurements of these isolated sub-microscopic fibrils revealed that they had identical crystalline structure as the original fibres.[24-26] During the sulfuric acid-catalysis, disordered or semi-crystalline regions of natural cellulose were preferentially hydrolyzed, whereas the crystalline regions that have a higher resistance to acid attack remained intact (**Figure 1.7**).[27] This process led to the discovery of the smallest possible nanocrystals of cellulose (**CNCs**) which are often referred to as microcrystals, whiskers, nanocrystals, nanoparticles, microcrystallites or nanofibers.[27,28] Owing to their nanoscale dimensions, cellulose nanocrystals display exceptional mechanical properties, high stiffness of the crystalline regions and high aspect ratio.[28] Most importantly, nanocellulose is renewable.

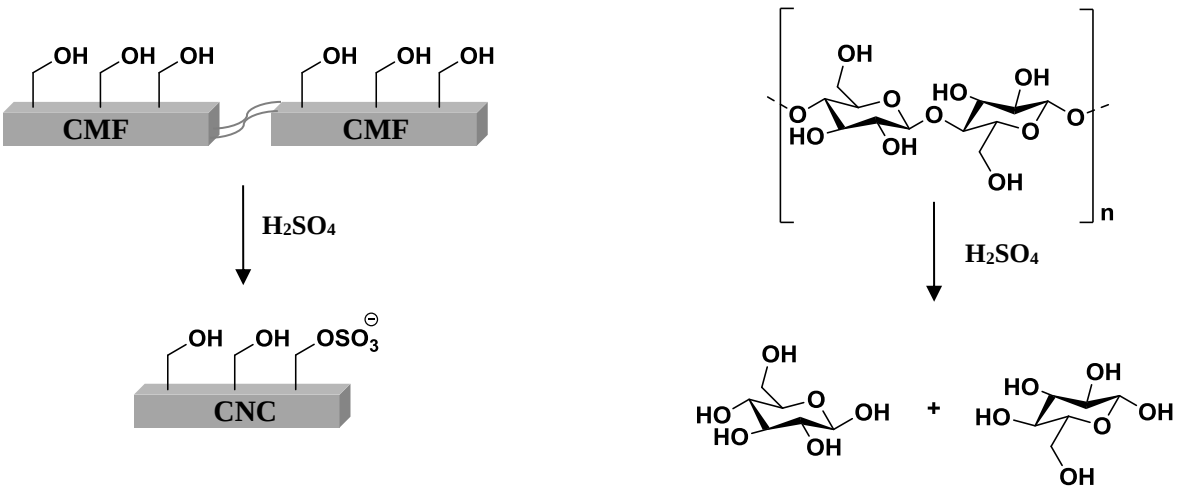


Figure 1.7 Illustration of the sulfuric acid-catalysed cellulose microfibrils (CMF, crystalline and amorphous regions) hydrolysis leading to CNs.

Recently, the fine structures of different types of cellulose nanocrystals from wood (W-NFC – wood cellulose nanofibers and W-CNC – wood cellulose nanocrystals) and bacteria (B-CNC – bacterial cellulose nanocrystals) were investigated at the individual fibril level by atomic force-,

cryogenic scanning electron- and transmission electron microscopy (AFM, cryo-SEM and TEM, respectively).[29] The TEM images in **Figure 1.8** show the morphology of the longer cellulose nanofibrils (NFC, **a**) and shorter cellulose nanocrystals (CNC, **b**). The main difference between nanofibrillated and nanowhiskers is in the aspect ratio (length/width) – NFC can have lengths of several micrometers, whereas CNC has lengths of several hundred nanometres. Moreover, the cryo-SEM image in **Figure 1.8** (**c**) shows that the fibers display right-handed chirality.

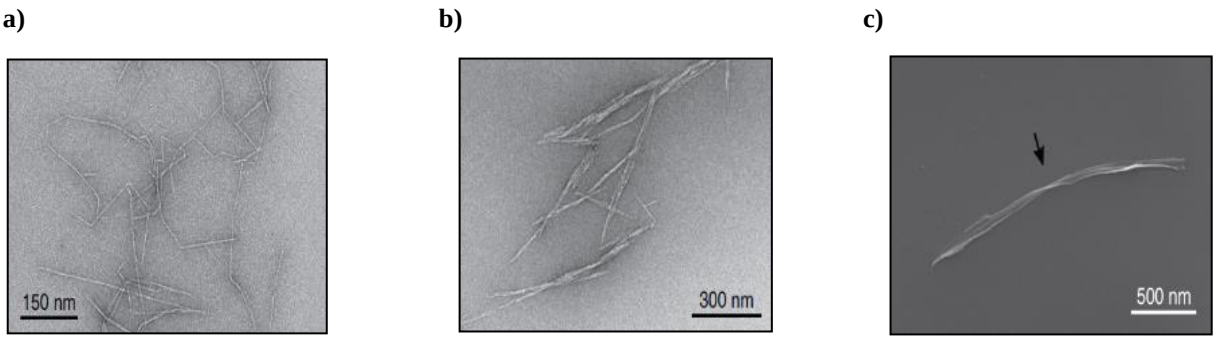


Figure 1.8[29] Microscopy images of different cellulose nanocrystals. **a)** W-NFC by TEM **b)** W-CNC by TEM **c)** B-CNC by cryo-SEM.

The rod-like structure of CNCs coupled with low-toxicity and nanoscale dimensions has inspired work on using CNCs for designing novel biomimetic materials, but also as potential nanocarriers for drug and fragrance delivery[30,31] and as bioimaging agents[32].

1. 3. 1 Strategies for the chemical modification of nanocellulose

Although conceptually simple, the chemical modification of cellulose nanocrystals provides a challenging task. Any devised synthetic transformation should not affect the rod-like shape of the nanoparticles. Moreover, during such a process, the crystalline integrity of the material should be maintained. That is to say, the envisioned chemical transformation should only take place at the surface of the nanoparticles, and not protrude into the core of the crystal. Notably, such reactions must target alcohol groups (primary and secondary) which are particularly hindered. Nevertheless, it should be borne in mind that the chemical transformation of particles imposes restrictive accessibility to the hydroxyl groups, which can result in moderate efficiency of modification. For

this reason, most investigators envisioned the modification of the superficial C_6 of the glucose unit[33], as this position is situated in a less encumbered environment and implicated in a less extensive hydrogen-bonding network. Accordingly, the reported chemical methods for the surface modification of cellulose nanocrystals include: TEMPO-mediated oxidation[34] (**Figure 1.9**, **a**), esterification[35] (**Figure 1.9**, **b**), halogenation[36] (**Figure 1.9**, **c**), $NaIO_4$ oxidation[37,38] (**Figure 1.9**, **d**), etherification[39] (**Figure 1.9**, **e**), cycloaddition[36] (**Figure 1.9**, **f**), silylation[40] (**Figure 1.9**, **g**) and cationization[41] (**Figure 1.9**, **h**). TEMPO (2,2,6,6-tetramethylpiperidine-1-oxyl)-mediated oxidation of CNCs was used for the selective conversion of the primary alcohols on the surface of the crystals into carboxylic acids.[27]

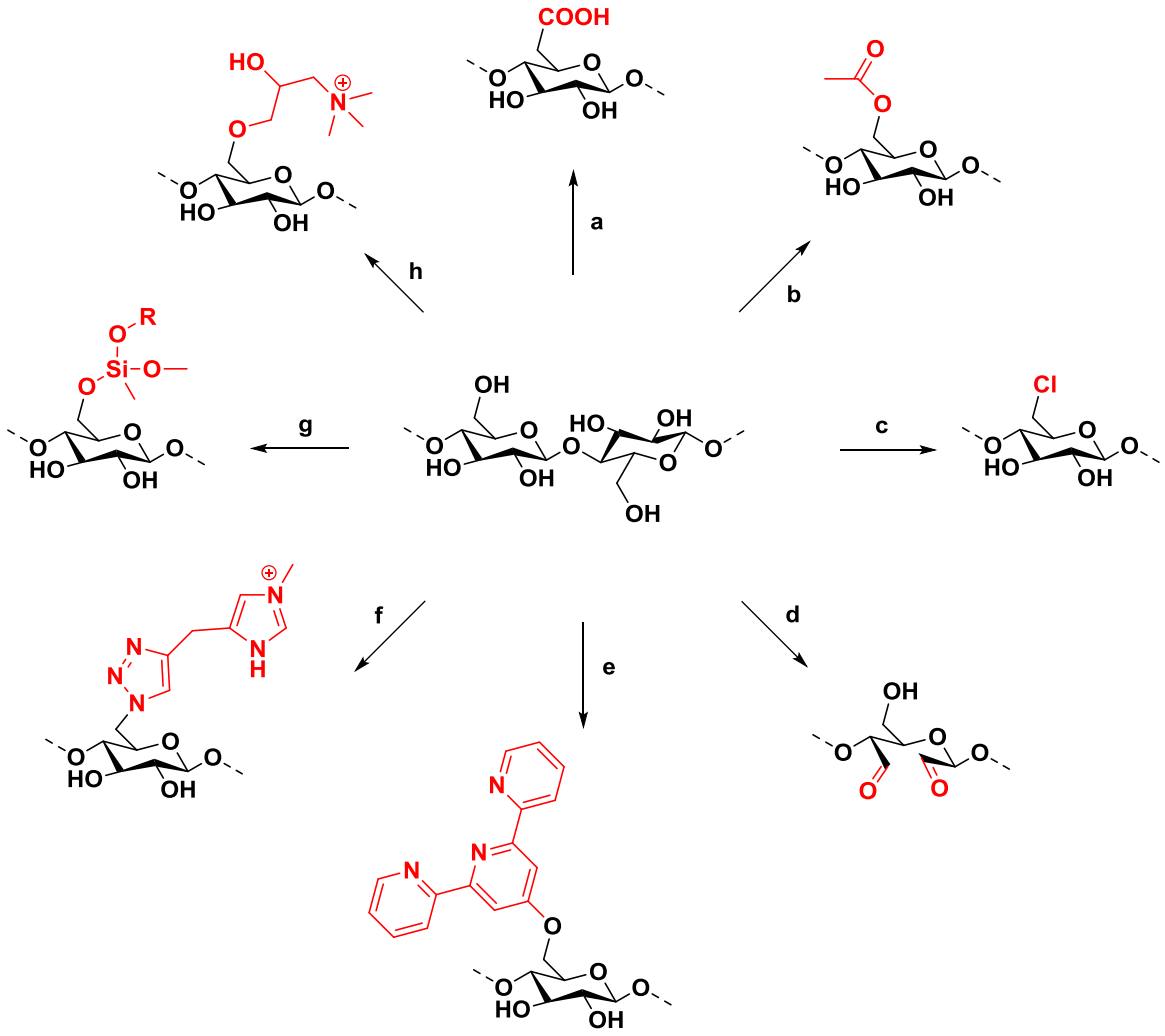


Figure 1.9 Scheme showing the chemical modifications at the surface of nanocellulose. **a)** TEMPO/NaOCl/NaBr **b)** acetic anhydride/acetic acid/toluene **c)** $SOCl_2$/pyridine **d)** $NaIO_4$ **e)** NaOH/4´-chloro-2, 2`:6`, 2``-terpyridine **f)** $SOCl_2$/pyridine; NaN_3/DMF; [MPIM][Br]/$CuSO_4$/NaAsc **g)** MTMS/HCl **h)** GTMAC/NaOH.

Recently, Kolmar and co-workers used the TEMPO-mediated oxidation of CNC as a means to introduce aldehyde functionalities on the surface of CNC, but only in a ratio of 1:19 to carboxylic acid groups.[42] By AFM measurements, these authors demonstrated that after TEMPO-mediated oxidation, the CNC preserved the form of whiskers and nanoscale dimensions (i.e. length of around 300 nm).[42] By contrast, acetylated CNC by treatment with acetic anhydride and acetic acid (**Figure 1.9**, **b**) resulted in a reduction of the diameters of the crystals, while the length of the nanowhiskers was preserved, as indicated by TEM and X-ray diffraction analyses.[35] This could indicate that the reaction proceeded from the surface of the nanoparticle to the crystal core and therefore resulted in damaging the crystallinity of the samples. In 2011, Eyley and Thielemans reported the chlorination of CNC by nucleophilic substitution of the primary hydroxyl group in the presence of thionyl chloride (**Figure 1.9**, **c**) without affecting the crystallinity of the sample as shown by X-ray diffraction analysis.[36] The halogenated CNC samples were then used as precursors for the introduction of imidazolium and tetrazole functionalities by azidation, followed by click chemistry (**Figure 1.9**, **f**). In addition, etherification *via* nucleophilic addition of deprotonated alcohol functional groups to EPTMAC (2,3-epoxypropyltrimethylammonium chloride) was also employed as a means to introduce cationic charges on the surface of CNC which resulted in the formation of a gel (**Figure 1.9**, **h**).[39] Finally, alkoxysilane chemistry was explored for the generation of hydrophobic and ultralight nanocellulose sponges by silylation of the alcohol functional groups (**Figure 1.9**, **g**).[40]

1. 3. 2 Nanocellulose – organic molecules nanocomposites

Different approaches were investigated for making novel hybrid materials of crystalline nanocellulose with organic fluorophores[43,44] or inorganic nanoparticles[45-47]. In 2007, Dong and Roman were among the first to describe fluorescently labelled cellulose nanocrystals (**Figure 1.10**, left) by reaction of the activated superficial hydroxyl groups with fluorescein-5'-isothiocyanate (FITC).[48] Fluorescently labelled cellulose nanocrystals bearing a positive charge (i.e. CNC-RBITC) can serve as bioimaging probes, as they were uptaken by human embryonic kidney (HEK-293) and *Spodoptera frugiperda* (Sf9) cells.[49] Due to their cationic charge of CNC-RBITC, these nanoparticles were found to penetrate the cellular membrane by making transient holes without substantial effect on the cell membrane integrity or indication of toxicity.[49]

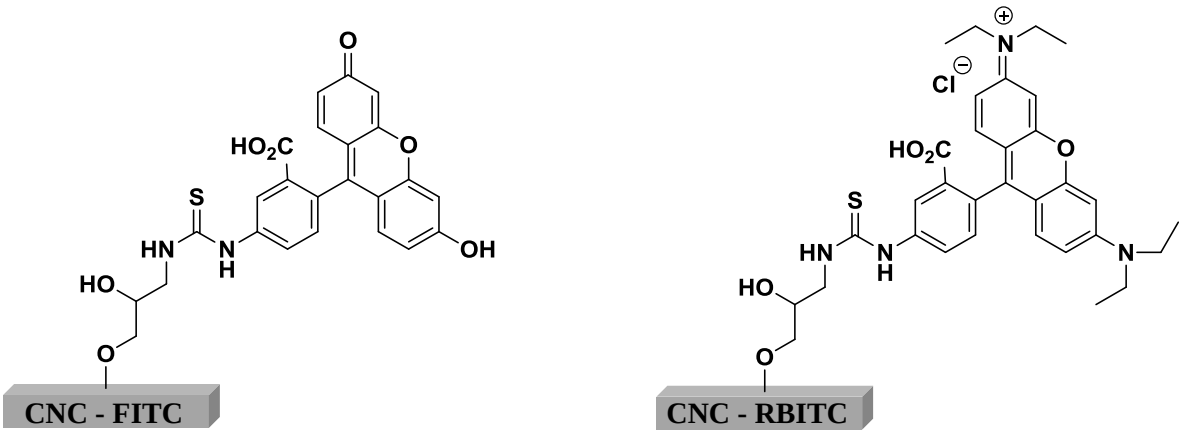


Figure 1.10 Illustration of the fluorescently labelled CNC: CNC-FITC (**left**) and CNC-RBITC (**right**).

Recently, the well-known sodium periodate oxidation of vicinal diols to dialdehyde functionalities (**Figure 1.9**, **d**) was employed for the covalent conjugation of CNC with Alexa Fluor dyes.[44] As shown in **Figure 1.11**, the aldehyde groups created on the surface of nanocellulose react with the amine functionalised Alexa Fluor derivative (AF546) by reductive amination.[44]

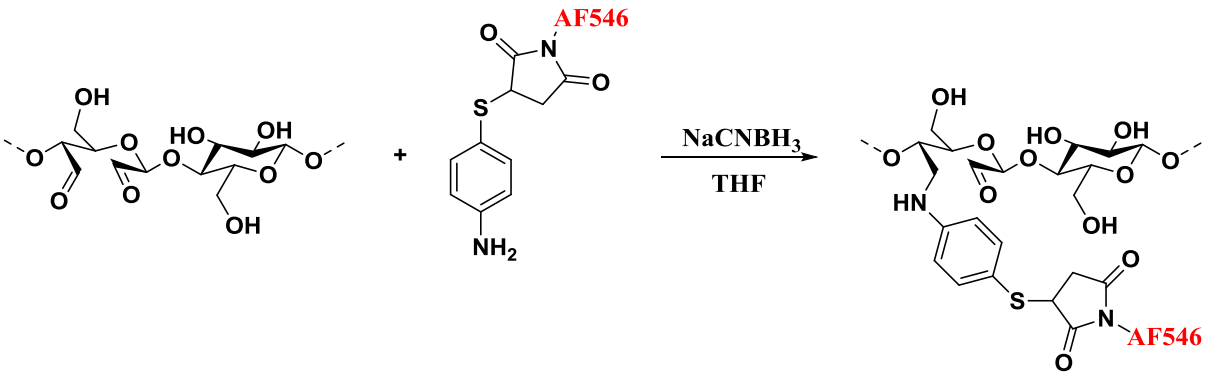


Figure 1.11[44] The aldehyde functionalities introduced on the surface of CNC by the sodium periodate oxidation of CNC react with the amine-containing Alexa Fluor derivative by reductive amination.

In addition, a hydrazide Alexa Fluor derivative (AF633) was attached at the reducing-ends of CNC *via* their respective hydrazone (**Figure 1.12**).[50] After being administered in mice it was observed that the fluorescently labelled CNC specifically targeted the bones possibly by the

interaction of Ca^{2+} in the bone matrix with the negatively charged surface of CNC.[50] This property of CNC paves the way for potential applications of CNC conjugates in biomedicine.

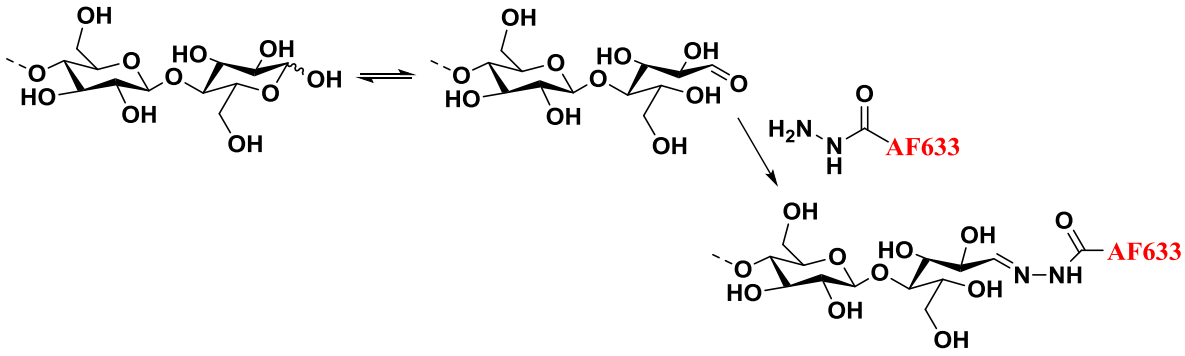


Figure 1.12 Hydrazide Alexa Fluor (AF633) was attached to the reducing ends of crystalline nanocellulose *via* hydrazone conjugation.

1. 3. 3 *De novo* hybrid materials: nanocomposites of proteins with nanocellulose

Cellulases, cellulose-degrading enzymes, consist of a catalytic domain and a cellulose-binding domain (CBD) connected by an extended polypeptide linker.[51] As shown in **Figure 1.13**, in cellulases, the role of CBD (yellow) is to bring the catalytic part of the enzyme in close proximity to the substrate surface. Following nature's example, Linder and co-workers emulated the concept of using multifunctional enzyme design, to interlink proteins to cellulose nanofibrils.[52] Their strategy relies on replacing the catalytic domain of cellulase with hydrophobins[53], while maintaining the polypeptide linker (**Figure 1.13**, right).

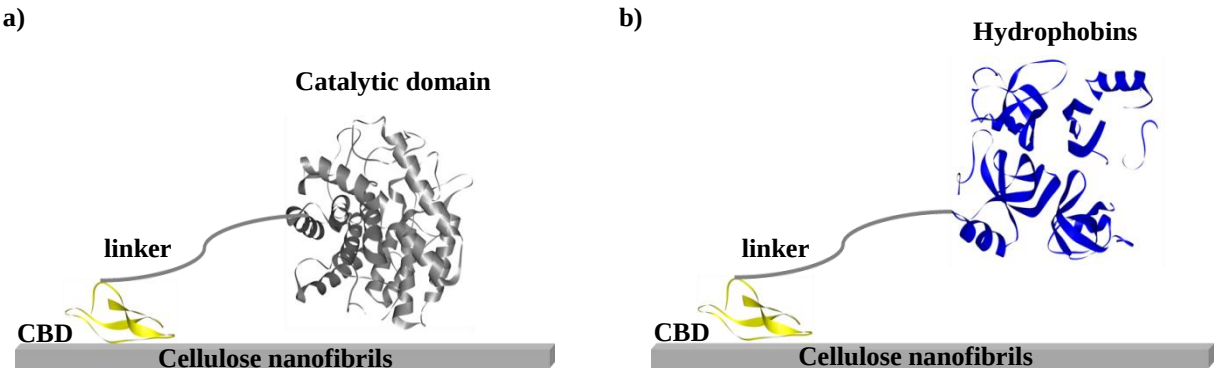


Figure 1.13 Schematic representation of the inter-linking of proteins with cellulose microfibrils. **a)** Prior to hydrolysis, cellulase binds to cellulose through a wedge-shaped domain (CBD, yellow, pdb: 2MWK[54]). **b)** The concept of coupling enzymes modules was used to create new fusion CBD-hydrophobin (pdb: 2FZ6[55]) proteins for inter-linking cellulose nanofibrils.

Hydrophobins are amphiphilic proteins secreted on the spores of filamentous fungi where they serve as surfactants and adhesive agents.[56] Owing to their self-assembly at interfaces and amphiphilic structure they form stable foams and films. In an effort to build a model system for molecular biomimetic materials based on nanocellulose, fusion proteins were absorbed onto NFC.[52] The bi-functional fusion proteins (CBD-HFB) consist of one hydrophobin part (HFB) connected to two cellulose binding domains (double CBD) from *Trichoderma reesei* (**Figure 1.14**, **a**). Nanocellulose films were then prepared by drying the aqueous dispersions of fibrils and proteins (i.e. CBD-HFB) in water. The authors outlined a model in which the proteins would act as an adhesive matrix between the nanocellulose fibrils. As shown by the schematic representations in **Figure 1.14**, there are two possibilities by which the molecular cross-linking of the NFC network is directed by the CBD-HFBI protein complex. The first model proposes that the

CBD-domains attach to the NFC surface and an inter-fibril linkage is then mediated by the hydrophobin oligomerization (**Figure 1.14**, **b**). The second model suggests that the molecular cross-linking of the NFC fibrils occurs through CBDs (**Figure 1.14**, **c**). It turned out that this novel hybrid material displays superior mechanical properties to pristine nanofibrils. The CBD-HFB-NFC hybrid films absorb more water than the original NFC films; however, the absorbed water is mainly localized on the surface of the proteins. In the case of NFC films, the water is localized in between the inter-linked nanofibrils. Specifically, the inter-locking of proteins led to increased resistance of the hybrid material to plastic deformation or irreversible transformation in the entangled structure caused by strain.

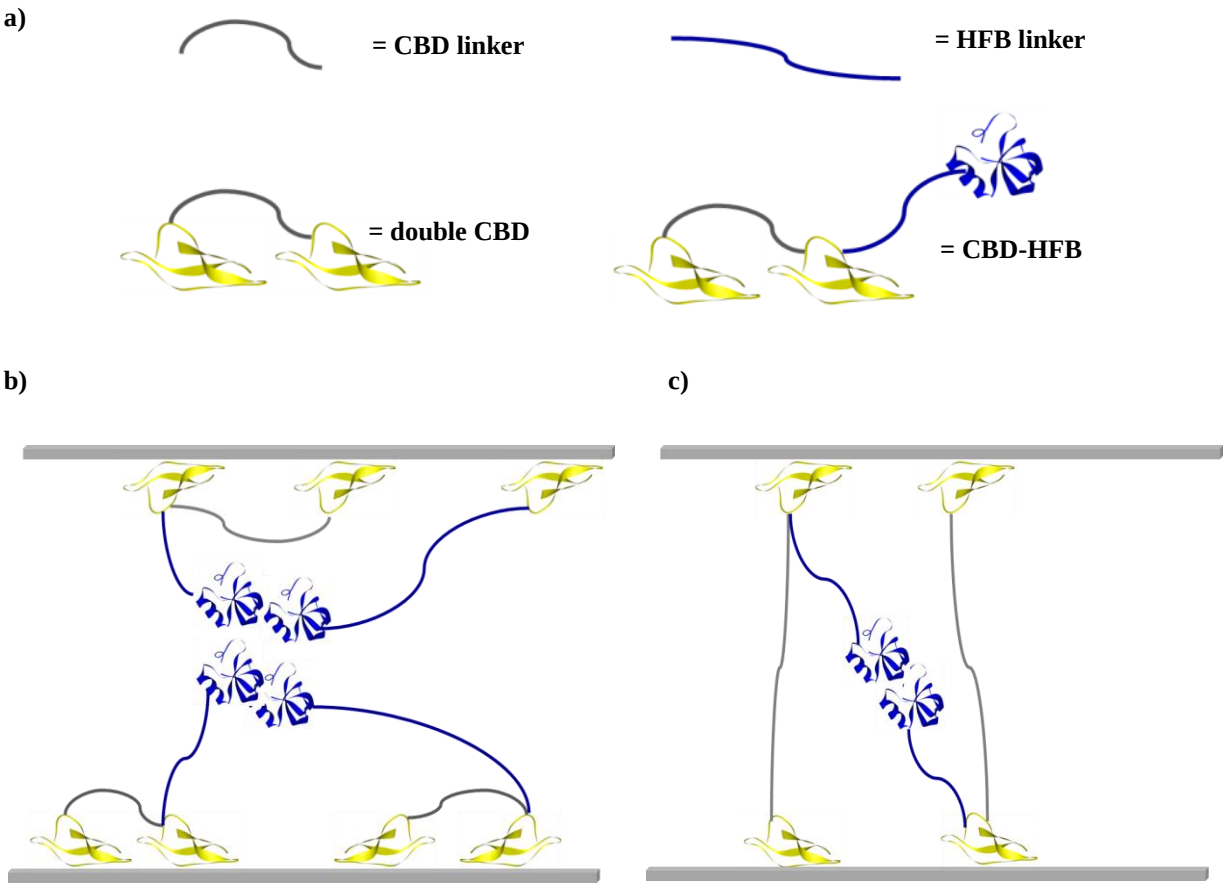


Figure 1.14 Proposed molecular structures of the NFC fibrils inter-linked with the CBD-HFB proteins. **a)** Model of the engineered bi-functional CBD-HFB complex.[57] **b)** The first possibility envisions that both CBDs bind to the same NFC fibril, while the hydrophobins interact with each other. **c)** The second possibility is that the molecular cross-linking of the NFC fibrils occurs through CBDs, while hydrophobins interact as in the first model.

Investigation of the properties of nanocellulose/protein hybrids is an exciting research area just at the beginning of its exploration. It therefore requires the development of selective surface chemical methods to govern the specifics of recognition between the proteins and nanoparticles. Currently, the covalent linkage of proteins on cellulose nanocrystals relies, as indicated in most published examples[58], on amide bond formation between the nucleophilic lysine residues on the surface of proteins and activated carboxylic groups introduced at the surface of nanocellulose. Specifically, by using the EDC (carbodiimide) and sulfo-NHS (sulfo-*N*-hydroxysuccinimide) crosslinking method, Linder and co-workers demonstrated the attachment of alkaline phosphatase (AP) and anti-hydrocortisone antibody to the surface of nanofibrilated cellulose (NFC) as shown in **Figure 1.15**, **a**.[59] In addition, the authors described two other strategies for the covalent immobilization of AP relying on epoxy amination (**Figure 1.15**, **b**) and arylhydrazone conjugation (**Figure 1.15**, **c**).

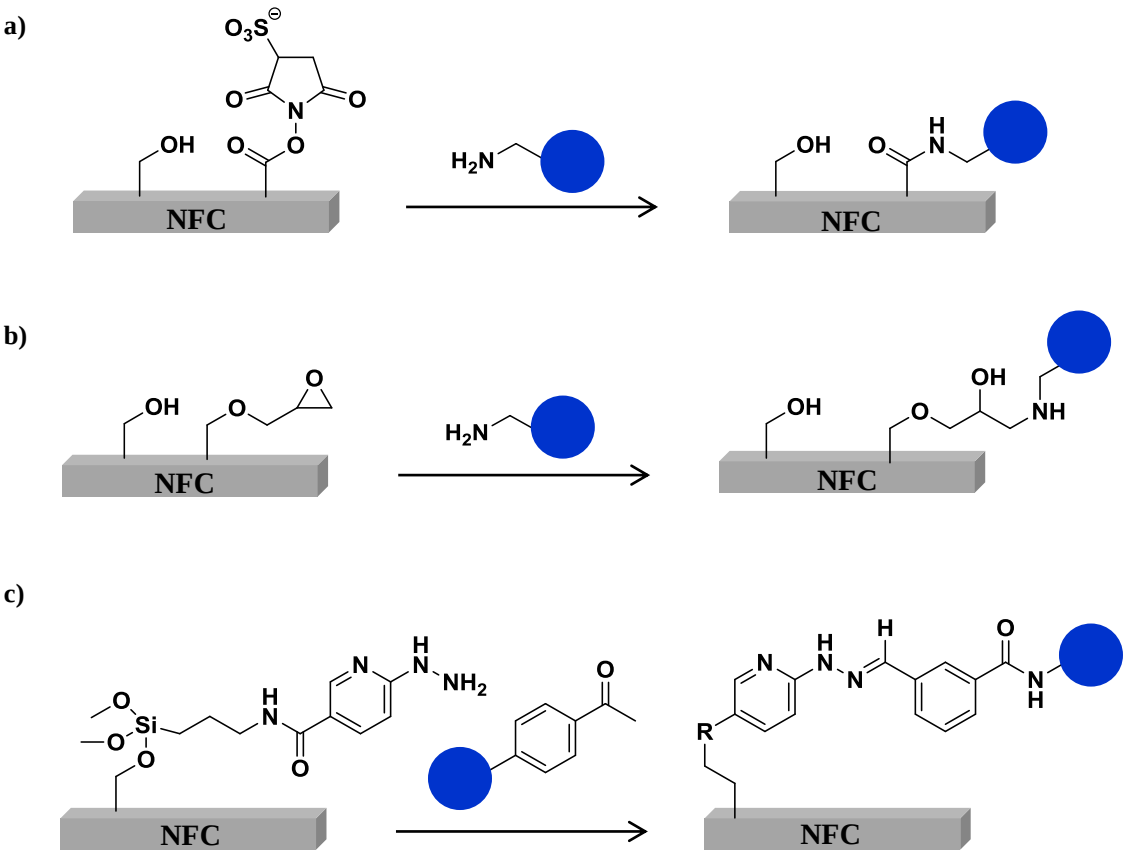


Figure 1.15 Chemical strategies for the covalent linkage of alkaline phosphatase (AP) with nanofibrilated cellulose (NFC). **a)** The carboxylic acid functionalities introduced by TEMPO-oxidation at the surface of NFC were activated with EDC/sulfo-NHS prior to reaction with lysine residues on the surface of AP. **b)** Reaction of lysine residues on the surface of AP with epoxy-NFC by a ring opening reaction. **c)** Aldehyde functionalities

were inserted on the surface of AP by reaction of the lysine residues with succinimidyl-4-formylbenzoate. Thereafter, they were conjugated to silylated NFC containing aryl hydrazine groups.

Although the aforementioned approaches do not require extensive protein modification steps, they do suffer from poor selectivity and therefore prohibit the construction of defined composites. On the other hand, the authors claim that the conjugation of AP to NFC by multi-point cross-linking resulted in increased stability of the enzyme.[59] However, attachment of proteins on surfaces through several lysine residues also leads to protein inactivation.[60] For example, in a separate study, Kolmar and co-workers compared the activity of galactose oxidase (GOase) when immobilised on CNC *via* the unspecific EDC/NHS coupling and through single-point attachment.[42] They observed a reduced activity of the galactose oxidase (GOase) when the enzyme was attached by multiple lysine residues (i.e. *ca.* 60% of that achieved by GOase which was specifically attached).[42]

Another approach available for making protein-nanocellulose conjugates employs the use of gold nanoparticles. Gold nanoparticles possess a high surface to volume ratio[61] and therefore novel hybrid nanocomposites of CNC functionalised with gold nanoparticles (AuNPs) could serve as a support for enzyme immobilization with high loading.[62] The developed strategy is outlined in **Figure 1.16**. AuNPs/CNC conjugates were synthesised by reduction of gold (III) chloride trihydrate with sodium borohydrate in a suspension of CNC.[63] The stable colloids were thereafter reacted with lipoic acid, resulting in thioctic acid-derivatized gold-nanocellulose clusters. Activated esters were introduced by chemical modification of the ligand terminus. The following proteins: cyclodextrin glycosyl transferase (CGTase), alcohol oxidase and glucose oxidase (GOx)[64] were covalently coupled to the AuNPs/CNC hybrids *via* interaction of the surface exposed lysine residues with the activated esters.

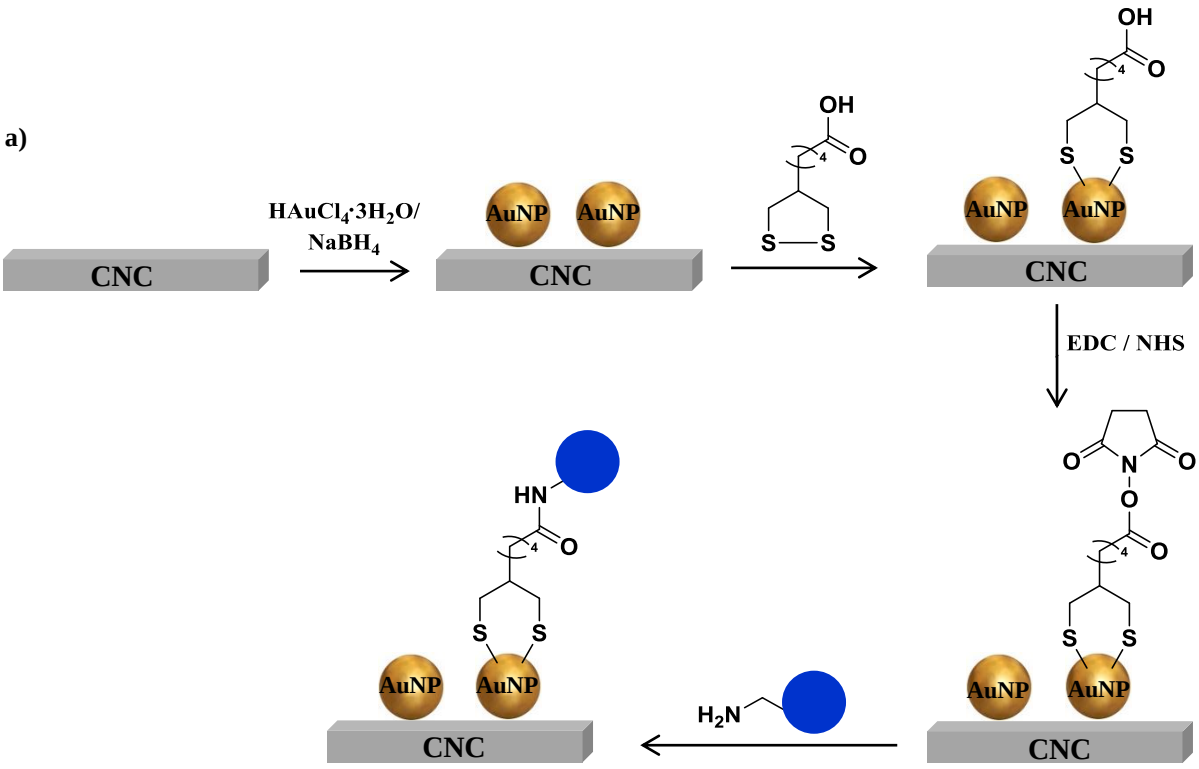


Figure 1.16 Covalent immobilization of alcohol and glucose oxidases onto CNC-AuNPs conjugates. Reaction of lysine residues on the surface of CGTase or GOx with activated NHS esters.

A lengthy and elaborate synthesis aiming the attachment of *β*-casein at the reducing ends of crystalline nanocellulose has been worked out by the group of Kadla.[65] In this process, the copper(I)-catalysed azide-alkyne cycloaddition or copper 'click' chemistry[66] was employed for the linkage of alkyne functionalised *β*-casein to azide functionalised reducing ends of CNC (**Figure 1.17**). This method necessitates extensive chemical modification on both nanocellulose and *β*-casein, long reaction times and synthesis of additional linkers between the protein and nanoparticles.

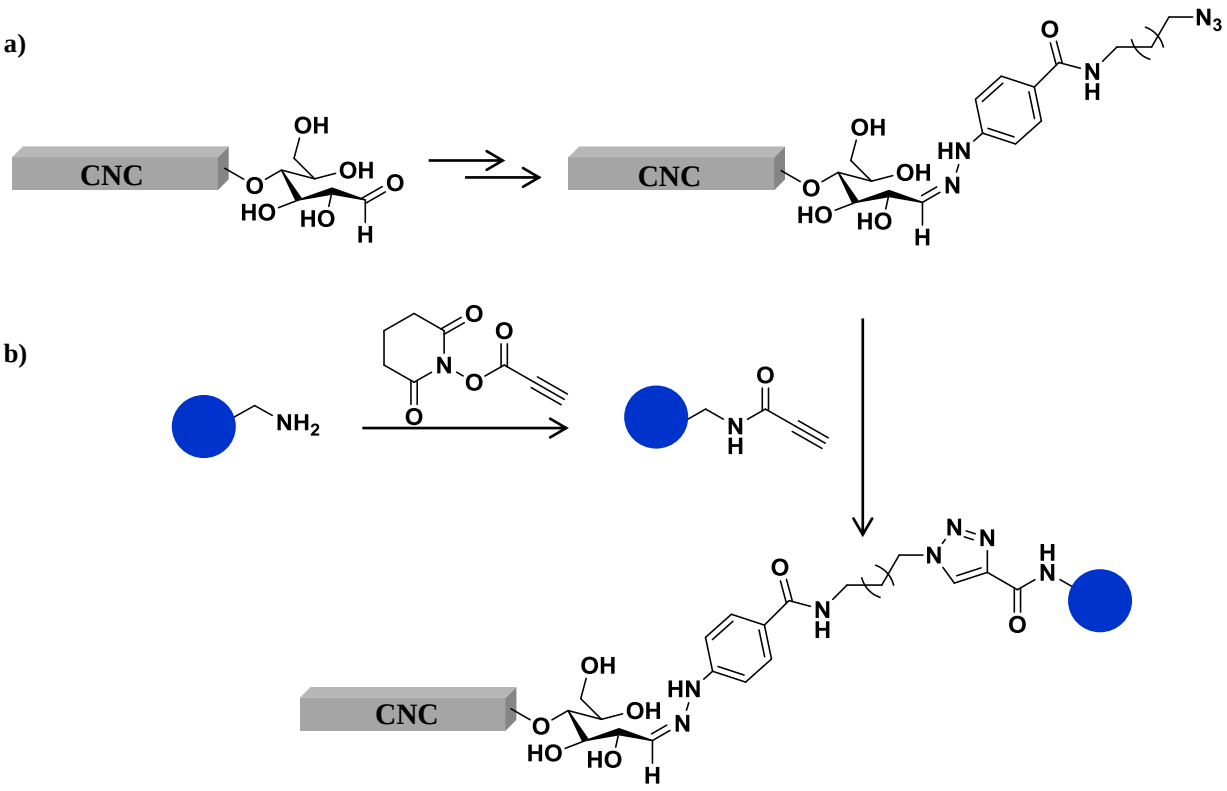


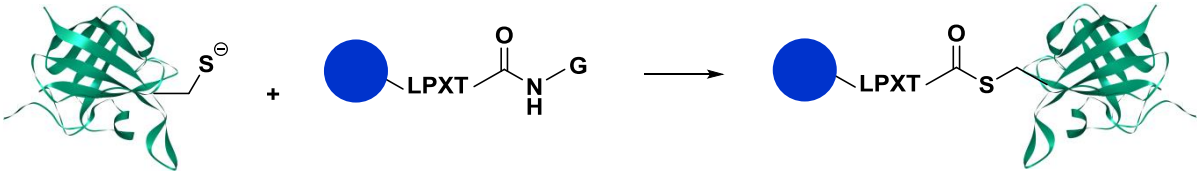
Figure 1.17 Covalent immobilization of *β*-casein at the reducing ends of CNC. **a)** Chemical functionalisation of CNC at the reducing ends by hydrazone conjugation with *p*-hydrazinylbenzoic acid, followed by activation of the carboxylic acid group with EDC/NHS to introduce azide functionalities. **b)** Reaction of lysine residues on the surface of *β*-casein with activated NHS esters to introduce terminal alkyne groups. Conjugation of the modified *β*-casein with CNC by the 'click' reaction.

The abovementioned methods for constructing protein-nanocellulose conjugates by covalent interactions have in common disadvantages such as the generation of heterogeneous mixtures of conjugates due to random orientation of the immobilized proteins on the surface of CNC and inactivation of the linked proteins. Furthermore, as these methods feature modifications of the amino-bearing residues (i.e. lysines) another disadvantage is the change in the overall surface electrostatic charge required for maintaining protein folding, which could affect the stability of the modified proteins.[67] Therefore, there is a need for developing new approaches which enable chemoselective conjugation of proteins to nanocellulose.

A chemoselective strategy for covalently immobilizing proteins onto crystalline nanocellulose was described by Kolmar and co-workers.[42] The described chemoenzymatic method combines the selective oxime-ligation used for decorating the surface of CNC with peptide linkers, with coupling of recombinant proteins by the sortase-mediated protein ligation.[68] Sortase A (SrtA), a

transpeptidase from *Staphylococcus aureus*, cleaves surface proteins at a LPXT sequence near the *C*-terminus (**Figure 1.18**, **a**).[69] The protein acyl fragment is then transferred to a co-substrate amine of a pentaglycine peptide of the peptidoglycan layers of *S. aureus* (**Figure 1.18**, **b**).[69] The transpeptidation reaction catalysed by sortase, has inspired chemical biologists to use this catalyst as an *in vitro* tool for making protein-peptide[70], protein-sugar[71] conjugates or linking drugs to therapeutic antibodies (ADCs)[72].

a)



b)

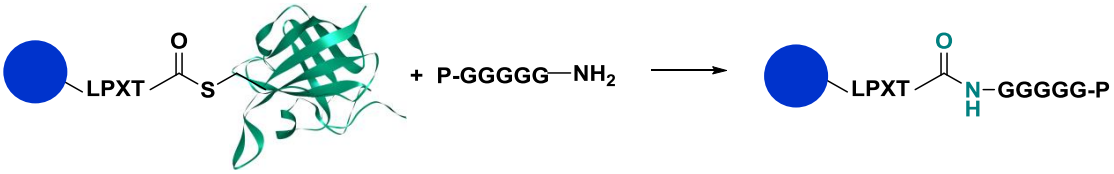


Figure 1.18 Schematic representation of the catalytic steps performed by sortase *S. aureus* (pdb: 1IJA[73]). **a)** Nucleophilic attack by cysteine thiolate (Cys_{184}) results in cleavage of the Thr-Gly bond within the LPXTG sorting motif. **b)** Nucleophilic attack by the terminal amine of the Gly residue results in amide bond formation between the protein bearing the sortase motif and the penta-glycine peptide.

When sortase-mediated ligation was employed for the selective assembly of proteins with nanocellulose, the following steps were performed. Firstly, synthetic peptide linkers bearing the penta-Gly sequence and aminooxy moiety at the side-chain of a lysine residue were conjugated *via* oxime-ligation to the surface of crystalline nanocellulose. Prior to the peptide attachment, the surface of CNC was decorated with aldehyde and carboxylic acid groups by the TEMPO-mediated oxidation (**Figure 1.19**, **a**). Secondly, the sortase recognition motif (i.e. LPETG) was introduced at the *N*- or *C*-terminus of recombinant proteins by cloning manipulation (**Figure 1.19**, **b**). Finally, the sortase-tagged proteins namely green fluorescent protein (tGFP), antigen-binding shark vNAR and galactose oxidase (GOase) were site-directly linked to the surface of nanocellulose by the calcium dependent sortase A. The CNC-protein conjugates were analysed by flow cytometry and fluorescence microscopy. Accordingly, it was proven that tGFP was selectively conjugated to the functionalised CNC as "the fluorescent signal of CNC-tGFP is about

100 times stronger than that of unmodified CNC". On the other hand, the analysis of the shark vNAR-CNC conjugates by use of fluorescently labelled antibodies proved challenging due to the tendency of lysozyme to "form non-specific interactions" with CNC. Arguably, this could have occurred through electrostatic interactions between the negatively charged carboxy-CNC surface and the six lysine residues on the surface of lysozyme. Interestingly, the direct coupling of the tagged proteins with the TEMPO-oxidised material was reported in the aforementioned article to have occurred "to a very low extent". This assumption is considerably qualitative and cannot be interpreted based on the reported flow-cytometric data. Nevertheless, this generic method is powerful as it could be extended to immobilise any recombinant protein bearing the sortase recognition sequence.

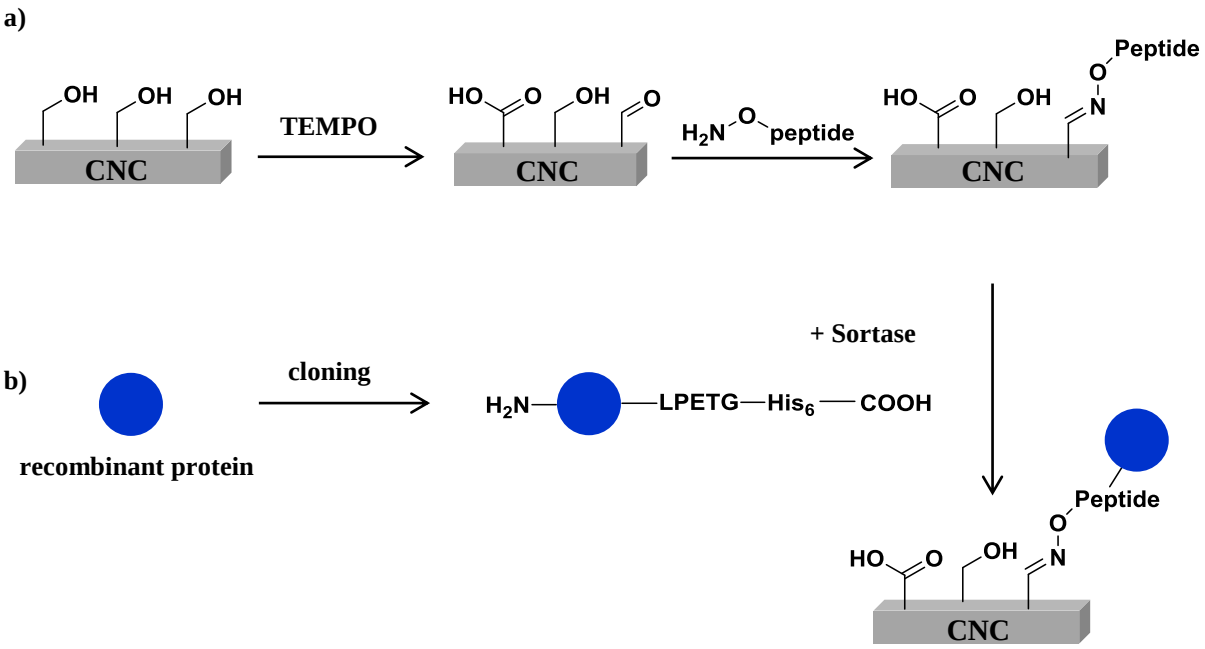


Figure 1.19 A chemoenzymatic strategy for linking recombinant proteins to CNC. **a)** Schematic representation of the TEMPO-mediated oxidation resulting in the introduction of aldehyde functionalities on the surface of CNC. Peptide linkers bearing aminooxy groups were selectively conjugated to CNC by the oxime ligation. **b)** Sortase recognition motifs were introduced at the *N*-termini of proteins by cloning manipulations. Recombinant proteins were coupled to CNC-peptide by the sortase-mediated ligation.

1. 4 Formylglycine generating enzyme as a tool for protein engineering

When nanocellulose is decorated with electrophilic functional groups such as aldehyde or carboxylic acids, specific conjugation to proteins is challenging due to interactions with the lysine residues on proteins surface. We envisioned a reverse strategy. Our perspective entails the decoration of nanocellulose surface with nucleophilic functionalities which interact selectively with electrophilic moieties introduced on proteins. One way to introduce electrophilic amino acids with molecular accuracy is by using the formylglycine generating enzyme (FGE).[74]

In bacteria, archaea and eukaryotes, formylglycine generating enzyme (FGE) catalyses a unique co- or post-translational modification of sulfatases. FGE recognises a short amino-acid motif, CXPXR, located 50-80 residues away from the *N*-termini of sulfatases, while the 'sulfatase polypeptide is largely unfolded' and selectively converts the cysteine residue.[75] More specifically, FGE catalyses the oxidation of the cysteine residue within the recognition sequence to a C_{α}-formylglycine (FGly) amino acid derivative. FGly is an essential catalytic residue in the active site of sulfatases, where it is present in its hydrated form.[75] In humans, mutations of the gene encoding for FGE, resulted in the inability to generate FGly residues on sulfatases, which has been associated with a rare genetic disease - multiple sulfatase deficiency (MSD).[76,77]

The discovery of FGE has prompted the establishment of two research areas. The first area concerns the understanding of the complex mechanistic pathway catalysed by this enzyme at a molecular level and its intriguing activation of oxygen without the use of a co-factor.[78] Concomitantly, the second area addresses the FGly residue as an unprecedented tool for selective protein labelling. This area flourished with the discovery that FGE can insert the aldehyde functionality, which is singular in the chemical repertoire of the natural amino acids, *in vivo* into recombinant proteins bearing the sulfatase recognition sequence.[79] Consequently, the potential applications are diverse and range from protein-protein[80], hydrazide-biotin[81], glycan aminooxy[82] and PEG-protein to antibody-drug conjugates.[83]

We anticipated that a chemoenzymatic method mediated by FGE for selectively conjugating proteins bearing aldehyde functionalities to crystalline nanocellulose could be developed (**Figure 1.20**). An insight into this method was provided, starting from simple glucose-aldehyde containing proteins conjugates to engineered orthogonality in supramolecular systems.

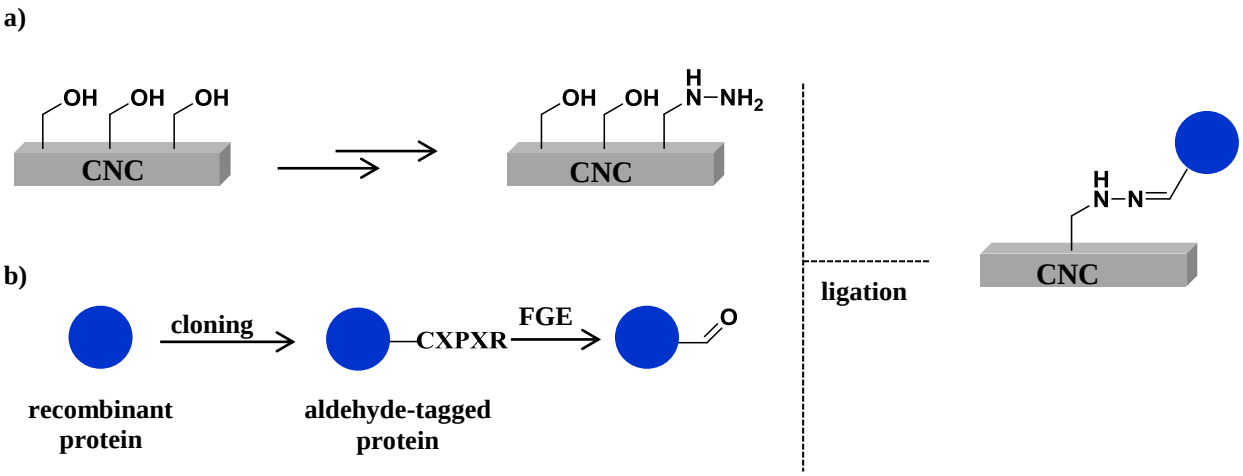


Figure 1.20 A chemoenzymatic strategy for linking recombinant proteins to CNC. **a)** Schematic representation of the chemical functionalisation of CNC resulting in the introduction of hydrazine functionalities on the surface. **b)** FGE recognition motifs were introduced at the *N*-termini or internal sites of recombinant proteins by cloning manipulations. Aldehyde-tagged proteins were converted *in vitro* by FGE prior to conjugation to CNC *via* hydrazone ligation.

1. 5 Aims of the thesis

The aim of this thesis was to elaborate a chemoselective method for the conjugation of recombinant aldehyde-tagged proteins to chemically functionalised crystalline nanocellulose.

The starting point of the theme is shown in Chapter 2 and comprises the development of a facile *in vitro* assay for the site-specific tagging of recombinant proteins at the *N*-termini and internal sites by using FGE.

The successful regioselective synthesis of a hydrazine glucose derivative and its chemoselective conjugation to aldehyde-tagged proteins is described in Chapter 3.

Chapter 4 emphasises the immobilization of fluorescent proteins at the surface of the hydrazine-functionalised nanocellulose. Detailed characterization of the protein-nanocellulose nanomaterials proves the chemoselectivity of the method.

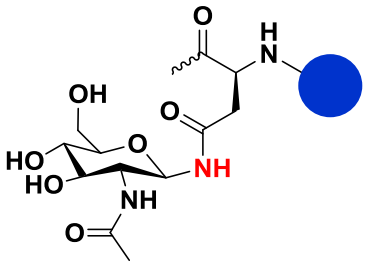
The application of this method is described in Chapter 5, where active enzymes and a human antibody were immobilised on crystalline nanocellulose in a site-specific fashion. The results obtained in this thesis demonstrate that this selective method is valuable for conjugating proteins to crystalline nanocellulose with atomic precision. It paves the way for many exciting applications in biomedicine including the development of novel tailored hybrid materials and the site-specific immobilization of recombinant proteins.

Chapter 2 - *In vitro* activity of FGE on aldehyde-tagged proteins

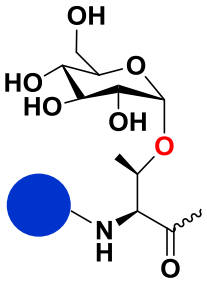
2. 1. 1 Glycosylation of proteins

Post-translational modifications (PTMs) are fundamental for broadening proteins structure and their biological function. Protein glycosylation or covalent attachment of sugar moieties to proteins, is ubiquitous in eukaryotes[84] where it plays essential roles in signalling[85], hormone action[86] and cancer progression[87]. Phosphorylation[88], acetylation[89] and methylation[90] expand the natural amino acids repertoire and dynamically trigger new functions.

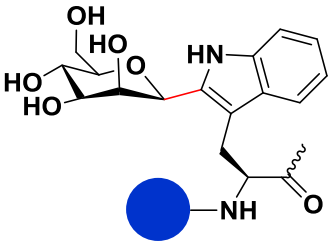
Post-translational modifications not only create diversity, but also provide us with new tools to site-specifically modify proteins.[91] As most therapeutically relevant proteins contain PTMs such as glycosylation to ensure proper cellular targeting[92], it is of great interest to emulate PTMs *in vitro*. There are five main different modes by which in nature carbohydrates are site-specifically attached to proteins or lipids: *N*-linked (by which the glycan is attached to the amide side chain of an asparagine residue, **Figure 2.1**, **a**), *O*-linked (by which the glycan is linked to the hydroxyl group of a serine or threonine residue, **Figure 2.1**, **b**), *C*-mannosylation (by which the carbohydrate is linked to a tryptophan residue, **Figure 2.1**, **c**), glypation (when a glycosylphosphatidylinositol tag is attached at the *C*-termini of proteins, **Figure 2.1**, **d**) and phospho-serine glycosylation (by which the glycan is connected to a serine residue by a phosphate group, **Figure 2.1**, **e**).[84,91]



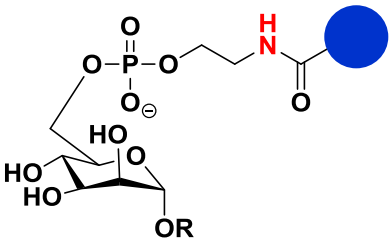
a) ***N*-glycan linkage**



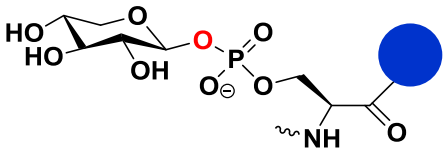
b) *O*-glycan linkage



c) *C*-glycan linkage



d) Phosphate-*N*-glycan linkage



e) Phosphate-glycan linkage

Figure 2.1 Representative examples of the five different covalent linkage types of sugar-amino acids in protein glycosylation. **a)** *N*-acetyl-glucosamine-asparagine. **b)** Glucose-threonine **c)** Mannose-tryptophan **d)** Glycosylphosphatidylinositol (GPI) - *C*-terminus of proteins **e)** Xyltol-phosphate-serine.

Nature's protein-sugar conjugates unveil challenging chemical diversity. The difficulty in deciphering the function of glycosylation arises therefore from the combination of the structural diversity of proteins with the intrinsic complexity of carbohydrates. For this reason, chemical and enzymatic tagging methods that duplicate natural structures have been developed.[93] Among them, one major strategy is the chemical ligation between tagged proteins and glycan (**Figure 2.2**). Site-specific chemical ligation relays on site-directed gene mutagenesis of tagged proteins and synthetic chemistry required for the modification of sugars.

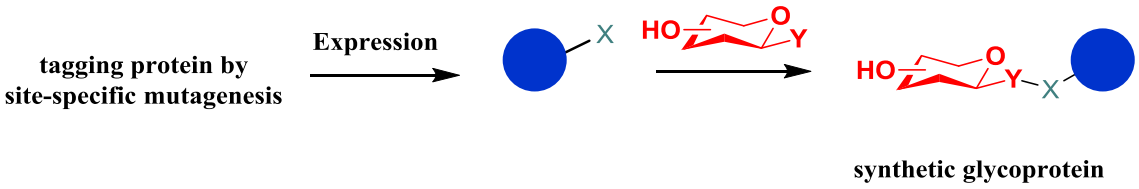


Figure 2.2[93] Chemoselective ligation between the tagged protein and carbohydrate. A natural or unnatural amino-acid is incorporated by site-specific mutagenesis at the *locus* of protein tagging. The tagged protein is then conjugated with the chemically modified carbohydrate in a site-specific fashion.

2. 1. 2 Creating synthetic glycoproteins

We would like to explore the chemoselective ligation approach used for mimicking glycosylation for the specific attachment of recombinant proteins to nanocellulose. One of the first examples of this combined approach was developed by Davis *et al*.[94] According to their method, natural cysteine (Cys) and unnatural amino acids such as azidohomoalanine (Aha) were incorporated in the structure of *β*-galactosidase and subtilisin, respectively at glycosylation sites. Chemoselective ligation reactions such as azide-alkyne Huisgen cycloaddition (**Figure 2.3**) and disulfide exchange (**Figure 2.4**) were then performed.

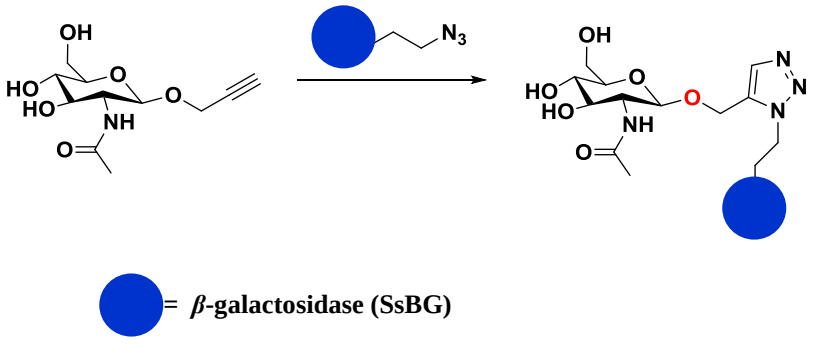


Figure 2.3 Chemoselective conjugation of sugar derivative of *N*-acetylglucosamine (GlcNAc-*O*-CH_2-alkyne) with unnatural amino-acid azidohomoalanine (Aha).[94]

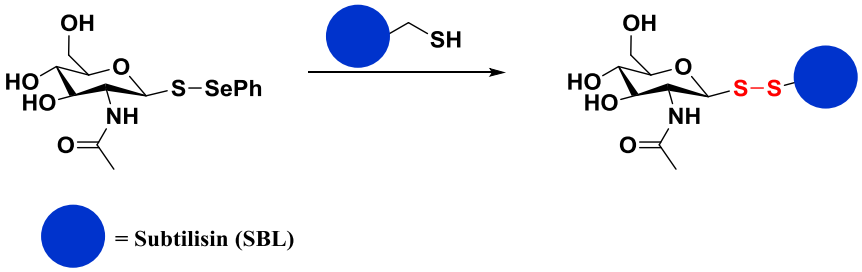


Figure 2.4 Site-specific glycoconjugation of cysteine with synthetically modified *N*-acetylglucosamine-thiol (GlcNAc-S-SePh).[94]

In 2003, FGE was assigned as the enzyme catalysing a novel co- or post-translational modification in sulfatases.[95] Driven by the chemoselective (mutually compatible) aldehyde functionality introduced by FGE, Bertozzi *et al.* pioneered the use of FGE as a tool for site-selective modification of proteins.[79] They designed a method known as the 'aldehyde-tag system' which adopts the *in vivo* co-expression of recombinant proteins with an additional prokaryotic FGE from *Mycobacterium tuberculosis*. Specifically, the "aldehyde-tag system" was used for generating a variety of protein-conjugates such as: biotin hydrazide and aminooxy polyethyleneglycol (AO-PEG) with maltose binding protein (MBP) and green fluorescent protein (GFP), aminooxy Alexa Fluor 488 with human immunoglobulin (hIgG).[81] Among them, of significant importance is the generation of homogeneous glycosylated proteins where the 'genetically encoded aldehyde tag' served as a means of site-specific conjugation. To a well-established therapeutic, the human growth hormone (hGH), a *C*-terminal aldehyde tag was appended by cloning manipulation. The aldehyde-tagged hGH was produced by co-expression with FGE *M. tuberculosis*, purified and then conjugated to synthetic aminooxy glycans such as 2, 3-sialyllactose by oxime ligation (**Figure 2.5**).[82] The oxime-linked glycan resembles structurally the natural *N*-glycan linkage (**Figure 2.1**).

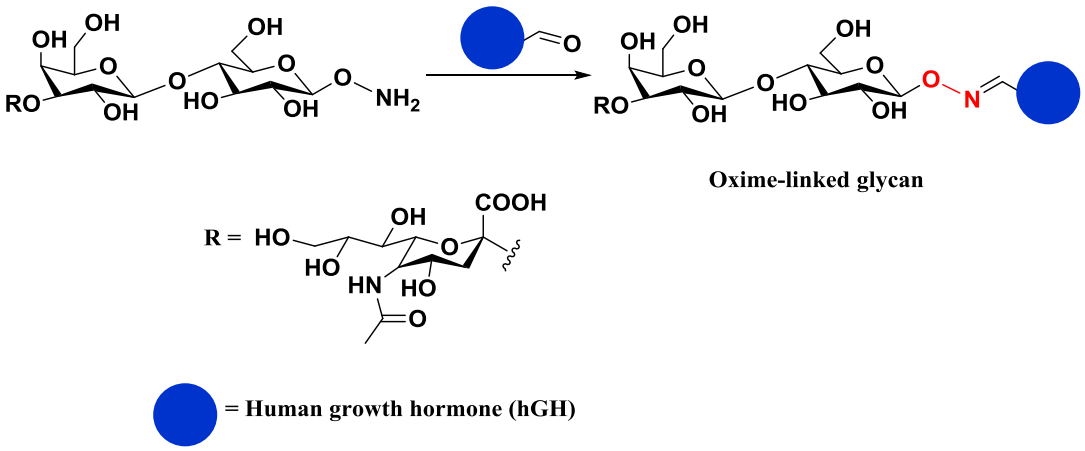


Figure 2.5 Chemoselective ligation of synthetic aminooxy glycan (2,3-sialyllactose) with aldehyde containing human growth hormone (hGH) to yield oxime-glycosylated conjugates.[82]

However, it was observed that by co-expression, the FGly conversion yields may vary between different protein substrates.[80,96,97] One reason could be the low expression levels of FGE *M. tuberculosis* when co-expressed with maltose binding protein (MBP)[80], EcMutS[96] and aldehyde-tagged Fc fragment of IgG1[97]. Another reason for low yielded conversions is the instability of the aldehyde functionality which could be modified *in vivo* by other cellular enzymes[97] (such as the aldehyde dehydrogenase from *E. coli* which catalyses the oxidation of acetaldehyde to acetyl-CoA using NAD^+ and coenzyme A[98]). Incomplete conversion by FGE *in vivo*, leads to heterogeneous mixtures of labelled proteins which require additional purification steps. The holy grail of protein labelling is to achieve a chemoselective method which can be adapted for homogeneous modification of any recombinant protein.

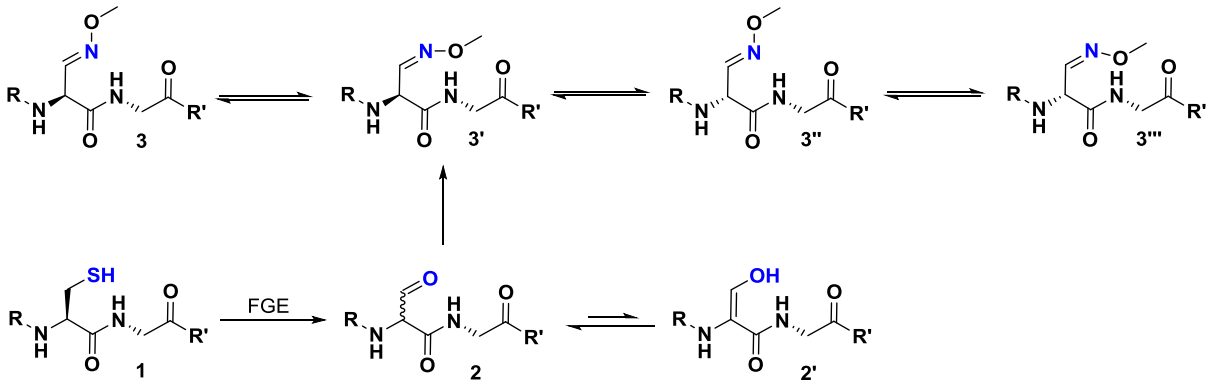
In our lab, it was found that *in vitro* addition of FGEs from *Thermomonospora curvata* and *Mycobacterium smegmatis* with copper (I) increases the enzymatic catalytic efficiency more than 20-fold on peptide substrates.[99] Moreover, a mutant of FGE (i.e. FGE-4C), in which all the cysteine residues outside the active site were mutated (i.e. C187A, C231A, C284S, C298A = FGE-4C) displays 10-fold higher catalytic efficiency than the wild type.[99] The combination of the copper pre-treatment with FGE-4C, makes Cu(I) - FGE-4C a promising catalyst (with 200-fold increased activity) for modifying aldehyde-tagged proteins *in vitro*. We therefore took advantage of the superior catalytic performances of Cu(I) - FGE-4C to build a system which would allow fast *in vitro* insertion of the aldehyde functionality on recombinant proteins. *Ab initio*, we explored the enzyme function for oxidising *N*-terminal aldehyde-tagged recombinant proteins; then, we expanded the use of FGE on internally inserted aldehyde-tags.

2. 2 Results

2. 2. 1 *In vitro* activity of FGE-4C on *N*-terminal aldehyde-tagged proteins

To demonstrate the *in vitro* efficiency of FGE-4C *T. curvata* on recombinant aldehyde-tagged proteins, we firstly tested the conversion on a peptide bearing the sulfatase motif from *Thermomonospora curvata*: GTPLCGPSR. The peptide was first incubated with FGE-4C and then treated with methoxyamine, a reagent which selectively reacts with the aldehyde functionality in the peptide (**Figure 2.6**, **a**). The cysteine to formylglycine conversion was monitored by LC-MS (**Figure 2.6**, **b**) and the mass corresponding to the oxime-peptide was confirmed (calc. *m/z* 898.5, obs. *m/z* 898.5). The diastereomeric peptides peaks (**3**, **3`**, **3``**, **3```**) may suggest loss of the chiral information at the C_{α}-formylgylcine centre due to aldehyde-enol (**2**, **2`**) tautomerisation under acidic conditions.

a)



b)

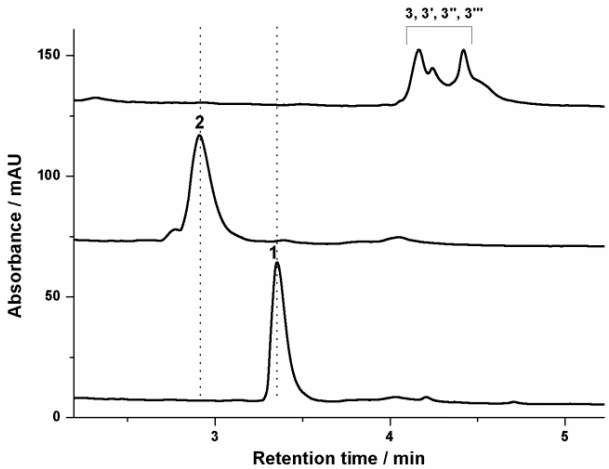


Figure 2.6 FGE-4C activity on peptide substrate. **a)** Reaction scheme of thiol-peptide with FGE, followed by labelling with methoxyamine. **b)** HPLC chromatograms of thiol-peptide (**1**), aldehyde-peptide (**2**) and mixture of oxime-peptide diastereoisomers (**3**) at 230 nm.

We then constructed a general pET28 expression plasmid (ald-gfp) that inserts the recognition motif (GTPLCPSR) for FGE at the *N*-terminus and the His_6-tag at the *C*-terminus. Using this vector we recombinantly produced aldehyde-tagged protein versions of green fluorescent protein (Ald-GFP), lumazine synthase from *Saccharomyces cerevisiae* (Ald-YLS), lumazine synthase from *Aquifex aeolicus* (Ald-ALS) and small-laccase (Ald-SLAC, see **Chapter 5**). Insertion of the aldehyde-tag at the *N*-termini of the aforementioned proteins did not affect the production yields of these proteins in *Escherichia coli*, up to 15 mg of purified proteins were obtained per litre culture.

With the aldehyde-tag we also introduced an arginine residue which is a trypsin cleavage site at the *N*-termini of the proteins. Consequently, we demonstrated the chemospecificity of the FGE-4C oxidation on the designed protein substrates by tryptic digestion. In a test reaction, 100 µM of Ald-YLS substrate was converted by 10 µM of Cu(I) - FGE-4C in 2 h at rt; addition of trypsin to the protein mixture digested Ald-YLS to completion in 3 h at rt (**Figure 2.7**).

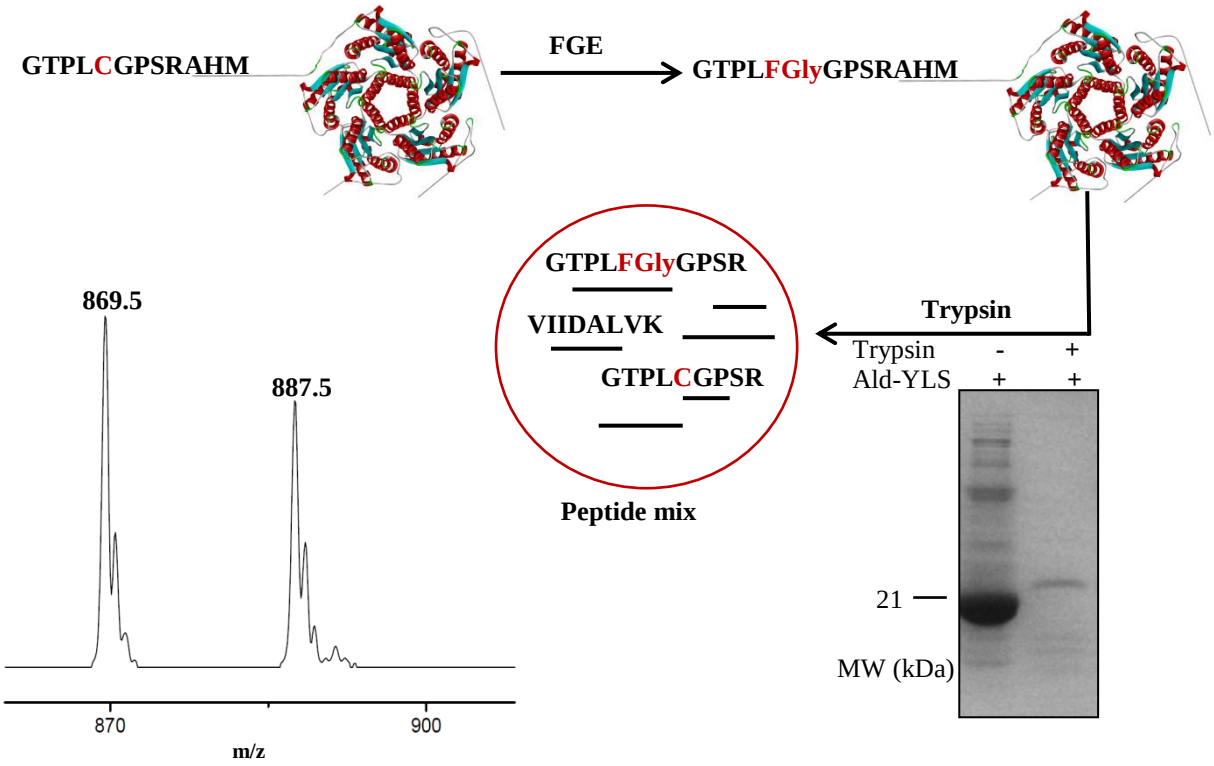


Figure 2.7 Schematic representation of the tryptic digestion of Ald-YLS. Crystal structure of lumazine synthase from *Saccharomyces cerevisiae* (pdb: 1EJB[100]). Colour code: peptide tags which were inserted at the *N*-terminus

of the protein sequence are highlighted in grey. SDS-PAGE analysis indicates that Ald-YLS was fully digested by trypsin.

By LC-MS, we were able to identify the *N*-terminal 9-residue tryptic fragments corresponding to the mass of the aldehyde- and diol-peptide fragments (i.e. peptide-aldehyde calc. *m/z* 869.4, obs. 869.5; peptide-diol calc. *m/z* 887.5, obs. *m/z* 887.5, **Figure 2.7**), as in water the FGly-containing peptide could also be found in a hydrated state, as its geminal diol form.[101] We observed that the expected mass of the diol-containing peptide and unconverted Cys-peptide as detected by LC-MS are identical (i.e. calc. *m/z* 887.4, obs. *m/z* 887.5). Therefore, to further confirm the presence of the diol functionality, the tryptic digested fragments of FGE-4C treated Ald-YLS were measured by high-resolution mass spectroscopy (HRMS). The exact mass of the fragments was confirmed (i.e. peptide-aldehyde calc. *m/z* 435.227, obs. *m/z* 435.227; peptide-diol calc. *m/z* 444.233, obs. *m/z* 444.232 g/mol).

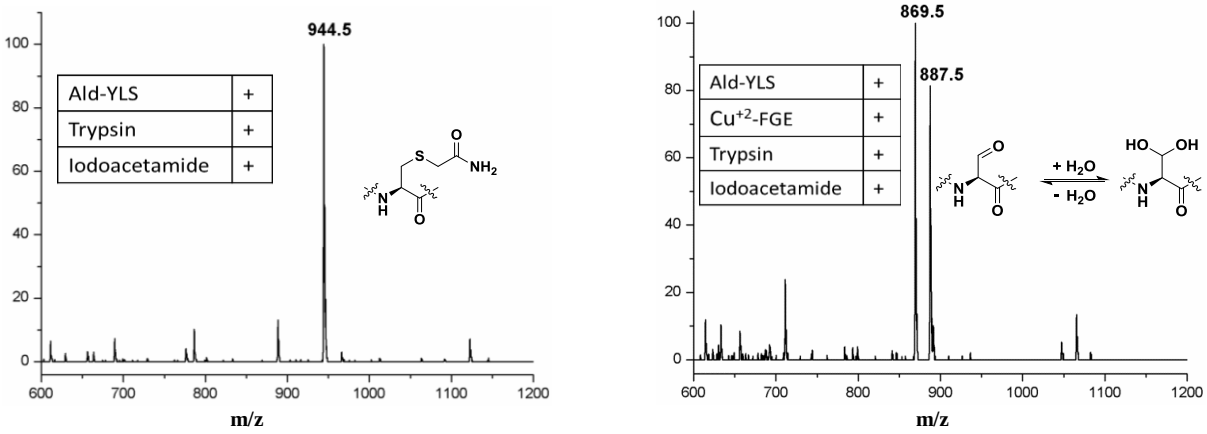


Figure 2.8 Left. Protein was reduced with DTT, alkylated with IAA and digested with trypsin. **Right.** Protein was reduced with DTT, incubated with FGE, alkylated with IAA and digested with trypsin. Reaction mixtures were analysed by LC-MS.

Complete *in vitro* conversion by FGE-4C on Ald-YLS was also demonstrated by alkylation of the unreacted Cys-peptide with iodoacetamide (IAA). As shown in **Figure 2.8**, alkylation of Ald-YLS with IAA followed by tryptic digestion and LC-MS analysis indicate the expected peptide mass (calc. *m/z* 944.5, obs. *m/z* 944.5). After alkylation of Ald-YLS, which was previously treated with FGE-4C, with IAA and tryptic digestion we identified the peptide fragments corresponding to the FGly- and diol-peptide and no alkylated unconverted Cys-peptide.

When monitoring the conversion of Cys to FGly by FGE-4C on Ald-YLS by using HRMS, we detected the expected aldehyde-containing protein, highlighted by * in equilibrium to its diol

form, as shown in **Figure 2.9**. Instead, by using a bioorthogonal reagent to label the formed FGly, and thus driving the thermodynamic equilibrium to the aldehyde form, we could accurately follow the *in vitro* conversion over time by HRMS (**Figure 2.10**).

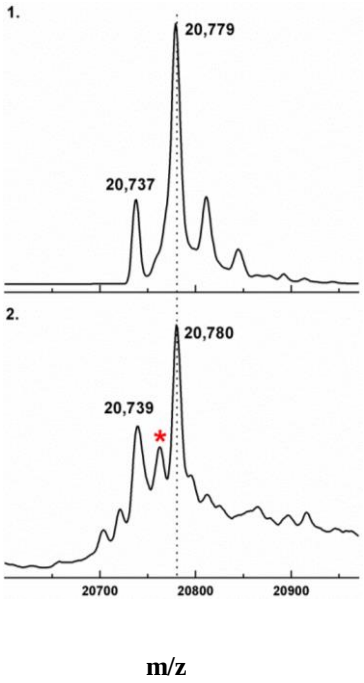


Figure 2.9 Mass spectrometry analysis of Ald-YLS after treatment with FGE-4C. **1.** Protein mass spectrometry of Ald-YLS after 1 min incubation with tFGE-4C. Ald-YLS displays an average mass of 20,737 Da and 20,779 Da (Δ_{mass} = 42 Da). **2.** After 3 h incubation time with FGE-4C a new species of average mass of 20,762 Da, highlighted by * is observed (Δ_{mass} = 18 Da).

The protocol used for monitoring the aldehyde insertion is described. To a solution of Ald-GFP in buffer containing 50 mM Tris, 50 mM NaCl, 50 mM EDTA, was added 0.5 mM DTT at pH 8.0. The samples were incubated at room temperature for 5 min and oxidation was initiated by addition of Cu^{2+} - FGE-4C (7.5 µl, 100 µM). At various time points reaction aliquots (15 µl) were quenched by addition of 25 µl of labelling buffer containing *O*,*O*'-1,3-propanediylbishydroxylamine dihydrochloride or *O*-(4-nitrobenzyl)hydroxylamine hydrochloride (10 mM) in 100 mM sodium acetate buffer, pH 4.7 and 7 M urea. The formation of the corresponding protein-oxime conjugates was completed in 3 h at 35 °C, pH 5.3. 2 µl TFA (10%) was added (pH after TFA addition should be *ca*. 3) to the samples before being measured by high-resolution mass spectrometry.

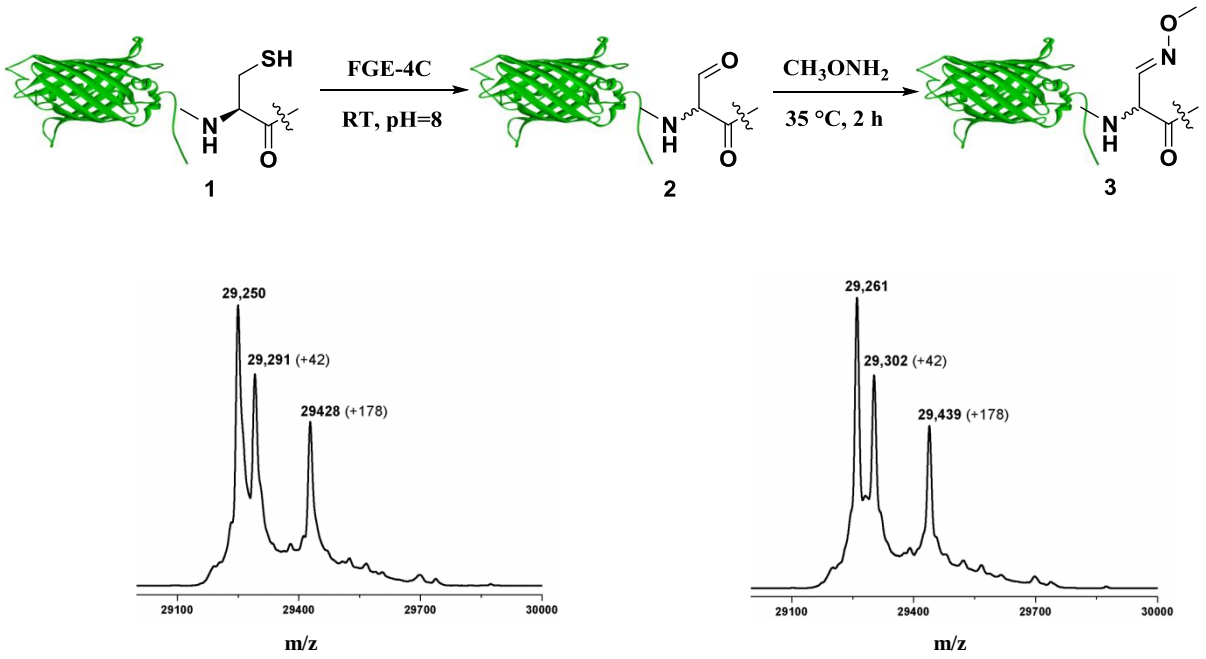
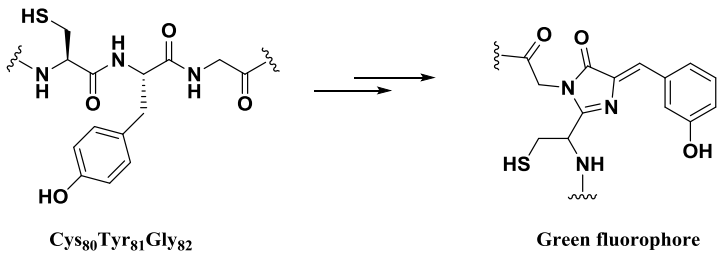


Figure 2.10 FGE-4C activity on Ald-GFP. Reaction scheme of Ald-GFP with FGE-4C, followed by labelling with methoxyamine. **Left.** Mass spectrometry of Ald-GFP after 1 min oxidation with FGE-4C and labelling with methoxyamine. Expected mass: 29,270 Da, found mass: 29,250 Da (Δ_{mass} = 20 Da, due to posttranslational modification on GFP, see **Scheme 2.1**). **Right.** Mass spectrometry of Ald-GFP after 1 h incubation with 0.1 equiv FGE-4C and labelling with methoxyamine. Expected mass: 29,261 Da, found mass: 29,261 Da.



Scheme 2.1 Autocyclisation of the peptidyl triad ($Cys_{80}Tyr_{81}Gly_{82}$) in GFP to the green fluorophore accounts for the observed -20 Da decrease in mass. For an in-depth mechanistic description of the autoconversion of $Cys_{80}Tyr_{81}Gly_{82}$ in the green fluorophore, the reader is invited to read the work of Getzoff and co-workers.[102]

2. 2. 2 *In vitro* activity of FGE-4C on lumazine synthase from *Aquifex aeolicus*

We next sought to place the FGly residue into naturally occurring protein cage nanoparticles.[103] This is an exciting goal as the aldehyde functionality provides us with chemical handles to site-specifically immobilise protein capsids on nanoparticles such as nanocellulose (see **Chapter 4**) and thus develop novel multifunctional bio-conjugates. For this purpose, we chose the lumazine synthase from *Aquifex aeolicus* (**Figure 2.11**), which is a hyperthermophilic protein capsid constructed by the self-assembly of multiple identical subunits (i.e. 60 or 180).[103] Its hollow spherical architecture has been used as a template for encapsulation of HIV protease[104] and as a modular platform for targeted drug delivery[105]. In an effort to expand the utility of this versatile nanocompartment, we attached the FGE recognition sequence at its *N*-terminus. As the monomeric building blocks in lumazine synthase from *S. cerevisiae* and *A. aeolicus* display very similar overall fold (**Figure 2.13** shows the superposition of the two structures), we expected Ald-ALS to be a substrate for FGE. However, incubation of Ald-ALS with FGE followed by labelling with methoxyamine did not display the expected calculated mass by HRMS (**Figure 2.13**, right).

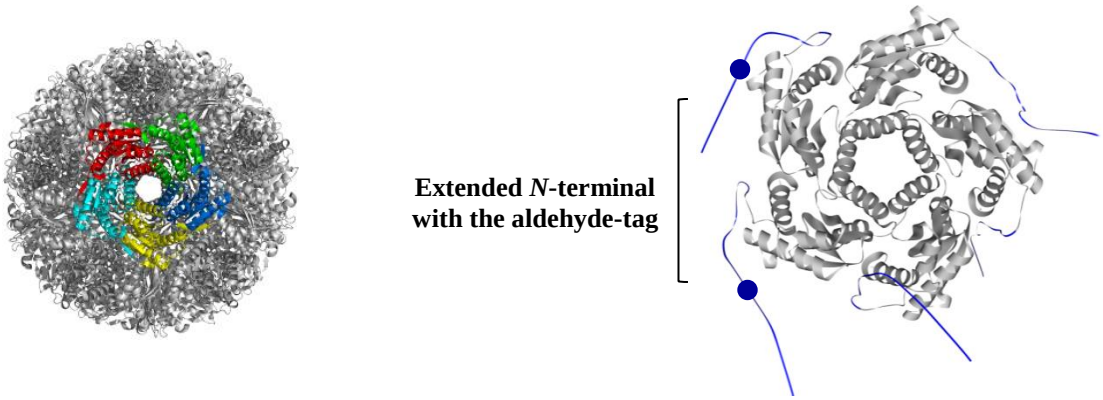


Figure 2.11 Left. Crystal structure of lumazine synthase from *Aquifex aeolicus* (pdb: 1HQK[106]). **Right.** Crystal structure of lumazine synthase from *Saccharomyces cerevisiae* (pdb: 1EJB[100]). Colour code: *N*-terminus of the protein sequence is highlighted in blue and the dot indicates the locus where the aldehyde-tag was engineered.

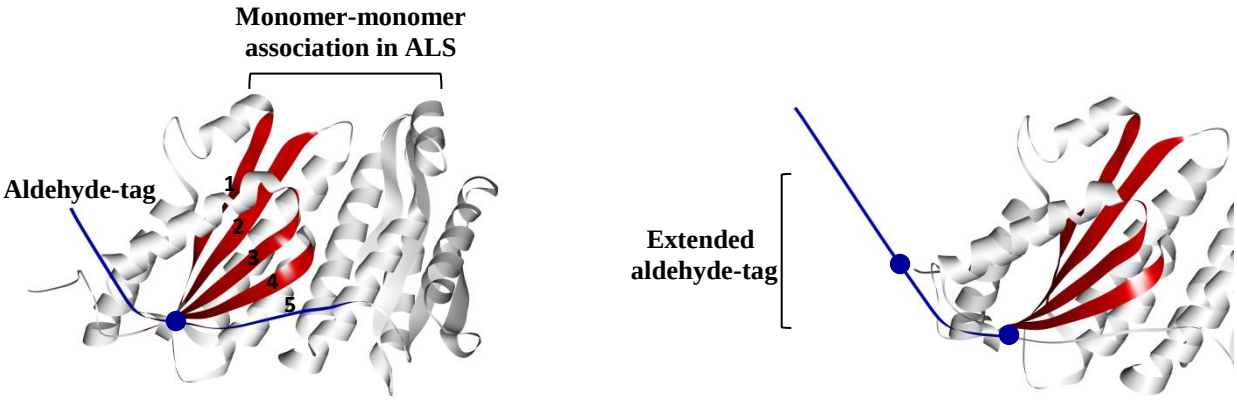


Figure 2.12 Left. Crystal structure of two monomeric subunits of lumazine synthase from *Aquifex aeolicus* (pdb: 1HQK[27]). Colour code: highlighted in red are the *β*-sheets (labelled **1**, **2**, **3**, **4**) of one monomer interacting with the *N*-terminal *β*-strand highlighted in blue (labelled **5**). The blue dot indicates the locus where the aldehyde-tag was engineered. **Right.** Crystal structure of one monomeric subunit of lumazine synthase from *Aquifex aeolicus*; shown in blue is the extended aldehyde-tag. The additional SGSGSG sequence was inserted in between the blue dots.

The difference in the two quaternary structures: pentameric for YLS versus 60-meric icosahedral capsids for ALS (**Figure 2.11**) could account for the restricted accessibility of the Ald-ALS tag in the active site of FGE. In addition, in ALS, the *N*-terminal *β*-strand interacts with the *β*-sheet of the adjacent monomer (**Figure 2.12**, left), thus extending the sheet from four to five strands. In YLS, the *N*-terminal peptide chain folds in a random orientation (**Figure 2.11**, right).

We anticipated that extension of the *N*-terminal tag by insertion of the SGSGSG peptide motif in between the aldehyde tag and the main protein body of ALS (**Figure 2.12**, right) will increase the flexibility of the tag and thus provide a better substrate for FGE. However, the Ald-SGSGSG-ALS variant was not produced in our *E. coli* recombinant system. It could be that the different conformation at the *N*-terminus of the Ald-SGSGSG-ALS protein prevents proper folding and packing of the lumazine synthase subunits to 60-meric icosahedral capsids.

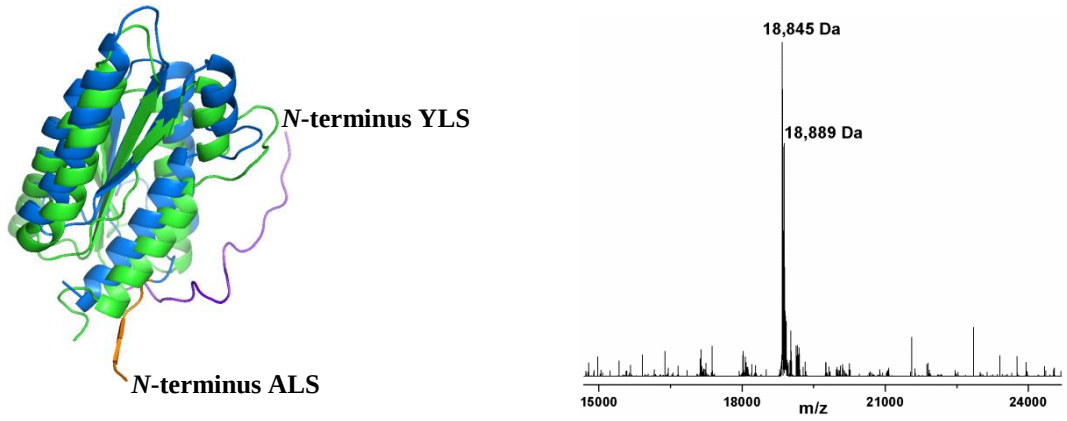


Figure 2.13 Left. Superposition of lumazine synthase monomers (ALS and YLS) from *A. aeolicus* and *S. cerevisiae*. Colour code: YLS is highlighted in blue and its *N*-terminus peptide in orange, while ALS is highlighted in green and its *N*-terminus peptide in purple. **Right.** Mass spectrometry of Ald-ALS after incubation with FGE-4C and labelling with methoxyamine. Expected mass: 18,859 Da, found mass: 18,845 Da, 18,889 Da.

Overall these observations indicate that it is a requirement that the *N*-terminus (to which we attach our aldehyde-tag) is on the surface of the protein and therefore accessible to FGE. However, there are many circumstances where the *N*-termini are buried in the core of the proteins where they play significant roles in oligomeric assemblies, ligand affinity and hyperthermostability.[107] To mention just a few examples: in ALS the *N*-terminus adopts an extended *β*-strand conformation which potentially stabilises the capsid formation and in haemoglobin from *Arabidopsis thaliana*[108] and eukaryotic pentameric ligand-gated ion channels, the *N*-terminal alpha-helical extension influences the quaternary structure.[109]

When the *N*-termini of proteins are not accessible for labelling, one could explore the possibility of inserting the aldehyde tag in internal regions of proteins. But can FGE *curvata* modify aldehyde tags placed in internal loops of proteins?

2. 2. 3 *In vitro* activity of FGE-4C on internal aldehyde-tagged proteins

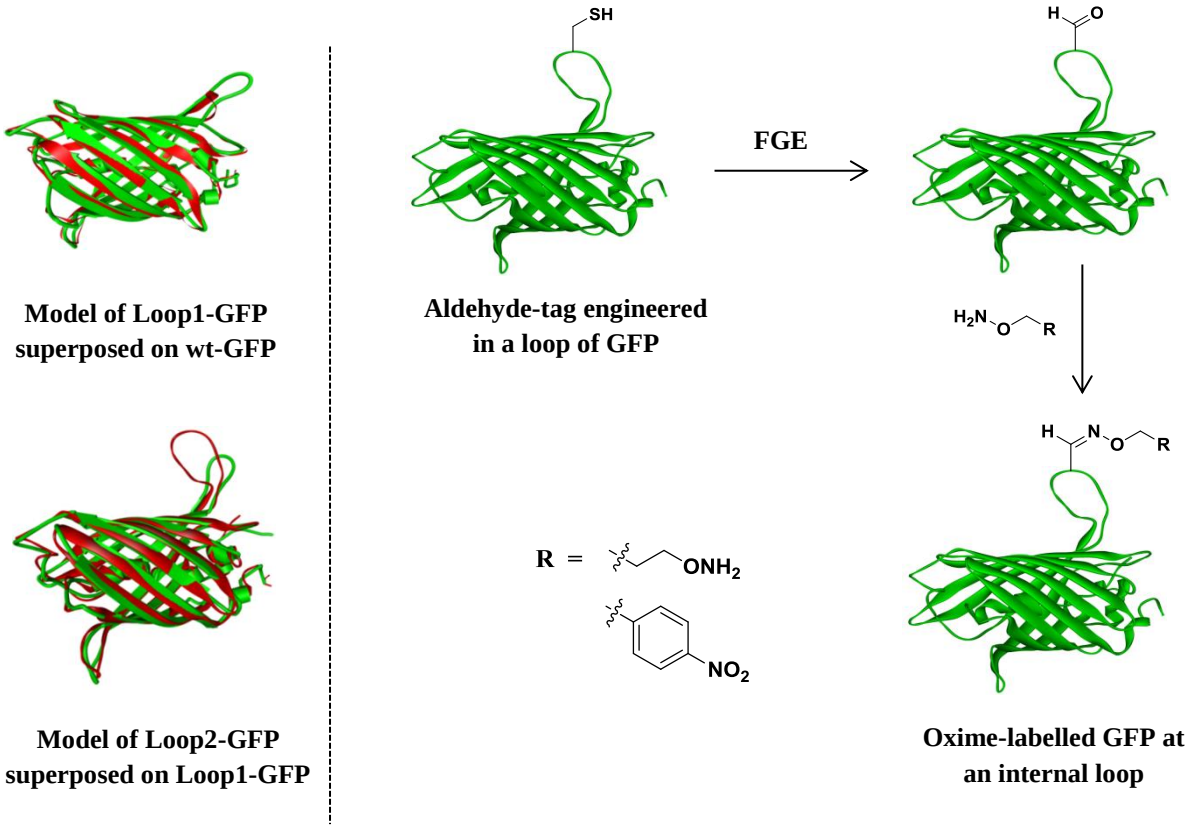


Figure 2.14 Left. Models of GFP structures with extended surface loops. Crystal structures of GFP (from *Aequorea victoria*, pdb: 1GFL[110]) with modelled loop1 peptide sequence. Colour code: GFP is shown in red, homology model Loop1-GFP is shown in green. Homology model of Loop2-GFP with modelled loop2 peptide sequence. Colour code: homology model Loop1-GFP is shown in green, homology model Loop2-GFP is shown in red. **Right.** Schematic representation of incorporation of the aldehyde tag at an internal site of GFP in order to make oxime-labelled proteins.

For addressing this question, GFP can serve as a good model as its crystal structure displays a barrel-like framework with eight loops connecting the *β*–strands of the barrel to one another. Out of these eight surface loops, two sites in GFP could tolerate insertion of 15 amino-acids long peptides of random sequences[111] without compromising fluorescence. We chose one of the smaller loops of GFP (**Figure 2.14**) to insert between Gln157-Lys158 residues the amino acids sequence (i.e. GTPLCGPSR) that was efficiently converted by FGE-4C on linear peptides (**Table 2.1**). We constructed the loop1-gfp clone and observed that the protein is similarly overproduced in *E. coli* as the wild-type, *ca.* 12 mg per one litre culture.

Table 2.1 – Aldehyde tags sequences inserted in an internal loop of GFP

Protein Construct	Sequence	Experimental Outcome
Wild-type GFP	D-K-Q-K-N-G	N/A
Loop1-GFP	D-K-G-T-P-L-C-G-P-S-R-N-G	*ca.* 30%
Loop2-GFP	D-A-G-K-G-T-P-L-C-G-P-S-R-A-G-N-G	>95%

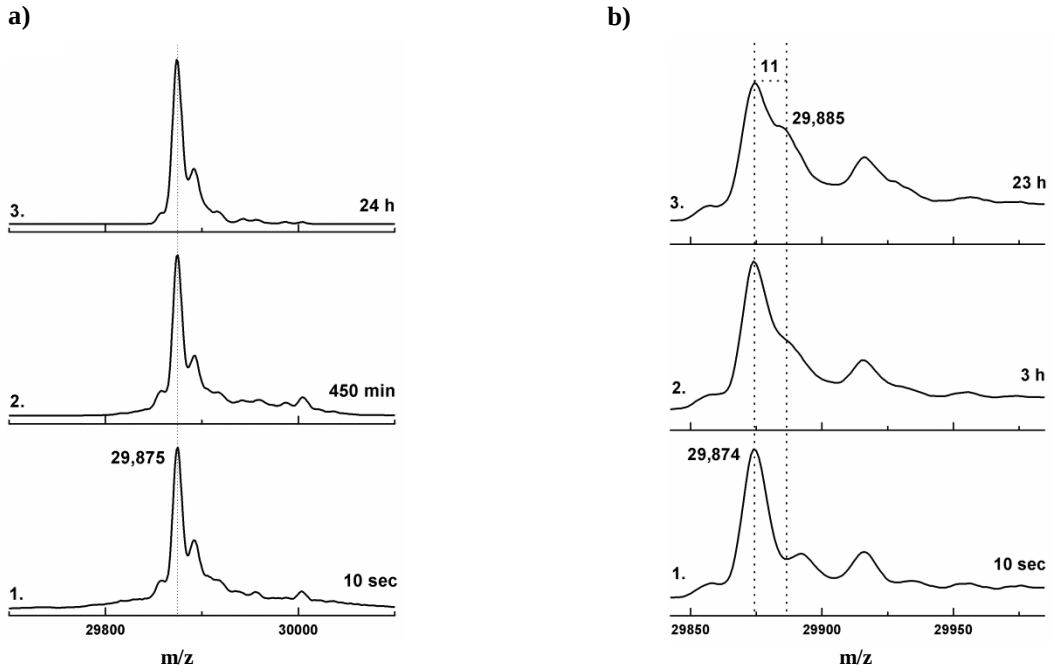


Figure 2.15 Mass spectrometry analyses of Loop1-GFP covalent modification. **a) 1.** Protein mass spectrometry of Loop1-GFP after 10 s incubation with FGE-4C. Loop1-GFP displays an average mass of 29,875 Da. **2.** After long incubation time with FGE-4C (i.e. 450 min) no new species of expected average mass of 29,945 Da is observed. **3.** Longer treatment of Loop1-GFP with FGE-4C (i.e. 24 h) does not change the reaction profile. **b) 1.** Protein mass spectrometry of Loop1 after 10 s incubation with FGE-4C. Loop1 displays an average mass of 29,874 Da (calc. 29,895 Da). **2.** After 3 h incubation time, a new species of average mass of 29,885 Da (calc. 29,886 Da) is observed (Δ_{mass} = 11 Da). **3.** Longer treatment of Loop1 (i.e. 23 h) with FGE-4C does not fully oxidise the protein.

To test the conversion of Cys to FGly, we incubated Loop1-GFP with catalytic amounts of recombinant FGE-4C and at different time points reaction aliquots were quenched by labelling methoxyamine hydrochloride in acetate buffer (pH 5.3). After the labelling reaction, we then analysed the products by HRMS. No conversion could be detected by HRMS (**Figure 2.15**, **a**) when Loop1-GFP was treated with catalytic amounts of FGE-4C. At the same time, we observed

that Loop1-GFP underwent minimal Cys to FGly conversion during treatment with stoichiometric amounts of FGE-4C as indicated by the increase in mass (**Figure 2.15**, **b**).

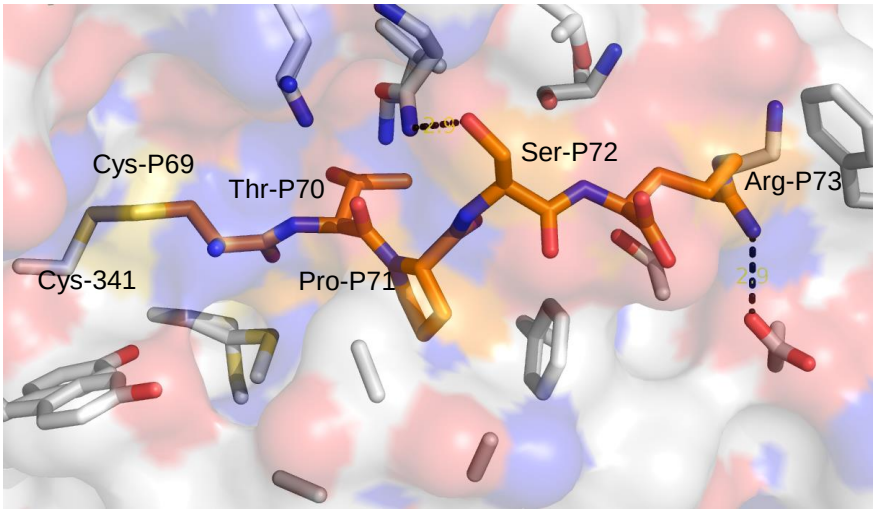


Figure 2.16 Substrate peptide CTPSR bound to mutant FGE C336S *Homo sapiens* (pdb: 2AIJ[112]). The peptide binds to Cys-341 *via* an intermolecular disulfide bond. Colour code: substrate peptide is highlighted in orange.

As we knew from previous experiments that Cu^{1+}-FGE-4C readily oxidizes peptide and *N*-terminal aldehyde-tagged proteins, we assumed that the conformation of our designed loop1 may be too restrained for fitting in the active site of FGE-4C. The crystal structure of FGE C336S *Homo sapiens* in complex with the peptide substrate indicates an extended conformation for the conserved binding CTPSR motif (**Figure 2.16**).[112] Therefore, we designed a more flexible loop which could provide better accessibility of the rigid CGPSR motif to the active site of FGE-4C. Insertion of additional Ala and Gly residues extends the size and flexibility of the loop, as indicated by the model prediction (**Figure 2.14**, left). We thus constructed the loop2-gfp clone and observed that the protein is folded and fluorescent (See Experimental) when produced, *ca.* 3 mg per one litre culture in *E. coli*.

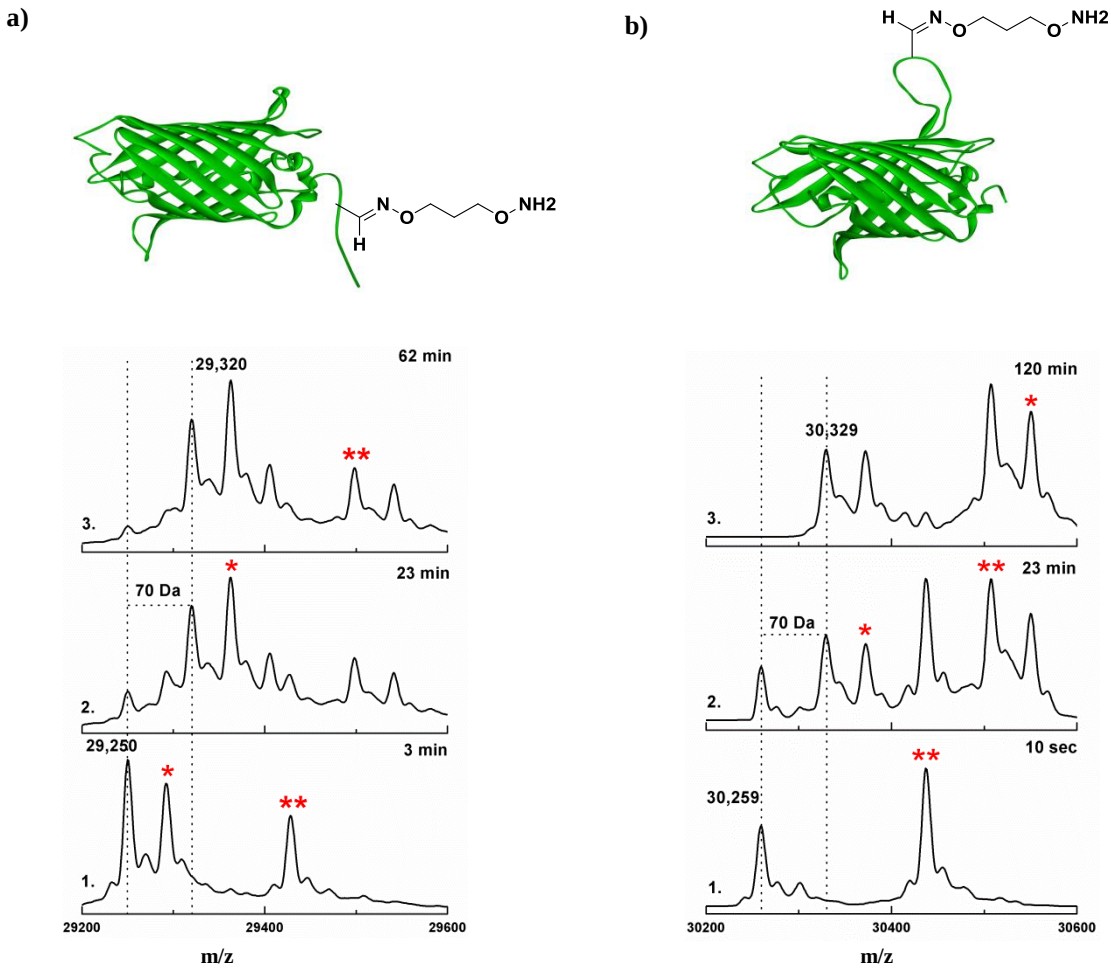
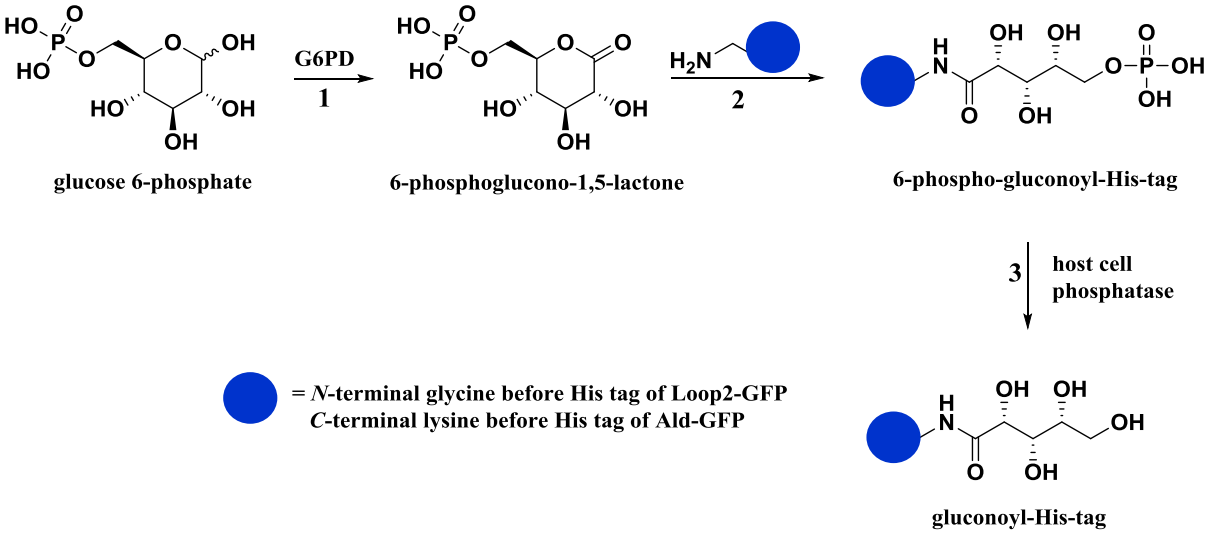


Figure 2.17 Mass spectrometry analyses of Ald-GFP and Loop2-GFP covalent modification. **a) 1.** Protein mass spectrometry of Ald-GFP after 3 min incubation with FGE-4C. Ald-GFP displays an average mass of 29,250 Da. (calc. 29,250 Da[102]) **2.** After incubation with FGE-4C (i.e. 23 min) a new species of average mass of 29,320 Da (calc. 29,320 Da) is observed (Δ_{mass} = 70 Da). **3.** Longer treatment of Ald-GFP with FGE-4C (i.e. 62 min) modifies all starting protein. **b) 1.** Protein mass spectrometry of Loop2-GFP after 10 s incubation with FGE-4C. Loop2-GFP displays an average mass of 30,259 Da (calc. 30,260 Da[102]). **2.** After longer incubation time with FGE-4C (i.e. 23 min) a new species of average mass of 30,329 Da (calc. 30,329 Da) is observed (Δ_{mass} = 70 Da). **3.** Treatment of Loop2-GFP with FGE-4C (i.e. 120 min) modifies all starting protein. Peaks highlighted by * (Δ_{mass} = 42 Da, 43 Da) could probably arise from the carbamylation of a lysine residue or the hydroxylamine functional group.[113] Peaks highlighted by ** (Δ_{mass} = 178 Da) could probably arise from gluconoylation.[114]

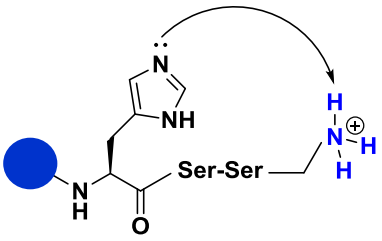
While for Loop1-GFP some conversion was observed in the presence of stoichiometric amounts of FGE, we wanted to see if catalytic conversion could take place on Loop2-GFP. We performed the experiments in parallel with Ald-GFP, and we were able to monitor the conversion over time. We observed full conversion on Loop2-GFP when using 0.1 equiv of tFGE-4C in *ca.* 3 h at room temperature.

2. 2. 3. 1 Suppression of gluconoylation

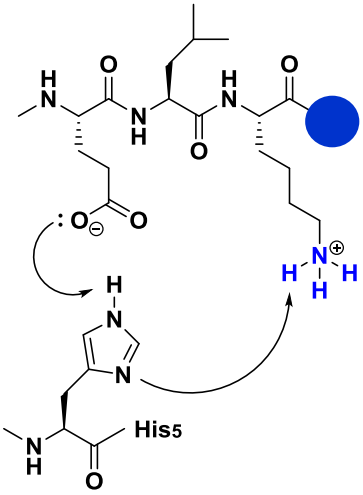
Upon improving the substrate scope of FGE-4C, we next sought to increase the purity of our aldehyde-tagged proteins (i.e. Ald-GFP and Loop-GFP) which consist partly (app. 35%) of posttranslationally modified proteins. This is indicated by the extra mass of 178 Da detected by HRMS (**Figure 2.17**) which is attributed to spontaneous *α-N*-6-phosphogluconoylation of recombinant proteins fused at the *N*-terminus to hexa-histidine affinity tags (His_6-tag).[114] The proposed chemical pathway responsible for the gluconoylation of Ald-GFP and Loop2-GFP is shown in **Scheme 2.2**.



Scheme 2.2[114] Proposed chemical pathway for the gluconoylation of the *N*-terminal glycine in Loop2-GFP and lysine in Ald-GFP. **1.** Enzymatic conversion of glucose 6-phosphate by *E. coli* glucose 6-phosphate dehydrogenase (G6PD) **2.** Primary amine of the *N*-terminal glycine (Loop2-GFP) or *C*-terminal lysine (Ald-GFP) undergoes nucleophilic attack at the lactone group, opening the ring. **3.** Cleavage of the phosphate functional group by *E. coli* phosphatase affords the sugar adduct responsible for the 178 Da increase.



Deprotonation of *N*-terminal glycine by His_6-tag



Deprotonation of lysine by *C*-terminal His_6-tag

Figure 2.18 Left. Deprotonation of the *N*-terminal primary amine group of glycine by adjacent histidine in Loop2-GFP. **Right.** Deprotonation of the side-chain of the lysine residue adjacent to histidine.

According to the proposal, Loop2-GFP undergoes *α-N*-6-gluconoylation at the *N*-terminal glycine residue which is preceded by two serine residues and the His_6-tag sequence (i.e. GSSHHHHHHS -). As shown in **Figure 2.18** (left), it could be that the role of histidine is to deprotonate the primary amine of the glycine residue, thus making it nucleophilic towards addition to 6-phosphoglucono-1,5-lactone. For Ald-GFP, we inserted the His_6-tag at the *C*-terminus of the protein. As we still observed gluconoylation, we proposed that in this case the lysine residue adjacent to the His_6-tag (- KLEHHHHHH) undergoes the reaction with 6-phosphoglucono-1,5-lactone (**Figure 2.18**, right).

As all our protein variants produced in BL21(DE3)pLysS *E. coli* strain exhibit significant product gluconoylation, we sought to suppress the posttranslational gluconoylation by investigating the production in other *E. coli* strains such as Top10(DE3) and JM83(DE3). We observed that when using the Top10(DE3) strain, the proteins are overproduced in high purity as indicated by SDS-PAGE analysis (See Experimental) and without additional sugar adducts as indicated by HRMS (**Figure 2.19**).

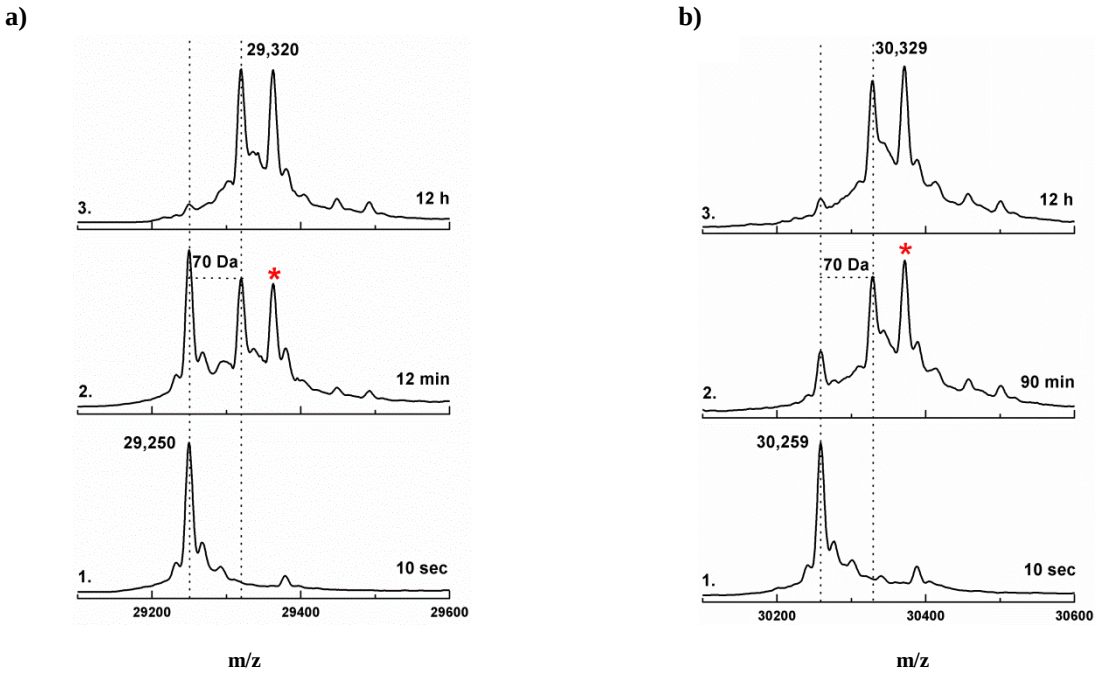


Figure 2.19 Mass spectrometry analyses of Ald-GFP and Loop2-GFP covalent modification. **a) 1.** Protein mass spectrometry of Ald-GFP after 10 s incubation with FGE-4C. Ald-GFP displays an average mass of 29,250 Da (calc. 29,250 Da). **2.** After incubation time with FGE-4C (i.e. 12 min) a new species of average mass of 29,320 Da (calc. 29,320 Da) is observed (Δ_{mass} = 70 Da). **3.** Longer treatment of Ald-GFP with FGE-4C (i.e. 12 h) modifies all starting protein. **b) 1.** Protein mass spectrometry of Loop2-GFP after 10 s incubation with FGE-4C. Loop2-GFP displays an average mass of 30,259 Da (calc. 30,260 Da). **2.** After incubation time (i.e. 90 min) with FGE-4C a new species of average mass of 30,329 Da (calc. 30,329 Da) is observed (Δ_{mass} = 70 Da). **3.** Treatment of Loop2-GFP with FGE-4C (i.e. 12 h) modifies the starting protein. Peaks highlighted by * (Δ_{mass} = 43 Da) could probably arise due to carbamylation of the hydroxylamine functional group.

2. 2. 3. 2 Protein carbamylation

The decrease in pH to 5.3 for the labeling reaction could affect the protein folding. As a result, for the labeling reaction we used *ca.* 4M urea to prevent protein precipitation and therefore get an accurate monitoring of the FGE conversion, which was performed at pH 8.0. It was previously reported that urea solution can cause carbamylation at the *N*-termini of proteins.[113] This could happen when urea dissociates to ammonia and cyanate; thus, nucleophilic attack by hydroxyl amine to the isocyanic acid affords carbamylation (**Figure 2.20**). In order to avoid carbamylation, for our monitoring experiments we used a different labelling reagent (i.e. *O*-(4-nitrobenzyl)hydroxylamine) which lacks the additional nucleophilic hydroxylamine group (**Figure 2.21**).

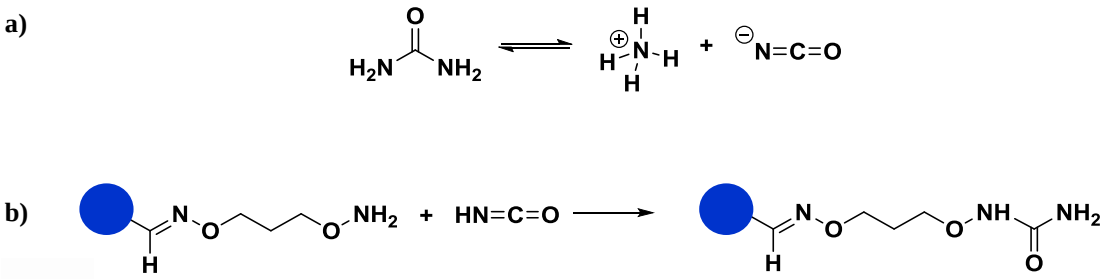


Figure 2.20 Possible mechanism of carbamylation of GFP variants labelled with *O*,*O*′-1,3 propanediylbishydroxylamine linker. **a)** The decomposition of urea into ammonia and cyanate. **b)** Carbamylation at the labelled protein residue by reaction with isocyanic acid.

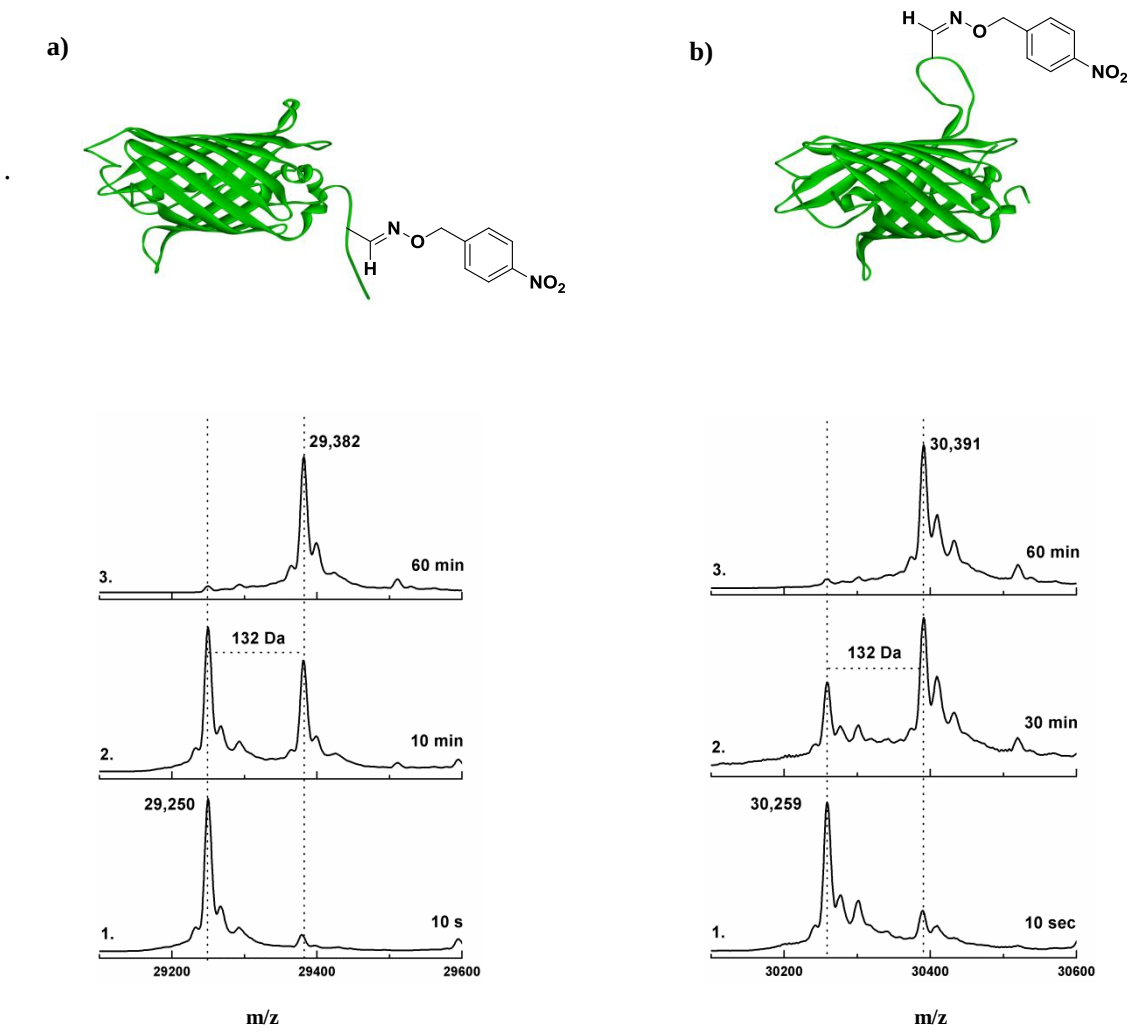


Figure 2.21 Mass spectrometry analyses of Ald-GFP and Loop2-GFP covalent modification. **a) 1.** Protein mass spectrometry of Ald-GFP after 10 s incubation with FGE-4C. Ald-GFP displays an average mass of 29,250 Da (calc. 29,250 Da). **2.** After incubation time with FGE-4C (i.e. 10 min) a new species of average mass of 29,382 Da (calc. 29,382 Da) is observed (Δ_{mass} = 132 Da). **3.** Longer treatment of Ald-GFP with FGE-4C (i.e. 60 min) modifies all starting protein. **b) 1.** Protein mass spectrometry of Loop2-GFP after 10 s incubation with FGE-4C. Loop2-GFP displays an average mass of 30,259 Da (calc. 30,260 Da). **2.** After incubation time (i.e. 30 min) with FGE-4C a new species of average mass of 30,391 Da (calc. 30,391 Da) is observed (Δ_{mass} = 132 Da). **3.** Longer treatment of Loop2-GFP with tFGE-4C (i.e. 60 min) modifies the starting protein.

2. 3 Discussion

We described a chemoenzymatic method for rapid *in vitro* protein labelling. The method firstly requires the targeted protein to be engineered with the aldehyde-tag at the *locus* where labelling is required. For this purpose, we constructed a general pET28 expression plasmid (ald-gfp) that inserts the recognition sequence for FGE at the *N*-terminus and a His-tag at the *C*-terminus. Using this vector we recombinantly produced a variety of proteins such as Ald-GFP, Ald-YLS, Ald-ALS, Ald-SLAC, Ald-Protein A and Ald-FGE-4C (see **Chapter 5**). Incubation of FGE with aldehyde-tagged proteins followed by the labelling reagent resulted in site-specific covalent modification.

After selective oxidation of aldehyde-tagged proteins by FGE, the afforded aldehyde functionality is detected in equilibrium with its geminal diol form. That is to say, under aqueous conditions the unhindered FGly residue undergoes $sp^2 \rightarrow sp^3$ hybridization by hydration at C_α (**Figure 2.22**, **a**). High-resolution crystal structure (1.3-Å resolution) of the bacterial sulfatase *Pseudomonas aeruginosa*, a native substrate of FGE, validates the evidence of the geminal hydrate form of FGly (**Figure 2.22**, **b**).[115] In addition, the electron density map of the geminal diol indicates that the active site environment of the sulfatase stabilises the (*S*) - configuration of the hydra-FGly residue (**Figure 2.22**, **b**).

On the other hand, when monitoring the conversion on peptide substrates, diastereomers of the oxime peptide were detected, suggesting epimerization at C_α as a result of enolization. It could be that by lowering the pH (i.e. from 8.0 to 5.3) during the oxime ligation, enolization of FGly is catalysed under acidic conditions. Faster enolization accounts for racemization at C_α and thus formation of the peptide diastereoisomers. The loss of chirality at the C_α was also observed on a different FGE peptide substrate which is deuterated at the C_α centre (**1**).[99] Upon reaction with FGE, the obtained FGly is converted to a cyclic imine (**2**) which is in equilibrium with its enamine tautomer (**3**). As a result, the major product of this reaction displays the mass of the racemic imine peptide (**4**, calc. *m/z* 988.5, obs. *m/z* 988.0) which through tautomerization exchanges the deuterium at the C_α with hydrogen.[99] Significantly, this epimerization reaction at C_α of FGly, which occurs in the oxime labelled aldehyde-tagged proteins has perhaps been underestimated as a source of protein isoforms.

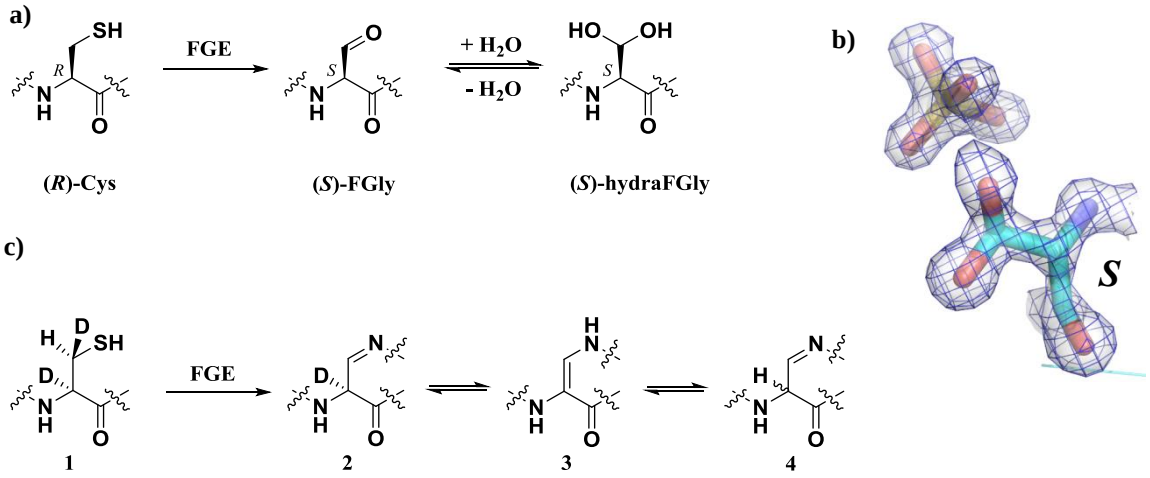


Figure 2.22 Stereochemistry at the FGly carbon centre. **a)** Reaction scheme of (*R*) - Cys with FGE, affording (*S*) - FGly in thermodynamic equilibrium with (*S*) - hydraFGly. **b)** A view of the FGly residue in the active site of arylsulfatase from *P. aeruginosa* (pdb: 1HDH[115]). The electron density points to the FGly hydrated form. **c)** Reaction scheme of the deuterated Cys with FGE, leading to imine (**2**) formation in equilibrium with its enamine (**3**) tautomer.

Most importantly, *in vitro* FGE catalytically oxidises the GFP irrespective of the aldehyde-tag being placed at the *N*-terminus or in a loop of the targeted protein. This allows the engineering of a multifunctional aldehyde-tagged GFP which finds applications in the immobilization of cross-linked proteins or protein-DNA conjugates on surfaces. We also found proteins such as Ald-ALS and Loop1-GFP that could not be efficiently labelled. While for Ald-ALS the accessibility to the aldehyde-tag is restricted by the proteins quaternary structure, for Loop1-GFP the conformation of the designed loop was determined to be too rigid.

While we were investigating the ability of FGE-4C *T. curvata* to introduce aldehyde functionalities on internal regions of GFP, the very same concept was published by Bertozzi and co-workers at the internal glycosylation site of Fc fragment of IgG1.[116] However, based on their highly-regarded "relative mass spectral abundances, they calculated the Cys-to-FGly conversion to be 76%" under stoichiometric amounts of FGE *M. tb.* at pH 9 for 20 h at 42 °C.[116] Our results based on mass spectra indicate >95% conversion using 0.1 eq of FGE for 3 h, at rt. The reason behind these results is the pre-activation treatment of FGE-4C variant with Cu^{1+}, which coupled with a number of point mutations, increases the catalytic activity of this enzyme by nearly 200-fold.[99] Recently, Rabuka *et al.* took advantage of their Cu^{1+} pre-activation of FGE *S. coelicolor* and *H. sapiens* to demonstrate fast *in vitro* catalytic conversion on aldehyde-tags inserted at internal sites of monoclonal antibodies.[117] Moreover, internally placed FGly was used as a

chemical handle to construct homogeneous anti-Her2 antibody trastuzumab conjugates[118] and selectively assemble aldehyde-tagged proteins with DNA scaffolds.[97]

2. 4 Experimental

Construction of expression plasmids for engineering *N*-terminal and internal aldehyde tags. pET28-Ald-gfp. The gene encoding the ald-gfp was amplified from vector pET-gfp by PCR which at the same time introduced the aldehyde tag by using the following primers: ald-gfp_s, gfp_s and ald-gfp_a. The NcoI and XhoI digested PCR fragment was ligated into vector pET28 which had been digested with the same restriction enzymes. The resulting plasmid pET28-Ald-gfp was then used as a template for the construction of all the aldehyde-tagged proteins. **pET28-Loop1-gfp.** The gene encoding loop1-gfp was obtained by introducing mutations of the loop designs into the gfp gene by site-specific mutagenesis by using the following primers: $loop1_s$, ald-gfp_a, $loop1_a$ and gfp_s. The NcoI and XhoI digested PCR fragment was then ligated into vector pet28-Ald-gfp, which had been digested with the same restriction enzymes to remove the ald-gfp gene.

Table 2.2 Oligonucleotides used in these experiments

Primer	Sequence (5'-> 3')
Ald-gfp sense	TATACATATGGGCAGCGGAGCATCCAAAGGAGAAGAA
Gfp sense	ATATCCATGGGAACACCACTATGCGGACCATCACGAGCACATATGGGCAGCGGAG
Ald-gfp antisense	TATACTCGAGTTTGTACAGTTCATCCATGC
Ald-yls sense	GATATACATATGGCAGTTAAAGGATTAGGCAA
Ald-yls antisense	GATATACTCGAGAAAAGCATTTTTACCGAACT
Ald-als sense	ACGATATCATATGGAAATCTACGAAGGTAAACTA
Ald-als antisense	GATATACTCGAGTCGGAGAGACTTGAATAAGT
Loop1-gfp sense	ACACCACTATGCGGACCATCACGAAATGGAATCAAAGTGAACTTC
Loop1-gfp antisense	TGATGGTCCGCATAGTGGTGTTCCTTTGTCTGCCATGATGTAT
Loop2-gfp sense	CCACTATGCGGACCATCACGAGCGGGCAATGGAATCAAAGTGAACTTC
Loop2-gfp antisense	TGGTCCGCATAGTGGTGTTCCTTTGCCCGCTTTGTCTGCCATGATGTAT
Ald-SG_3-gfp sense	TATACATATGGGCAGCGGAAGTGGTAGTGGTAGTGGTGCATCCAAAGGAGAAGAA

Expression and purification of Ald-GFP, Loop1-GFP and Loop2-GFP. Aldehyde-tagged variants were overproduced using the T7 promoter system in *E. coli* BL21 cells that had been transformed with the appropriate plasmid (pET28-ald-gfp, pET-loop1-gfp, pET-loop2-gfp). Cultures were grown at 37 °C in LB medium supplemented with kanamycin (50 mg/ml) and chloramphenicol (34 mg/ml) to an OD_{600} of 0.7 and then production of Ald-GFP was induced

with IPTG (0.1 mM) and cultures were grown at 28 °C. 16 h after induction, the cells were harvested by centrifugation at 8000 rpm and 4 °C for 20 min. The cell pellets were resuspended in 10 ml lysis buffer (50 mM sodium phosphate, 300 mM NaCl, pH 8.0) and lysed by sonication.

Cell lysates were cleared by centrifugation (8000 rpm for 20 min at 4 °C) and loaded onto 1 ml Ni^{+2}-NTA agarose resin in a gravity flow column. After washing with lysis buffer and lysis buffer containing 10 mM and 20 mM imidazole, proteins were eluted with lysis buffer containing 250 mM imidazole. Protein concentration was determined by nanodrop, and the yield of Ald-GFP is approximately 12 mg/ml. Proteins were dialysed into buffer containing 50 mM TrisHCl and 50 mM NaCl, pH 8.0. The proteins were flash-frozen and stored at −80 °C until needed. Similarly, Loop2-GFP was produced in *E. coli* BL21. Protein purity was assessed by SDS-PAGE and Loop2-GFP was further purified by size exclusion chromatography on a Superdex 75 HiLoad FPLC column using 50 mM TrisHCl, pH 8.0, containing 200 mM NaCl as a running buffer.

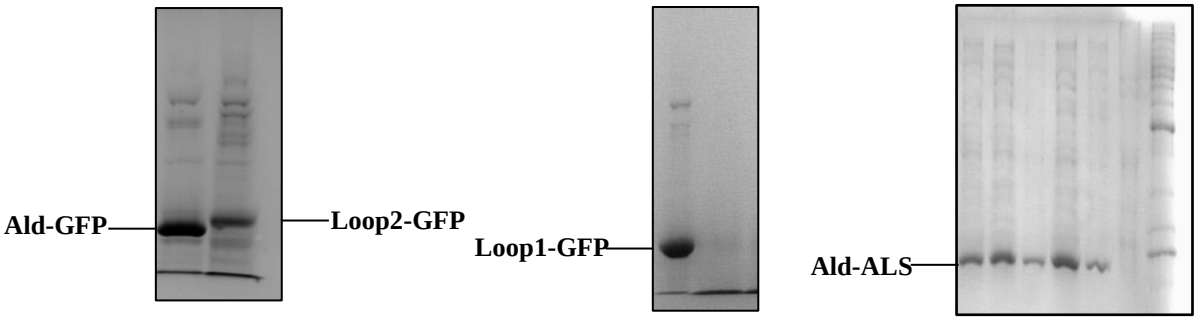


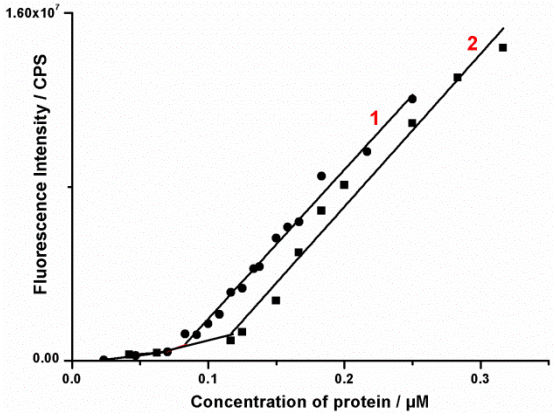
Figure 2.23 10-12% Tris SDS-PAGE.

Posttranslational modification of Ald-GFP, Loop1-GFP and Loop2-GFP. FGE-4C (134 µl, 150 µM) and $CuSO_4$ (20 µl, 1 mM $CuSO_4$ in 50 mM EDTA) were added to 46 µl 50 mM Tris, 50 mM NaCl, pH 8.0. Ald-GFP, Loop1-GFP or Loop2-GFP (56 µl, 135 µM) were added to 70.5 µl 50 mM Tris, 50 mM NaCl, pH 8.0 containing 0.5 mM DTT (1 µl, 250 mM) and 50 mM EDTA (15 µl, 500 mM). The samples were incubated at room temperature for 5 min and oxidation was initiated by addition of Cu activated tFGE-4C (7.5 µl, 100 µM). At various time points reaction aliquots (15 µl) were quenched by addition of 25 µl of labelling buffer containing *O*,*O*'-1,3-propanediylbishydroxylamine dihydrochloride or *O*-(4-nitrobenzyl)hydroxylamine hydrochloride (10 mM) in 100 mM sodium acetate buffer, pH 4.7 and 7M urea. The formation of the

corresponding protein-oxime conjugates was completed in 3 h at 35 °C, pH 5.3. 2 µl TFA (10%) was added (pH after TFA addition should be *ca.* 3) to the samples before being measured by high-resolution mass spectrometry.

Tryptic digestion of Ald-YLS. $CuSO_4 \cdot 5H_2O$ (1 mM in 500 mM EDTA) was incubated with FGE-4C (1 eq:1 eq) for 10 min on ice. After 1 min centrifugation, the supernatant was incubated with ald-YLS (1 eq:10 eq) and DTT (0.5 mM) and EDTA (50 mM) were added to the reaction mixture. Following 3 h incubation at RT, tryptic digestion was performed and the fragments were analysed by LC-MS

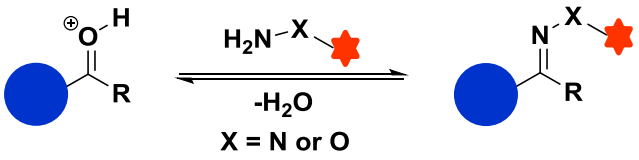
Fluorescence calibration curves of Ald-GFP and Loop2-GFP in the presence of cellulose nanocrystals.



Calibration curve **1**: the exemplary calibration of recombinant Ald-GFP in the presence of CNC (55.7 µg/ml). Calibration curve **2**: the exemplary calibration of recombinant Loop2-GFP in the presence of CNC (55.7 µg/ml).

3. 1. 1 Conjugation of biomacromolecules *via* hydrazone ligation

Chemoselective ligation reactions "were first described as the coupling of two mutually and uniquely reactive functional groups in an aqueous environment".[119] Among the first reactions to be explored by protein chemists as chemoselective was the condensation of ketones and aldehydes with α-effect nucleophiles (when considering the enhanced reactivity at the nucleophilic centre adjacent to atoms bearing unshared pairs of electrons).[120,121] Under acidic conditions (pH 4-6) aldehydes and ketones react with hydrazines to form hydrazones and with aminooxy groups to give the corresponding oximes (**Scheme 3.1**).[120]



Scheme 3.1 An amino acid bearing a carbonyl group is conjugated to α-effect nucleophiles *via* a chemoselective ligation reaction; when X is a nitrogen atom hydrazones are obtained, and when X is an oxygen atom oximes are formed.

The conjugation of hydrazines to aldehydes was utilised for the generation of a variety of bioconjugates.[122] For example, Meier and co-workers demonstrated the covalent attachment of proteins, namely yellow fluorescent protein (YFP), human epidermal growth factor receptor 2 (HER2) and human antibody IgG to artificial vesicles known as polymersomes *via* a stable bis-aryl hydrazone bond (**Figure 3.1**, a).[123] This method required the synthesis of formylbenzamide polymersomes and prior functionalisation of the abovementioned proteins with hydrazinonicotinamide.[123] The aryl hydrazone ligation was also used for the chemoselective covalent modification of quantum dots (QDs, CdSe-ZnS core-shell nanocrystals) with fluorescent peptides (**Figure 3.1**, b).[124] In order to increase the rate of the hydrazone formation, catalytic amounts of aniline were added to the reaction mixture of the peptide with the aldehyde-containing nanoparticles. Such conjugation chemistry fulfils the criteria of efficiency, biocompatibility and stability of the resulting bond.[123,124] Moreover, beyond being selective, the formed aryl hydrazone bond could be easily monitored by ultraviolet-visible (UV-vis) spectroscopy.[123,124]

Recently, Rabuka and co-workers elaborated a new bio-conjugation reaction known as the hydrazine-Pictet-Spengler (HIPS) ligation (**Figure 3.1**, c).[125] Maltose binding protein (MBP) and a therapeutic monoclonal antibody Herceptin (α-HER-2) that contain an FGly residue were tagged with hydrazine indoles.[125] The reaction pathway goes through a hydrazone intermediate resulted from the selective interaction of the aldehyde group with methyl substituted hydrazine. After condensation, the obtained hydrazone undergoes ring closure by formation of a C-C bond. Consequently, the HIPS ligation product is very stable (i.e. 5 days) in comparison to oxime-linked conjugates.[125]

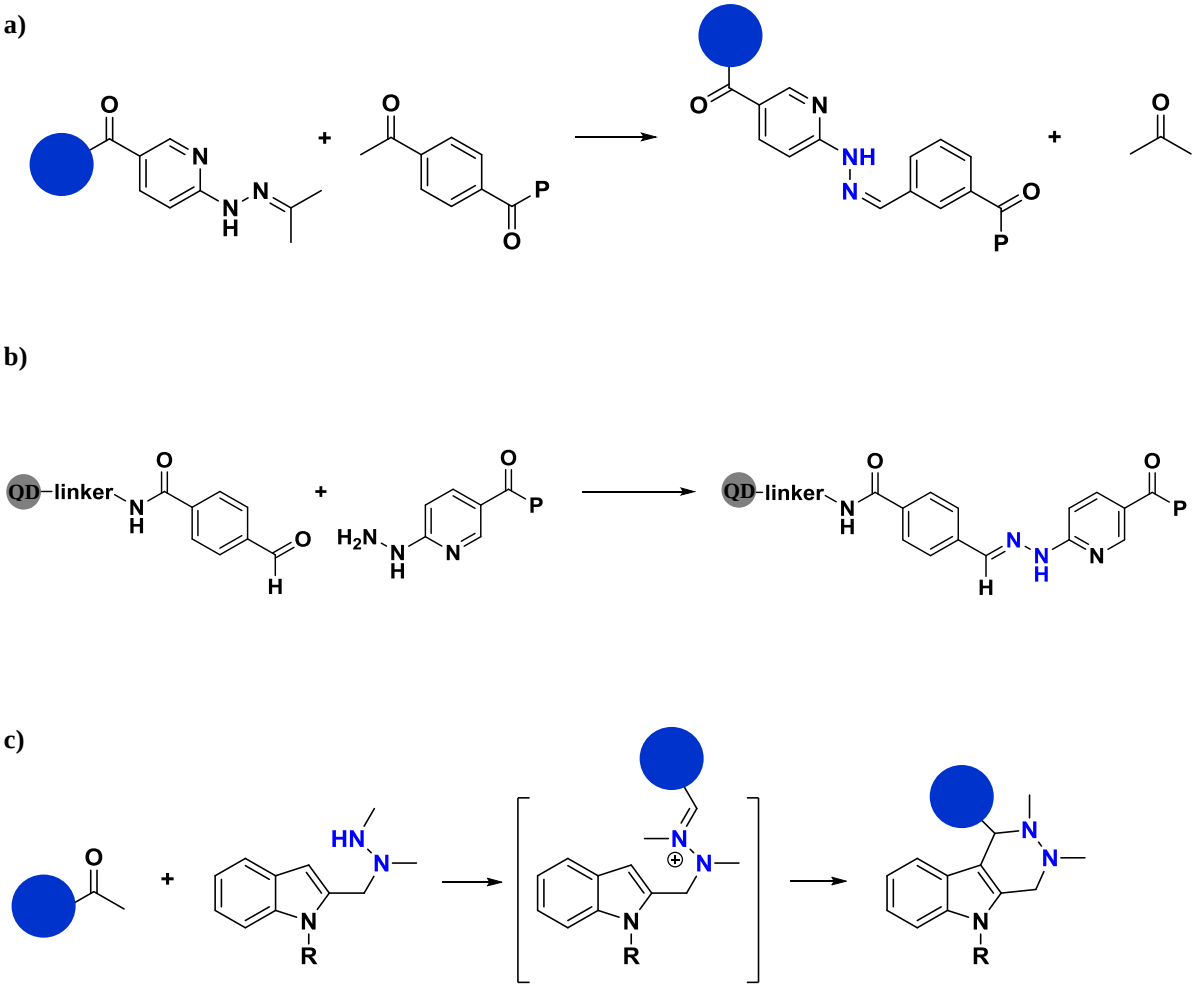
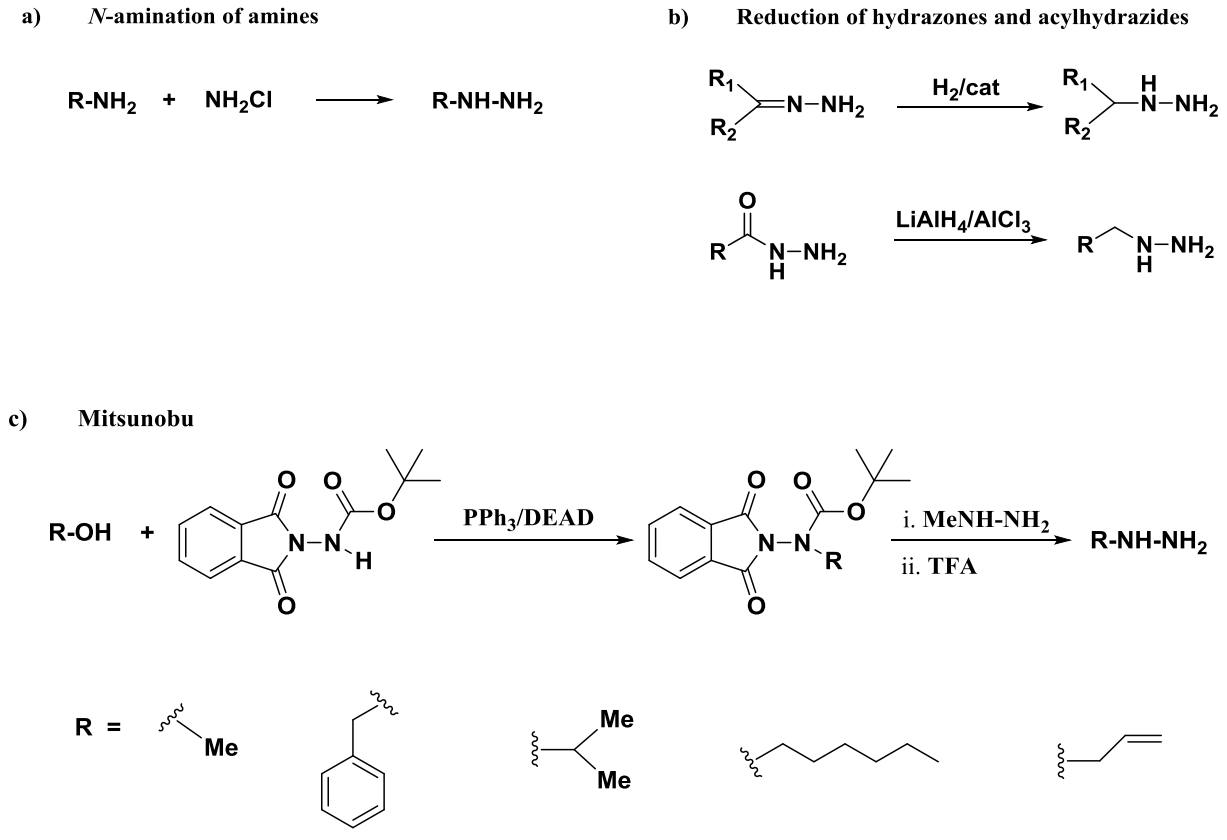


Figure 3.1 Reactions schemes showing the conjugation of biomacromolecules *via* hydrazone ligation. **a)** Conjugation of 4-formylbenzoate-modified polymerasomes with 6-hydrazinonicotinate hydrazone functionalised protein.[123] **b)** Covalent modification of quantum dots with peptides by arylhydrazone ligation.[124] **c)** Aldehyde-containing protein is conjugated by the Hydrazino-Pictet-Spengler (HIPS) ligation.[125]

3. 1. 2 Synthesis of hydrazines

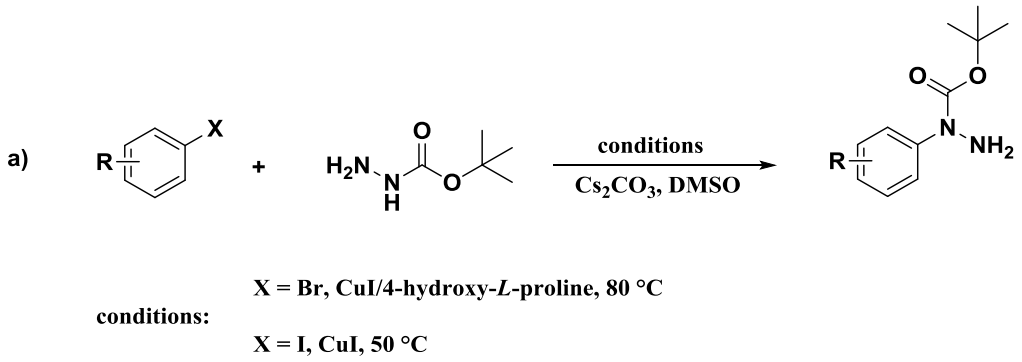
Owing to the importance of hydrazine derivatives, synthetic methods are continuously being developed. Among the first conventional methods to be explored for the preparation of monosubstituted alkylhydrazines was the direct amination by reaction between chloramine and primary amines (**Scheme 3.2**, **a**).[126] Reduction of hydrazones by catalytic hydrogenation or mild reducing agents such as $NaBH_4$ or $NaBH_3CN$ and reduction of hydrazides with $LiAlH_4$ directly provide the corresponding alkyl derivatives (**Scheme 3.2**, **b**).[126] In addition, Brosse and co-workers identified the potential use of *N*-substituted aminophthalimides as acid partners in Mitsunobu reaction for the conversion of alcohols into monosubstituted hydrazines (**Scheme 3.2**, **c**).[127] The method also requires a final dephthaloylation step to cleave the protected alkyl hydrazine.



Scheme 3.2 Conventional methods to synthesise monosubstituted alkyl hydrazine.

Effective systems which use the copper-catalysed C-N bond formation between aryl halides and *N*-Boc hydrazine were developed for the synthesis of *N*-aryl hydrazides (*N*-Boc aryl hydrazines)

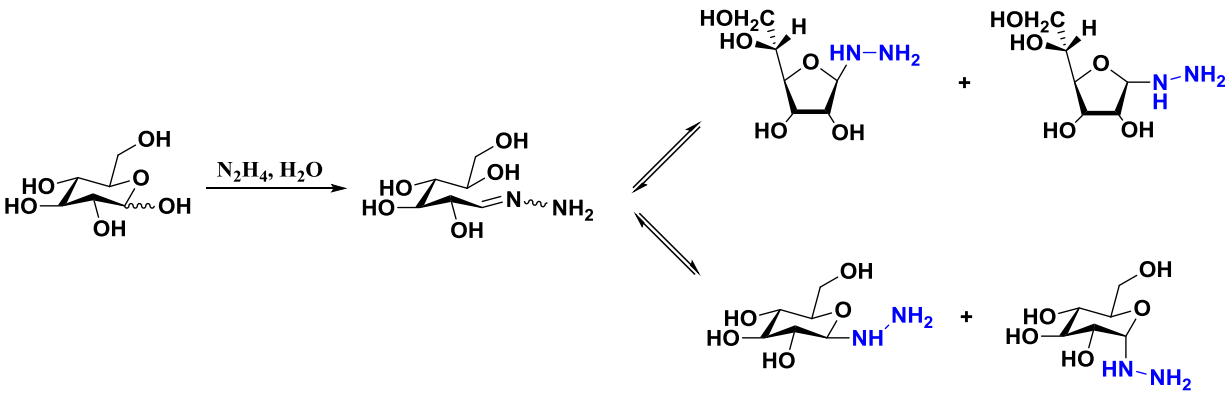
(**Scheme 3.3**, **a**).[128] The use of ligands such as 4-hydroxy-*L*-proline, picolinic acid and *N*, *N*-dimethyl glycine efficiently converted a multitude of aryl bromides and iodides to *N*-aryl hydrazides under mild conditions (50 °C to 80 °C).[128]



Scheme 3.3 Current methods to synthesise aryl hydrazines.

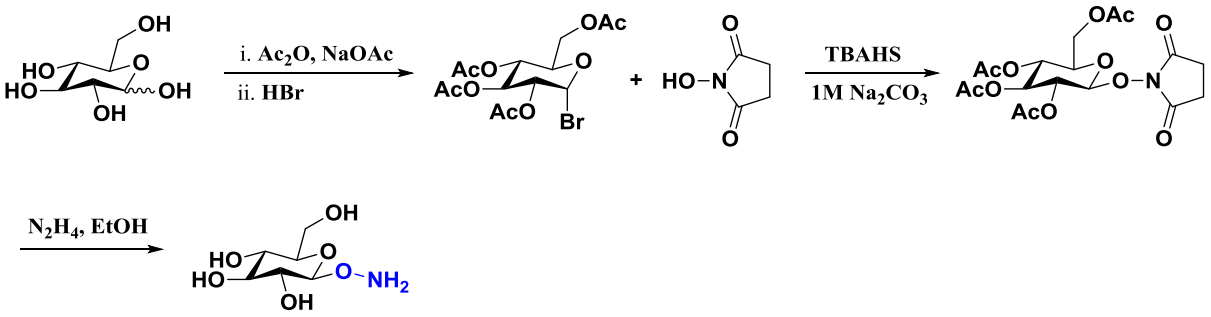
3. 1. 3 Synthesis of hydrazine/aminooxy glycans

Hydrazine functionalities were introduced in sugar molecules by conjugation of hydrazine hydrate at the reducing-ends, followed by ring closure. Reaction of glucose with hydrazine hydrate resulted in a mixture of glucose and fructose hydrazine functionalised saccharides. Notably, this strategy inserts hydrazine functions at the anomeric centre (**Scheme 3.4**).[129]



Scheme 3.4 Reaction of glucose with hydrazine hydrate results in the formation of a mixture of hydrazine functionalised saccharides.

The insertion of an aminooxy group at the reducing end of a mono- or disaccharide was accomplished in 4-synthetic steps.[130] The reaction preceded by protection of the hydroxyl groups, followed by selective halogenation at the anomeric centre (**Scheme 3.5**). Reaction of the brominated glycoside with *N*-hydroxysuccinimide resulted in the formation of *N*-hydroxysuccinimidoglycoside. Cleavage of the succinimide group with hydrazine hydrate yielded the desired saccharide.



Scheme 3.5 Synthetic pathway for the functionalisation of carbohydrates with aminooxy groups.

3. 2 Results

3. 2. 1 Synthesis of a hydrazine derivative of glucose

Our aim was to develop a model reaction for the selective functionalisation of nanocellulose with hydrazine without the requirement of multiple synthetic steps, protecting groups or harsh conditions (**Figure 3.2**).

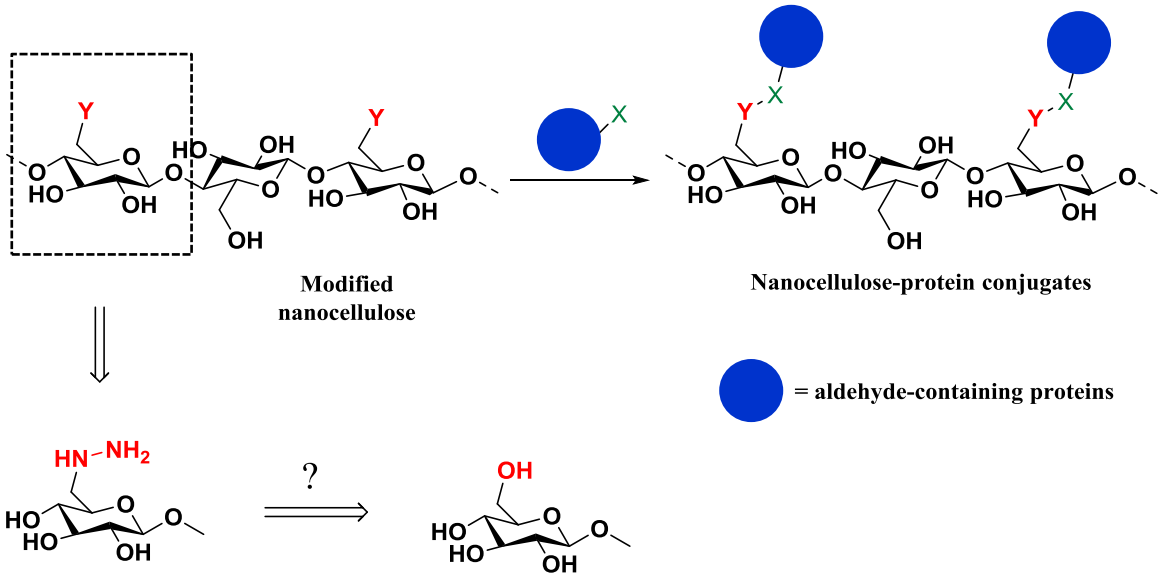
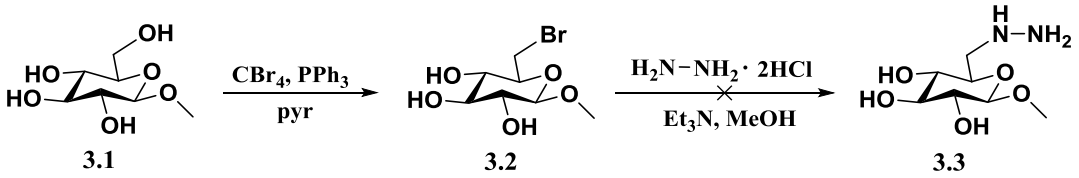


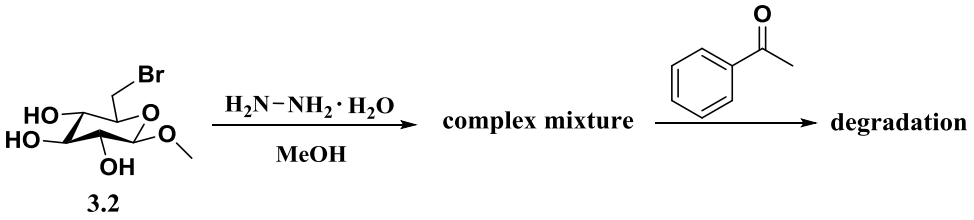
Figure 3.2 Chemoselective ligation between the tagged protein and nanocellulose. The tagged protein is then conjugated with the hydrazine modified nanocellulose in a site-specific fashion.

We designed our synthetic strategy of a minimal model of nanocellulose, to include nucleophilic substitution at C_6 of the methyl glycoside **3.1**. As the primary hydroxyl group is significantly less hindered and thus more nucleophilic than the other hydroxyl groups, it favours selective substitution in the presence of bulky reagents such as triphenylphosphine.[131] We selectively brominated the C_6 of methyl-*β*-D-glucopyranoside under previously reported Appel conditions (carbon tetrabromide/triphenylphosphine)[132] after attempts under *N*-bromosuccinimide/triphenylphosphine[133] conditions provided mixture of carbohydrates by TLC analysis.



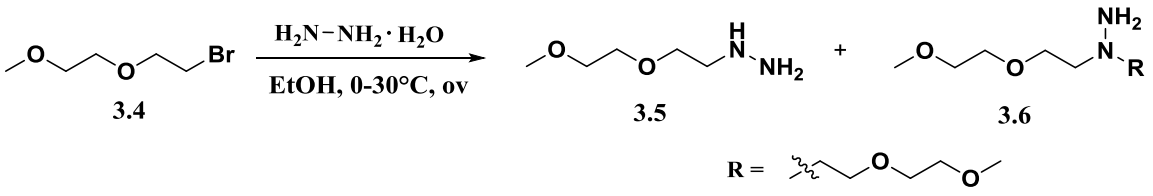
Scheme 3.6 Selective bromination of methyl-*β*-D-glucopyranoside.

We then envisioned transformation at C_6 atom by S_N2 nucleophilic substitution of the bromide into hydrazine. However, under conditions of large excess of hydrazine hydrochloride (8 equiv) and triethylamine, only the starting material was recuperated. Instead, when using hydrazine hydrate as the nucleophilic source, full conversion of the bromo-glucopyranoside was detected by ^{1}H NMR (a highfield shift for the H_A and H_B protons was detected, See Experimental) and ^{13}C NMR (C_6 shift from 33 ppm (C-Br) to 55 ppm (C-N) was detected). However, further elucidation by ^{1}H NMR of the yielded compound (after removal *in vacuo* of excess hydrazine) by reaction with benzaldehyde could not be accomplished due to degradation of the carbohydrate mixture.



Scheme 3.7 Reaction of Br-glucopyranoside with hydrazine hydrate.

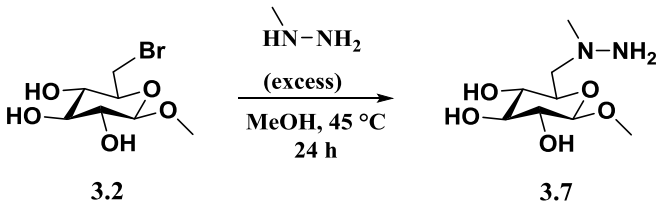
We addressed this question by using a simpler model. Accordingly, when compound **3.4** was treated with an excess of hydrazine hydrate we detected by ^{1}H NMR and ESI the formation of the mono-substituted (**3.5**) and di-substituted (**3.6**) corresponding hydrazines (calc. *m/z* 237, obs. *m/z*: 237 $[M+H]^+$) and hydrazones (after reaction with benzaldehyde, calc. *m/z* 223, obs. *m/z*: 223 $[M+H]^+$ and calc. *m/z* 325, obs. *m/z*: 325 $[M+H]^+$), suggesting that mono-substitution is difficult to control even under excess of hydrazine hydrate.[126]



Scheme 3.8 Reaction of PEG-Br with hydrazine hydrate.

We realised that in the presence of excess hydrazine hydrate compound **3.2** might undergo mono-, di-substitution similarly to compound **3.4**. Thus, when halogenated compounds (**3.2**, **3.4**) react with an excess of hydrazine monohydrate, the resulting mono-substituted hydrazine is a stronger nucleophile than hydrazine monohydrate and as a result di-substitution occurs. In order to overcome multimerization, we envisioned using a mono-substituted hydrazine instead. For this purpose, methylhydrazine hydrate is a good candidate as through inductive effects, the sp^3 lone pair on the substituted nitrogen becomes more basic and thus nucleophilic towards S_N2 displacement of the halogen (i.e. or through the electron donating character of the methyl group the conjugate acid ammonium ions of the secondary amine is better stabilised).

Substitution of bromide was driven to completion by excess of methyl hydrazine to afford compound **3.7**. After *in vacuo* removal of solvent and excess methyl hydrazine, ^{1}H NMR indicated the regioselective formation of the desired hydrazine substituted saccharide in a ratio of 1.4 : 1.0 to methyl hydrazine. Multiple attempts to separate hydrazone-*β*-D-glucopyranoside from the methyl hydrazine mixture, using co-evaporation, lyophilisation or reversed-phase C18 HPLC (See Experimental), failed to yield the desired molecule in high purity and the ^{1}H NMR spectrum still indicated impure compound in a ratio of 1.7 : 1.0 (sugar to methyl hydrazine).



Scheme 3.9 Selective bromide substitution by methyl hydrazine.

When monitoring compound **3.7** by ^{1}H NMR at pH 7, we observed no change in the proton signals over weeks (**Figure 3.10**) indicating that hydrazine-*β*-D-glucopyranoside is a stable molecule.

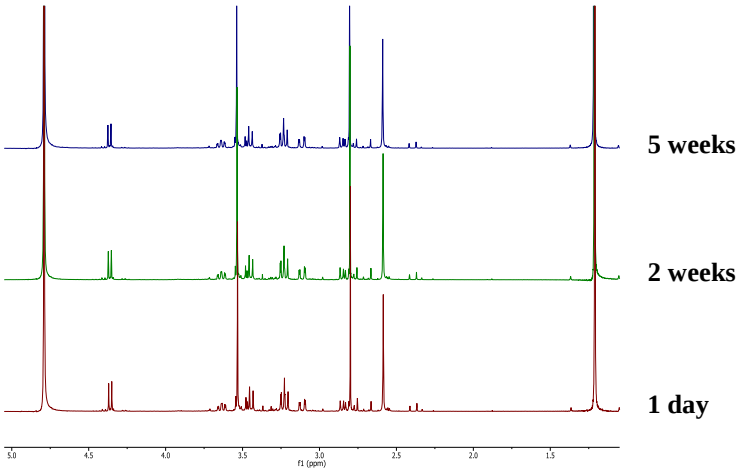
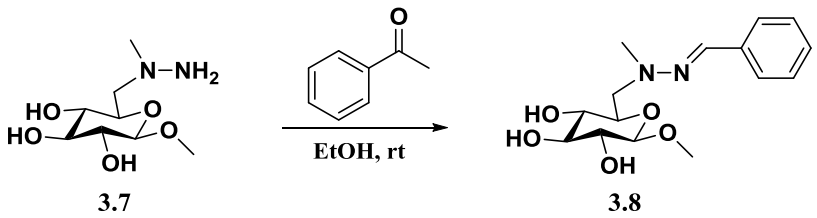


Figure 3.10 The stability of compound **3.7** was assessed by ^{1}H NMR (400 MHz, D_2O) in D_3PO_4 (40 mM) at pD 7; ^{t}BuOH (40 mM) with the characteristic signal at δ 1.17 ppm was used as a reference.

To demonstrate the regioselectivity of the nucleophilic substitution at the C_6 atom of methyl-*β*-D-glucopyranoside, we subjected the yielded compound **3.7** to benzaldehyde and purified the corresponding hydrazone **3.8** by column chromatography.



Scheme 3.11 Synthesis of the corresponding methyl glycoside hydrazone.

NOE analysis of the purified hydrazone **3.8** displays through space interactions between the *N*-substituted methyl protons (H_7) and hydrazone proton (H_8), thus providing direct evidence for substitution at C_6 by the methylated nitrogen of methyl hydrazine (**Figure 3.11**).

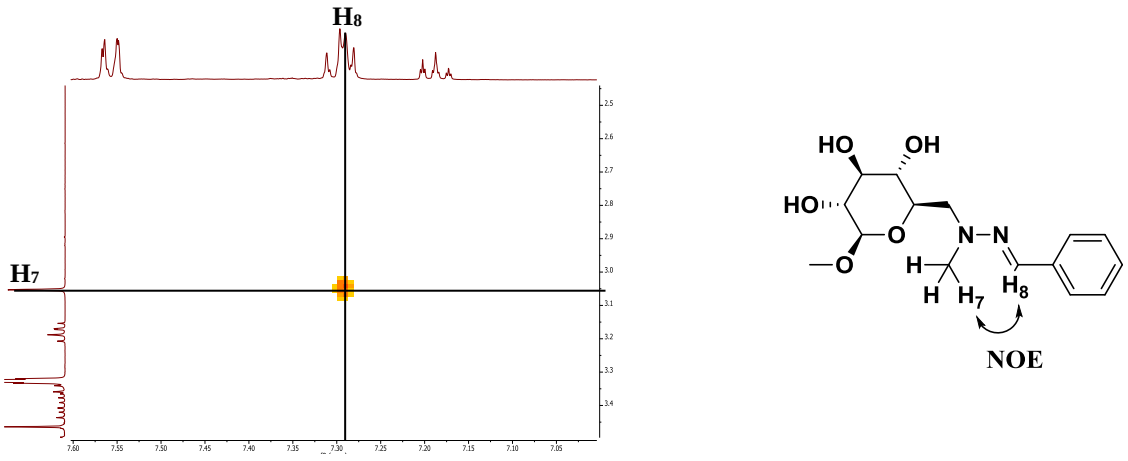


Figure 3.11 Zoom into the NOE spectrum of compound **3.8** showing the cross peak of *N*-substituted methyl protons (**H**$_7$) displaying through space interactions with hydrazone proton (**H**$_8$) and the *E* conformation of the hydrazone.

3. 2. 2 Sugar-protein conjugates

In order to illustrate the utility of the newly synthesised hydrazine glucopyranoside (**3.7**) for the chemoselective protein-sugar ligation, we subjected it to our aldehyde-containing proteins (Ald-GFP, Loop2-GFP and Ald-SLAC) previously described in **Chapter 2**. Aldehyde/hydrazine condensations display rather slow kinetics[120,121], necessitating acidic conditions (pH 4-6) and high concentrations of reactive saccharide in order to achieve complete protein labelling. We found by HRMS that the primary amine of the Tris buffer reacts with the FGly residue (calc. 29,334 Da, obs. 29,333 Da) and therefore for the conjugation reaction we used a phosphate buffer instead.

A typical reaction protocol is described below. To a buffer containing 50 mM Na_2HPO_4, 50 mM NaCl and 0.5 mM DTT, were added Ald-Laccase, Ald-GFP or Loop2-GFP (180 µM) and FGE-4C (18 µM). The enzymes were incubated at 32 °C for 3 h and then hydrazone ligation was initiated by addition of a quenching solution. Aldehyde-tagged protein mixtures (15 µl) were reacted with a quenching solution (25 µl) containing hydrazine-*β*-D-glucopyranoside (10 mM), NaOAc (150 mM), urea (6 M), Na_2HPO_4 (50 mM), NaCl (50 mM) for 4 h, at 32 °C and pH 5. At time various time points (30 s and 4 h) reaction aliquots were quenched by addition of 2 µl TFA (10%) to the samples, before being measured by high-resolution mass spectrometry.

Figures 3.12 and **3.13** show the HRMS spectra of the protein-sugar conjugates before and after selective hydrazone conjugation with the saccharide.

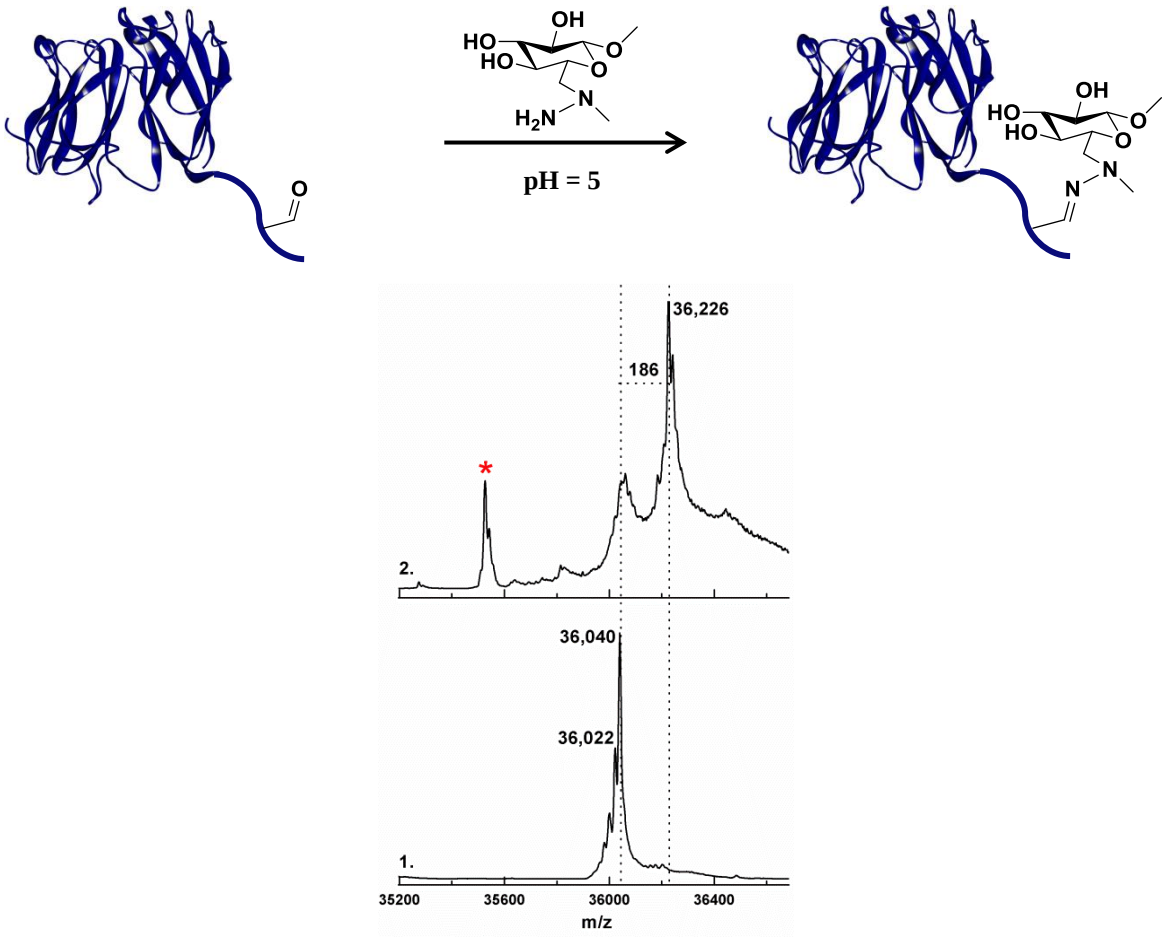


Figure 3.12 Ald-SLAC was oxidized *in vitro* by catalytic amounts of FGE-4C. Then, it was treated with hydrazine methyl *β*-D-glucopyranoside (**3.7**). The corresponding mass of the protein-sugar conjugate was detected by high-resolution mass spectrometry (HRMS). **1.** Ald-SLAC displays an average mass of 36,040 Da (calc. 36,000 Da). **2.** Following 4 h incubation with **3.7**, Ald-SLAC-sugar conjugate displays an average mass of 36,226 Da (calc. 36,186 Da). We detected additional peaks (highlighted by *) corresponding to cleaved Ald-SLAC.

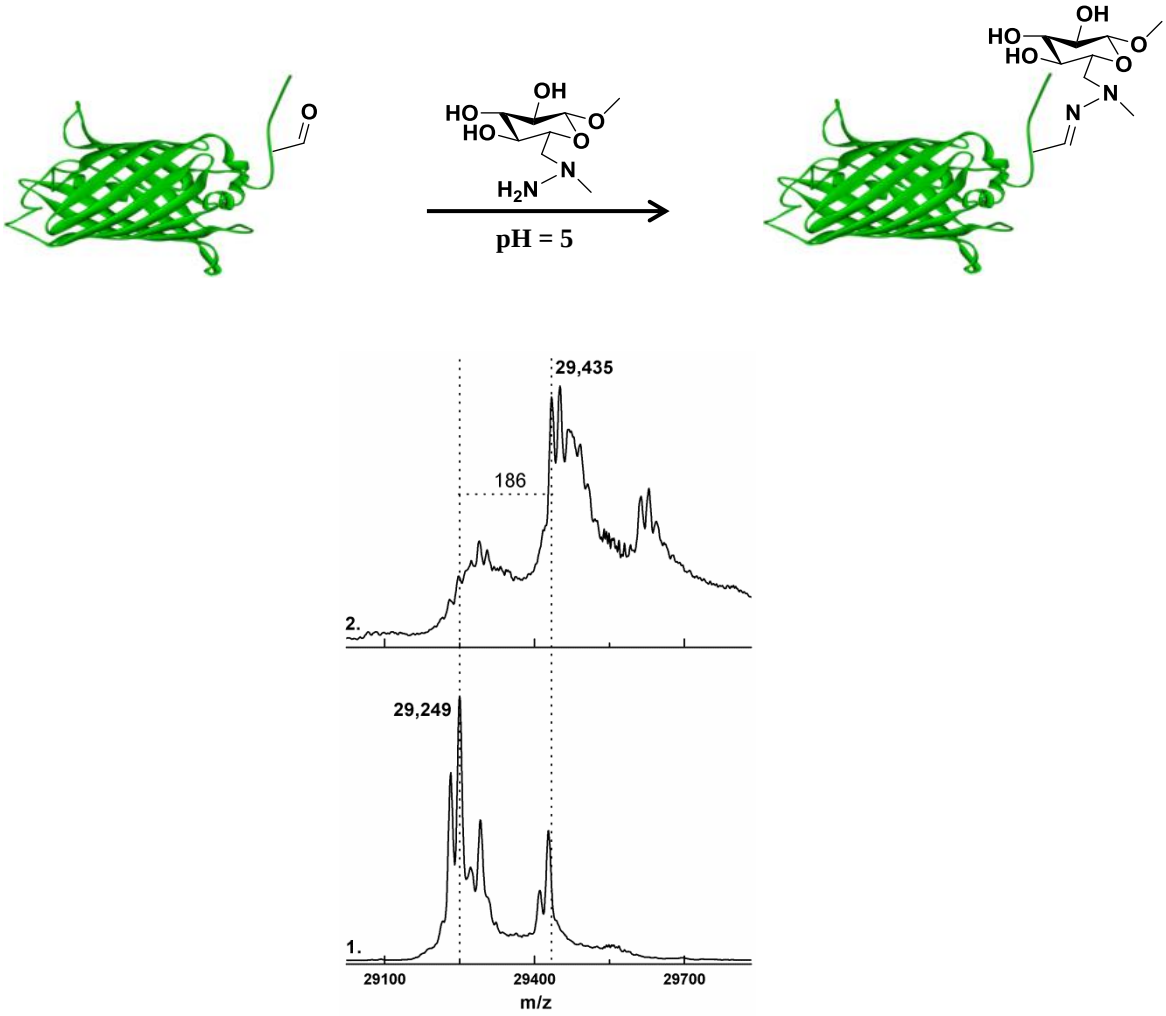


Figure 3.13 Ald-GFP was oxidized *in vitro* by catalytic amounts of FGE-4C. Then, it was treated with hydrazine methyl *β*-D-glucopyranoside (**3.7**). The corresponding mass of the protein-sugar conjugate was detected by high-resolution mass spectrometry (HRMS). **1.** Ald-GFP displays an average mass of 25,249 Da (calc. 25,250 Da). **2.** Following 4 h incubation with **3.7**, Ald-GFP-sugar conjugate displays an average mass of 29,435 Da (calc. 29,436 Da). We detected additional peaks (highlighted by *) corresponding to cleaved Ald-GFP.

3. 2. 2. 1 Chemoselectivity of the hydrazone ligation by tryptic digestion

Subsequently, we demonstrated the chemospecificity of the hydrazone glucopyranoside-protein conjugation by tryptic digestion. By LC-MS, we were able to identify the *N*-terminal 9-residue tryptic fragments corresponding to the mass of the thiol-peptide before oxidation with FGE-4C (i.e. calc. *m/z* 887.5, obs. *m/z* 887.5 $[M+H]^+$, 444.4 $[M+2H]^{+2}$, **Figure 3.14**, **a**), aldehyde- and diol-peptide fragments (i.e. peptide-aldehyde calc. *m/z* 869.4, obs. *m/z*: 869.5 $[M+H]^+$, 435.4 $[M+2H]^{+2}$; peptide-diol calc. *m/z* 887.5, obs. *m/z* 887.5 $[M+H]^+$, **Figure 3.14**, **b**) after oxidation with FGE-4C and the mass of the sugar hydrazone-peptide after oxidation with FGE-4C and reaction with hydrazine glucopyranoside (i.e. calc. *m/z* 1073.6, obs. *m/z*: 1073.5 $[M+H]^+$, 537.4 $[M+2H]^{+2}$, **Figure 3.14**, **c**).

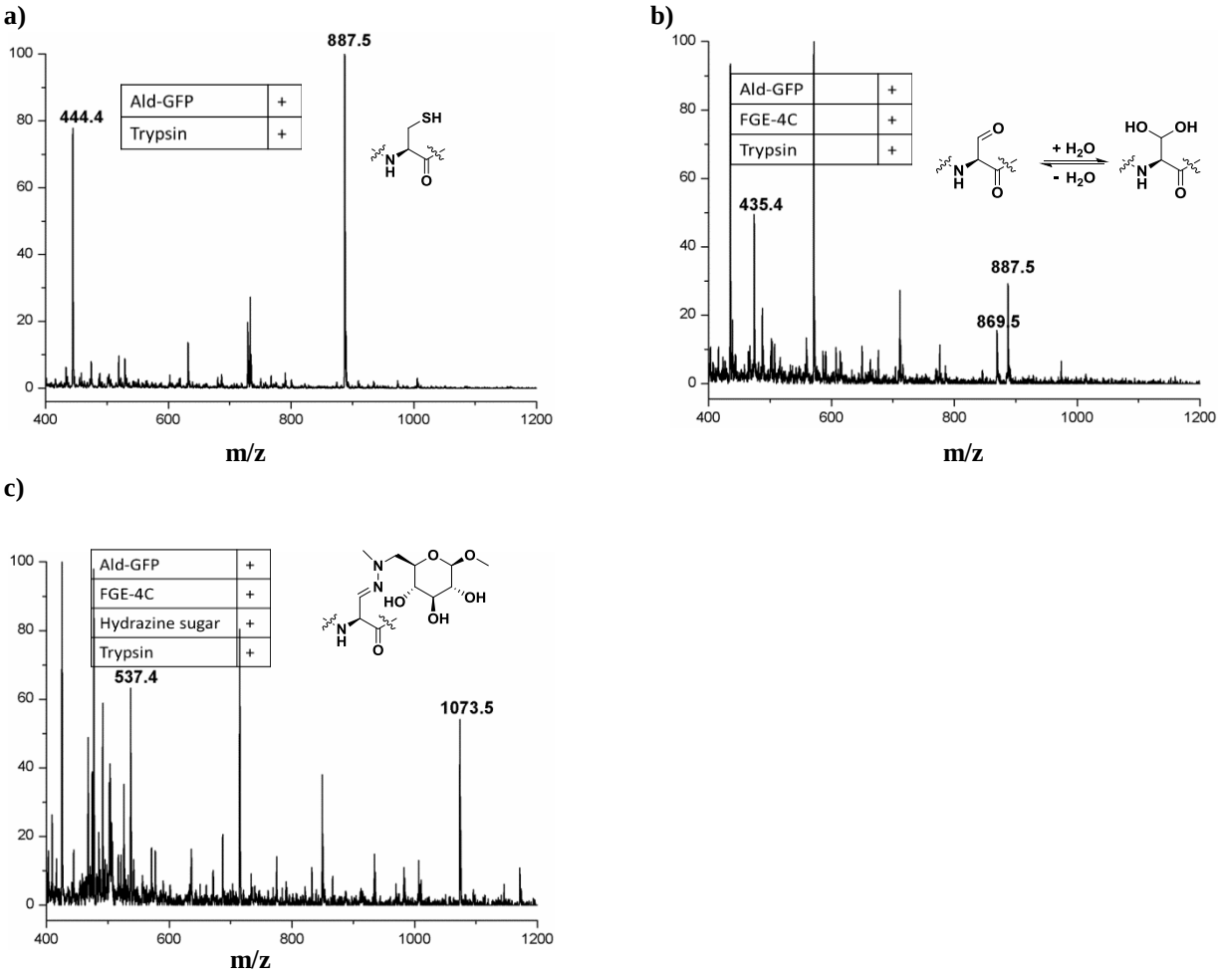


Figure 3.14 a) Ald-GFP was reduced with DTT, and digested with trypsin. **b)** Ald-GFP was reduced with DTT, oxidised with FGE-4C and digested with trypsin. **c)** After incubation with FGE-4C, aldehyde-containing GFP was reacted with compound **3.7** and then digested with trypsin. Reaction mixtures were analysed by LC-MS.

3. 2. 2. 2 Hydrazinolysis of amide bonds by methyl hydrazine

When assessing the chemoselective ligation between Ald-GFP or Ald-SLAC and hydrazine *β*-D-glucopyranoside by HRMS, we detected additional peaks (highlighted by *) corresponding to cleaved Ald-GFP and Ald-SLAC proteins (i.e. detected mass of the cleaved Ald-GFP is 28,776 Da, Δ_{mass} = 473 Da and of Ald-SLAC is 35,527 Da, Δ_{mass} = 513 (473+40) Da). We propose that the cleavage of the aforementioned proteins occurred as a consequence of having used hydrazine *β*-D-glucopyranoside containing methyl hydrazine as an impurity in a ratio of 1.4 eq : 1.0 eq. This proposal is in accordance with observations published by Ohshima and co-workers who described that unactivated amide bonds are efficiently cleaved by hydrazine hydrate in the presence of ammonium salts.[134] Under our ligation conditions the source of ammonium salt (i.e. NH_4Cl) required for the increased cleavage rates[134] is provided during the decomposition of urea into ammonia and cyanate (see **Chapter 2**). As shown in **Figure 3.15**, Ald-GFP and Ald-SLAC proteins might undergo hydrazinolysis of the amide bond which connects the $FGly_5$-Gly_6 residues in the presence of methyl hydrazine. However, this mechanistic proposal shown below does not explain why methyl hydrazine undergoes cleavage at the specific amide bond $FGly_5$-Gly_6.

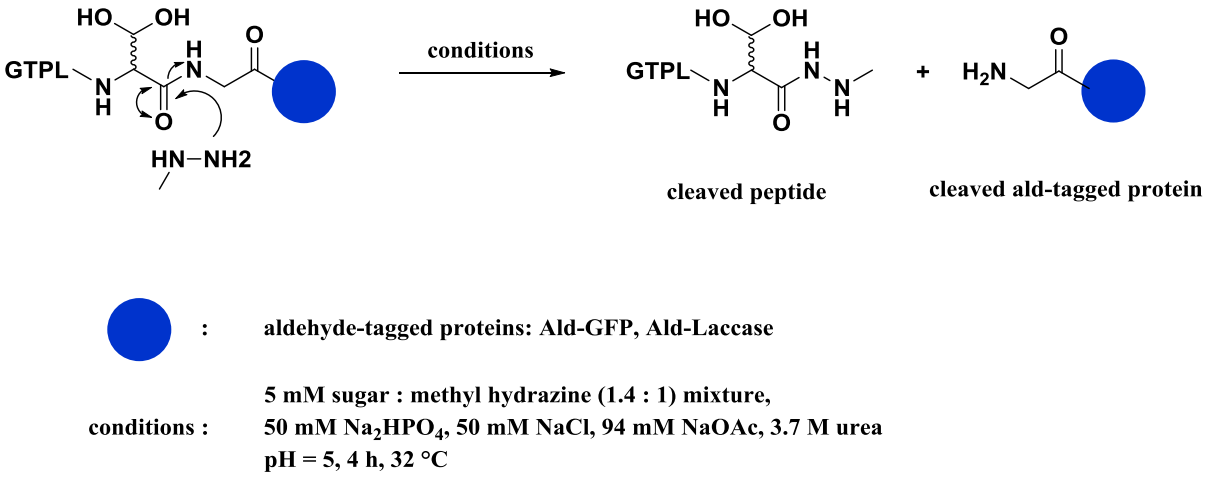


Figure 3.15 Proposed mechanism for the observed cleavage of amide bond between $FGly_5$-Gly_6 by methyl hydrazine. Hydrazinolysis of the amine bonds provides cleaved ald-tagged proteins and peptide (GTPLFGly) (mass of peptide - calc. *m/z* 474, obs. *m/z*: 473, representing the Δ_{mass} between the diol-containing protein and cleaved-protein).

Alternatively, we propose that the aldehyde functionality on the $FGly_5$ residue promotes the peptide cleavage by first reacting with methyl hydrazine and yielding the corresponding hydrazone protein (**Figure 3.16**). Then, the alkylated amine of the afforded hydrazone undergoes intramolecular nucleophilic addition to the amide bond which connects the $FGly_5$-Gly_6 residues. The resulted cleaved peptide bears a stable pyrazolone moiety. Substituted pyrazolones (i.e. thiopyrazolone) were used by Rabuka and co-workers for generating site-selectively modified antibody conjugates (α-HER2-thioPz-ADC).[135] Their strategy followed nucleophilic addition of thiopyralozone to the FGly residue, inserted at the C-terminus of the antibody, by a trapped-Knoevenagel condensation.

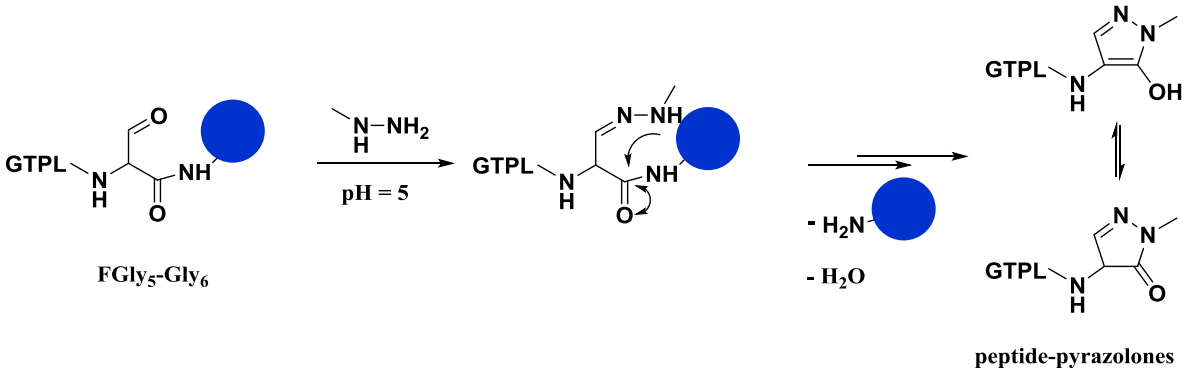


Figure 3.16 Proposed mechanism for the FGly promoted amide bond cleavage between $FGly_5$-Gly_6 by methyl hydrazine.

When monitoring the conversion of Loop2-GFP to Loop2-GFP-saccharide by HRMS, we detected the expected mass of the corresponding conjugate (calc. 30,445 Da, obs. 30,444 Da) (**Figure 3.17**). At the same time, we did not observe additional peaks corresponding to the hydrazinolysed protein. This observation suggests that when the aldehyde functionality is placed in an internal loop, the peptidic backbone is more rigid and therefore cannot undergo amide cleavage by intramolecular nucleophilic addition.

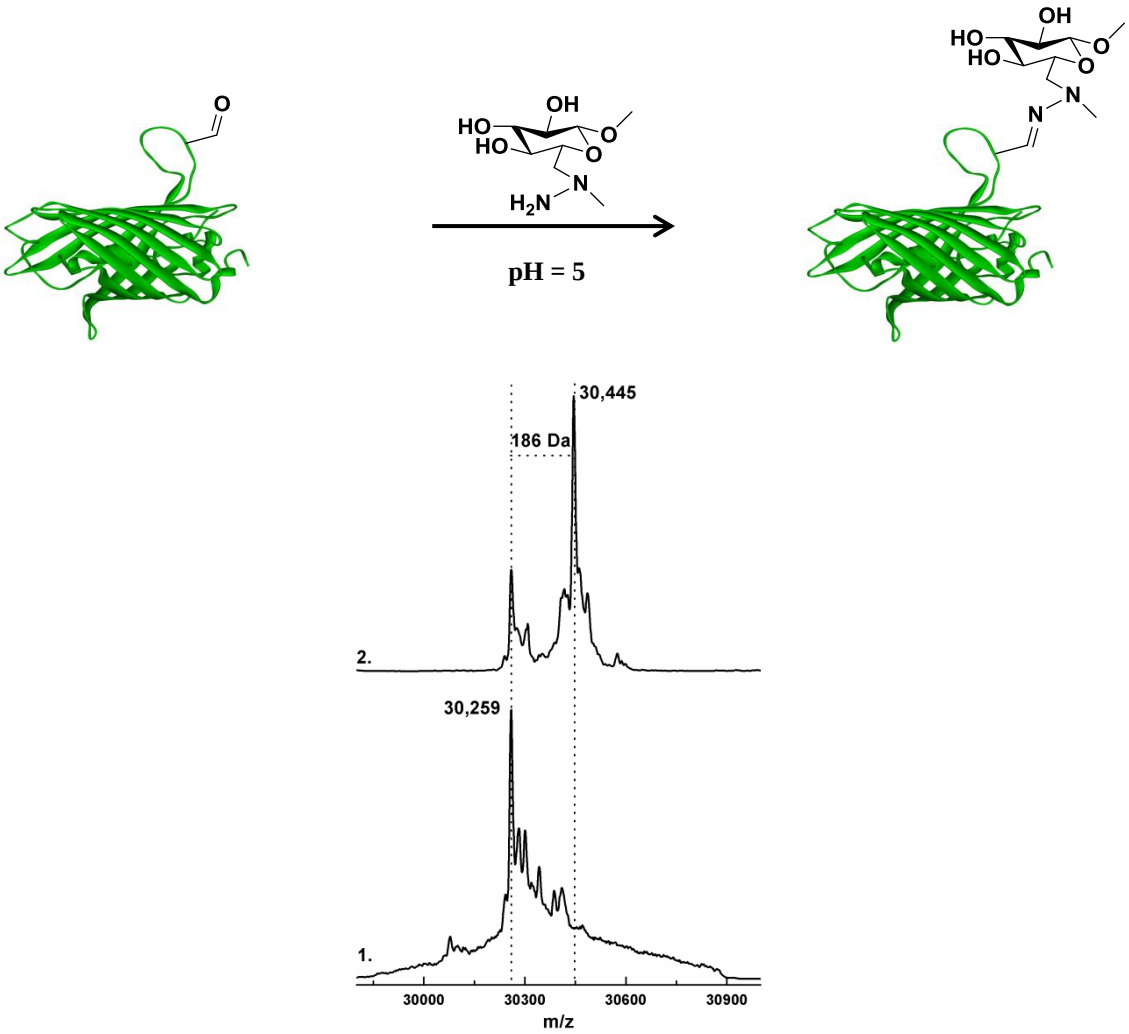


Figure 3.17 Loop2-GFP was oxidized *in vitro* by catalytic amounts of FGE-4C. Then, it was treated with hydrazine methyl *β*-D-glucopyranoside (**3.7**). The corresponding mass of the protein-sugar conjugate was detected by high-resolution mass spectrometry (HRMS). **1.** Loop2-GFP displays an average mass of 36,259 Da (calc. 36,260 Da). **2.** Following 4 h incubation with **3.7**, Loop2-GFP-sugar conjugate displays an average mass of 30,445 Da (calc. 36,446 Da).

3. 3 Discussion

To generate tailored nanocellulose-protein conjugates, we required a strategy for selectively synthesising hydrazine carbohydrates without the use of protecting groups and minimal number of synthetic steps. Although many synthetic methods for the synthesis of mono-substituted alkyl and aryl hydrazines exist[126], they usually require the use of a Boc-protected hydrazine[128], metal catalysts[136], multiple synthetic steps[130] or harsh conditions. In principle, such synthetic strategies could not be employed for the modification of nanocellulose as the crystalline integrity and morphology of the material could be affected by acidic conditions. To combat this possibility, we foreseen the carbohydrate functionalization by direct attachment of a hydrazine group at the C_6 atom of the glycosyl monomer *via* nucleophilic substitution. Prior to hydrazination, selective halogenation at the C_6 atom of the glucopyranoside proceeded under mild bromination conditions.

When using hydrazine as the nucleophilic source to undergo bromide displacement, we observed that substitution gave rise to mixtures and resulted in degradation of the starting material. In fact, Mayr and co-workers demonstrated that the nucleophilicity of the α effect groups such as hydrazines and hydroxylamines are similar to that of alkylamines (i.e. the reactivity of hydrazine is similar to that of methylamine).[121] Even if hydrazine displays 100 times and hydroxylamine 3 times higher nucleophilicity with respect to ammonia, the displayed α-effect is smaller than the activating effect of the alpha-methyl group in methyl amine.[121] Regarding methylhydrazine, alkylation increases the nucleophilicity of the substituted nitrogen while decreasing the reactivity of the adjacent centre. When employing methylhydrazine as the nucleophilic source for undergoing halogen displacement, regioselective insertion of the desired hydrazine functionality was observed. The 2-step synthesis of compound **3.7** emphasises the applicability of this procedure for transforming crystalline nanocellulose into suitable coupling partners for aldehyde-containing proteins.

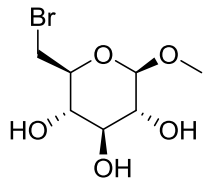
Novel synthetic glycoproteins were obtained by the highly selective coupling of the nucleophilic carbohydrate to the aldehyde-containing proteins (i.e. Ald-GFP, Ald-SLAC and Loop2-GFP). The ligation proceeded chemoselectively in aqueous buffer at pH 5, as demonstrated by HRMS analyses. It should be borne in mind that these reactions took place without prior need of purification steps of the hydrazine sugar. Nevertheless, hydrazone ligation of aldehyde-containing proteins to carbohydrates does require high concentrations of the sugar reactant (5 mM).[119,120] To overcome sluggish conversions one could explore the use of aniline, as catalytic amounts of this compound were reported to have increased the rates of oxime ligation.[137] New strategies for attaching sugars to proteins are of paramount importance for the generation of

glycoconjugates. The facile synthesis of hydrazine carbohydrate highlights the utility of this approach for the selective coupling of crystalline cellulose to proteins.

3. 4. Experimental

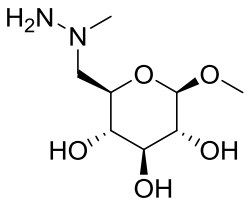
3. 4. 1 Chemical synthesis

(2S,3S,4S,5R,6R)-2-(bromomethyl)-6-methoxytetrahydro-2H-pyran-3,4,5-triol (1)[1]



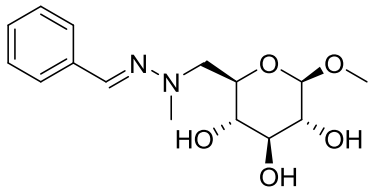
To a stirred solution of methyl β-D-glucopyranoside (0.40 g, 2.06 mmol, 1.0 **eq.**) in anhydrous pyridine (12 ml) at 0 °C under N_2 were added triphenylphosphine (1.08 g, 4.12 mmol, 2.0 **eq.**) and carbon tetrabromide (1.02 g, 3.07 mmol, 1.5 **eq.**). The reaction was heated up to 75 °C, stirred for 5 h and quenched at 0 °C by addition of methanol (10 ml). The reaction mixture was concentrated *in vacuo* and the crude material was purified by column chromatography (*eluent*: CH_2Cl_2/MeOH 20:1) to afford the title compound as a white solid (0.30 g, 1.17 mmol, 57 %). **^{1}H NMR** (400 MHz, CD_3OD) δ 4.20 (d, J = 7.78 Hz, 1H), 3.77 (dd, J = 11.04, 2.26 Hz, 1H), 3.56-3.52 (m, 4H), 3.42-3.33 (m, 2H), 3.26-3.24 (m, 1H), 3.16 (dd, J = 9.16, 7.78 Hz, 1H). **^{13}C NMR** (400 MHz, CD_3OD) δ 105.4, 77.9, 76.6, 75.2, 73.8, 57.4, 33.9. **HMRS** $[M+Na+H^+]^+$: 278.9843 (calc. 278.9839), 280.9823 (calc. 280.9818).

(2R,3R,4S,5S,6R)-2-methoxy-6-((1-methylhydrazinyl)methyl)tetrahydro-2H-pyran-3,4,5-triol (2)



To a stirred solution of **1** (0.20 g, 0.78 mmol) in anhydrous methanol (8 ml) at 0 °C was added an excess of methyl hydrazine (3×0.2 ml).The reaction under N_2 was heated up to 45 °C and stirred for 30 h. The reaction mixture was concentrated *in vacuo* and the resulting yellow oil was used for labelling aldehyde-containing proteins without further purification. *The main impurity is methyl hydrazine δ 2.78 (s, 3H), with the methyl hydrazine:sugar ratio being 1:1.4. **^{1}H NMR** (400 MHz, CD_3OD) δ 4.22 (d, J = 7.91 Hz, 1H), 3.58 (ddd, J = 9.79, 7.65, 2.51 Hz, 1H), 3.54 (s, 3H), 3.39-3.33 (m, 1H), 3.29-3.14 (m, 3H), 3.03-2.95 (m, 1H), 2.76 (s, 3H). **^{13}C NMR** (400 MHz, CD_3OD) δ 105.6, 77.8, 74.9, 74.5, 73.4, 62.4, 57.8, 47.58, 38.32. **LCMS** $[M+H]^+$: 223.39 (calc. 223.24). **HRMS** $[M+H^+]^+$: 223.1289 (calc. 223.1288), $[M+Na+H^+]^+$: 245.1106 (calc. 245.1108).

(2R,3S,4S,5R,6R)-2-((2-((E)-benzylidene)-1-methylhydrazinyl)methyl)-6-methoxytetrahydro-2H-pyran-3,4,5-triol (3)

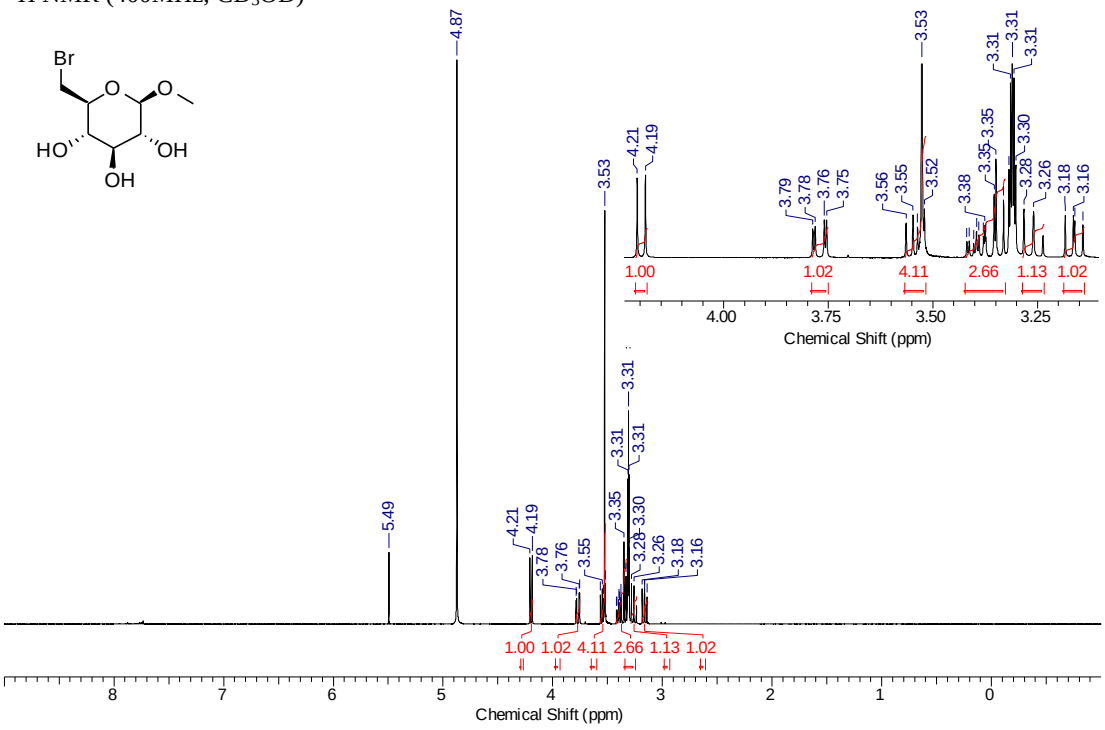


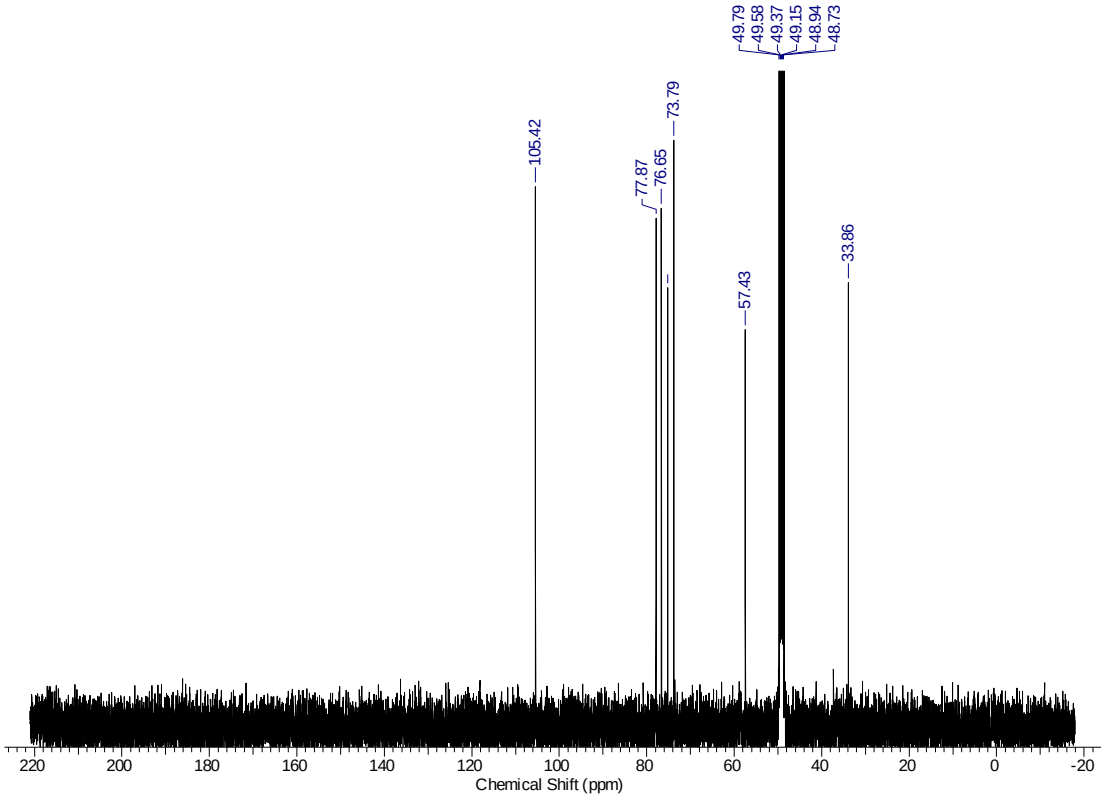
yield not determined

^{1}H NMR (400 MHz, CD_3OD) δ 7.57-7.52 (m, 2H), 7.32-7.25 (m, 1H), 7.21-7.15 (m, 1H), 4.10 (d, J = 7.65 Hz, 1H), 3.93 (dd, J = 14.37, 1.82 Hz, 1H), 3.53-3.48 (m, 1H), 3.45 (s, 3H), 3.44-3.34 (m, 2H), 3.20-3.13 (m, 2H), 3.04 (s, 3H). **^{13}C NMR** (400 MHz, CD_3OD) δ 138.8, 132.5, 129.5, 128.2, 126.7, 105.4, 78.1, 76.8, 75.3, 73.6, 60.5, 57.3, 39.6. **HRMS** $[M+H]^+$: 311.1604, (calc. 311.1601) and $[M+Na+H^+]^+$: 333.1425 (calc. 333.1421).

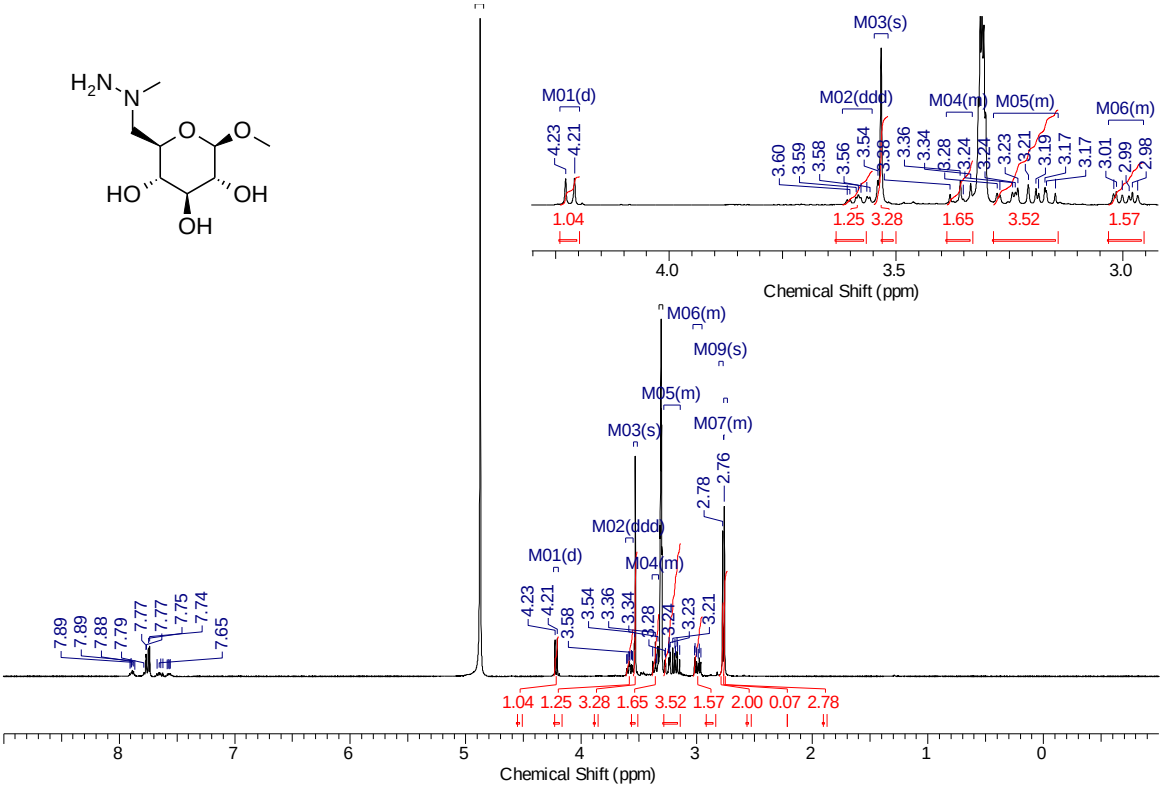
(2S,3S,4S,5R,6R)-2-(bromomethyl)-6-methoxytetrahydro-2H-pyran-3,4,5-triol (1).

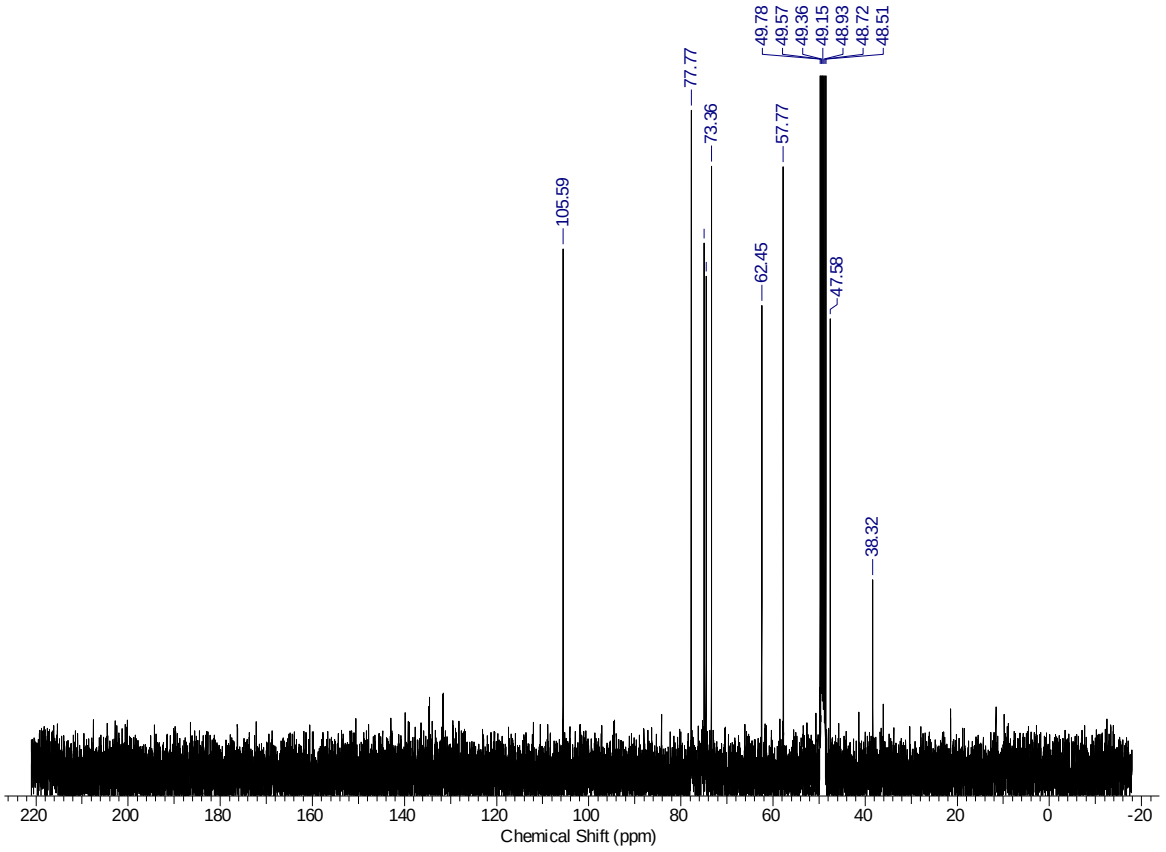
^{1}H NMR (400MHz, CD_3OD)



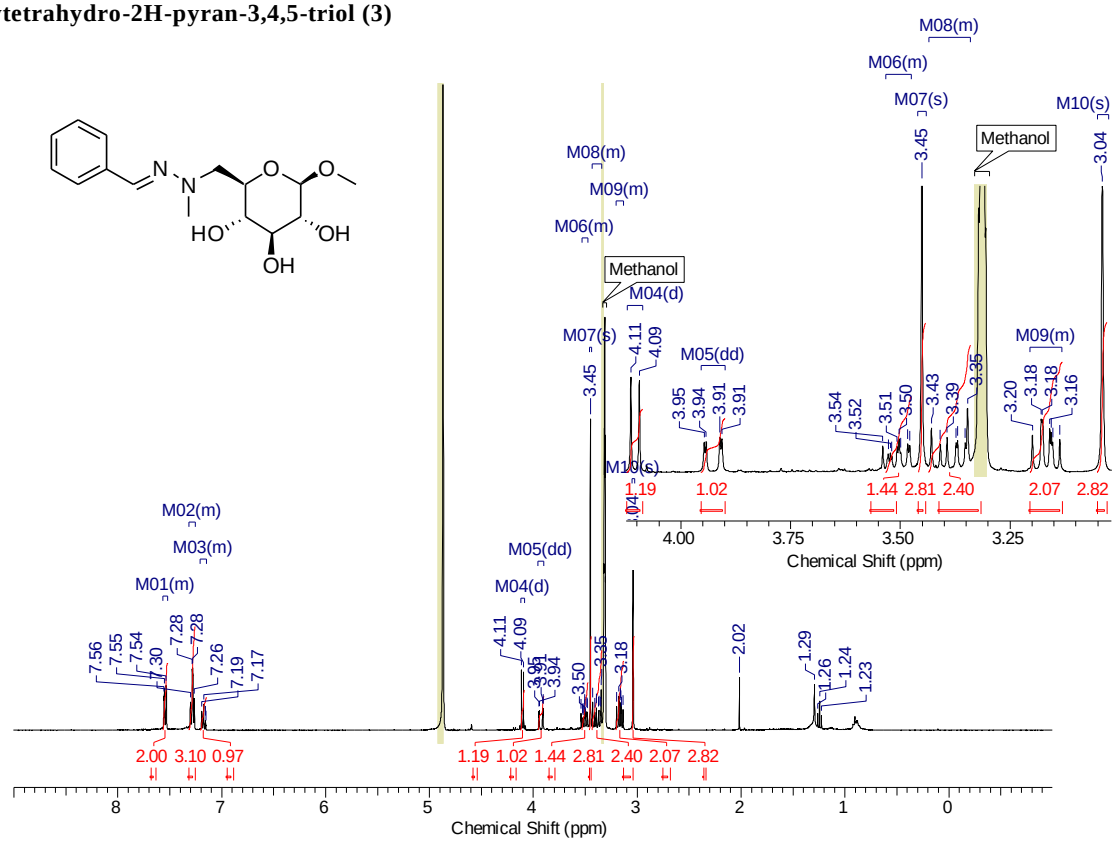


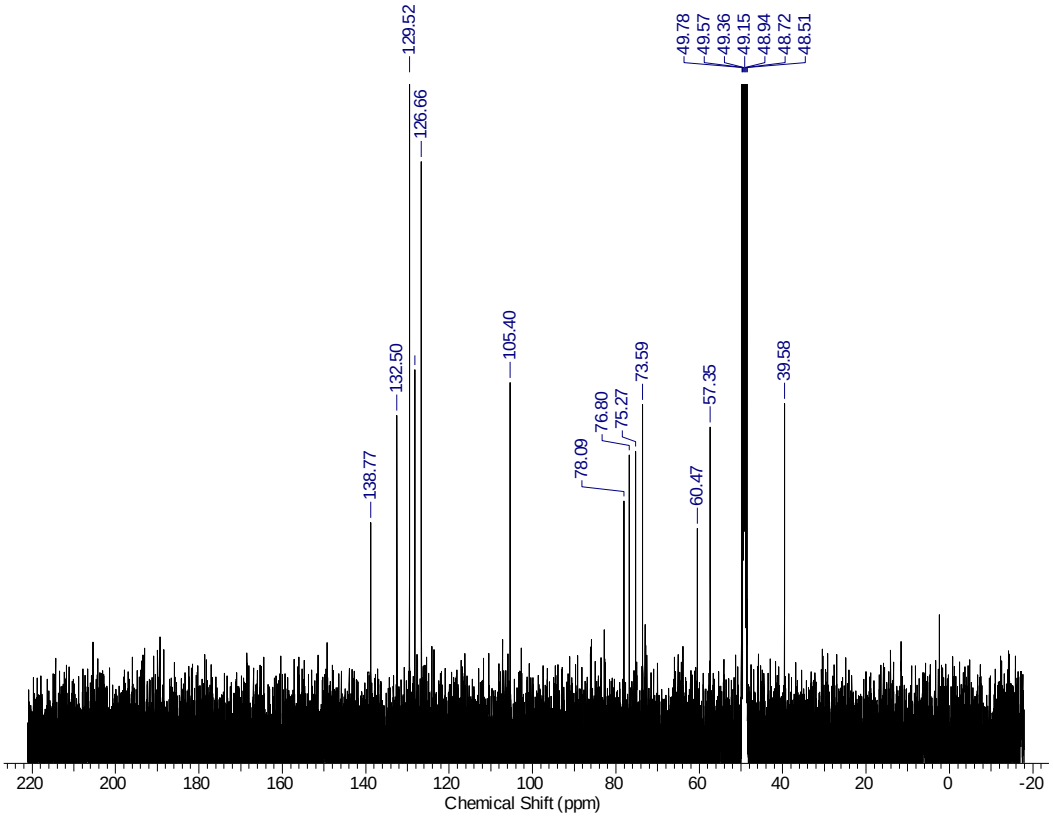
(2R,3R,4S,5S,6R)-2-methoxy-6-((1-methylhydrazinyl)methyl)tetrahydro-2H-pyran-3,4,5-triol (2). ^{1}H NMR (400MHz, CD_3OD)





(2R,3S,4S,5R,6R)-2-((2-((E)-benzylidene)-1-methylhydrazinyl)methyl)-6-methoxytetrahydro-2H-pyran-3,4,5-triol (3)





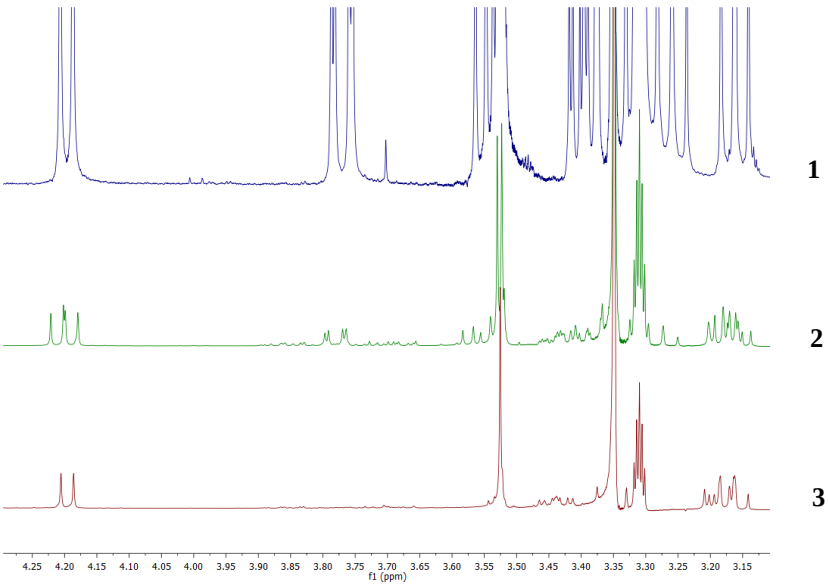


Figure 3.18 Reaction of compound **3.2** (**1**) with hydrazine hydrate was monitored by ^{1}H NMR. ^{1}H NMR spectrum **2** indicates reaction monitor and ^{1}H NMR spectrum **3** displays the conversion of the starting material.

4. 2 Fluorescent labelling of cellulose nanocrystals

Cellulose nanocrystals (CNCs) are rod-shaped nanoparticles obtained from native cellulose by acid hydrolysis.[138] The potential use of functionalised nanomaterials for bioimaging[139], sensor applications[140] and study of cell uptake has generated interest for labelling nanocellulose with fluorescent probes[141].

Towards this goal, Cranston and co-workers described the fluorescent labelling of CNCs with fluorescein derivative (DTAF) by a simple one-step procedure (**Figure 4.1**, left).[138] The DTAF-modified CNCs were characterised by atomic force microscopy (AFM) and it was found that the morphology of the pristine material was maintained. Moreover, the fluorescent conjugates were analysed by fluorescence spectroscopy and depending on the sulfonation degree of the starting material, the reported values for the degree of substitution range from 0.7 to 7.7 DTAF molecules per $\mathbf{10^5}$ anhydroglucose units (AGUs). The described fluorescent nanoconjugates might find use as optical markers by dispersing them on polyvinyl alcohol fibers.[138]

In a separate study, *N*-hydroxysuccinimide-modified rhodamine B was covalently attached to the surface of cellulose nanofibrils (CNFs).[140] The luminescent nanoparticles (**Figure 4.1**, right) with potential uses in sensor and biomedical applications, were characterised by fluorescence spectroscopy and confocal laser scanning microscopy. By UV/visible spectroscopy, the degree of substitution (labelling) of one rhodamine molecule to $\mathbf{1.65\times10^3}$ AGUs was calculated.[140]

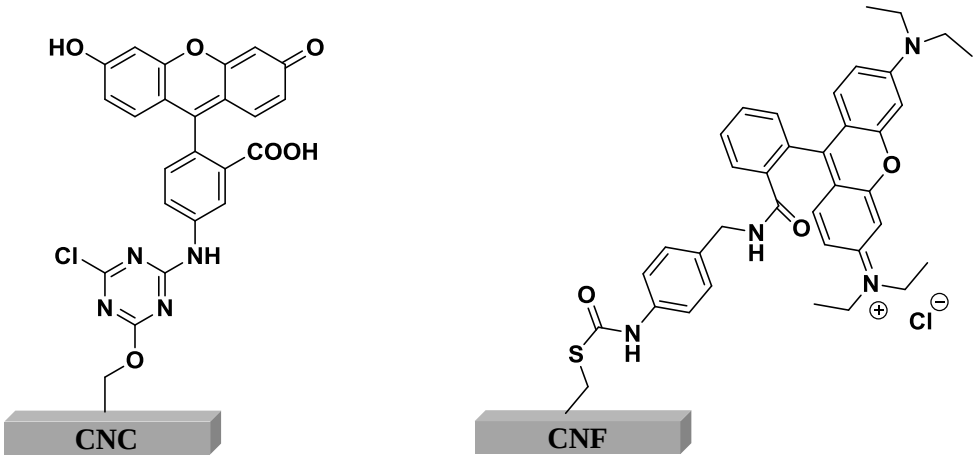


Figure 4.1 Illustration of fluorescently labelled nanocellulose: DTAF-CNC (left) and rhodamine B-CNF (right).

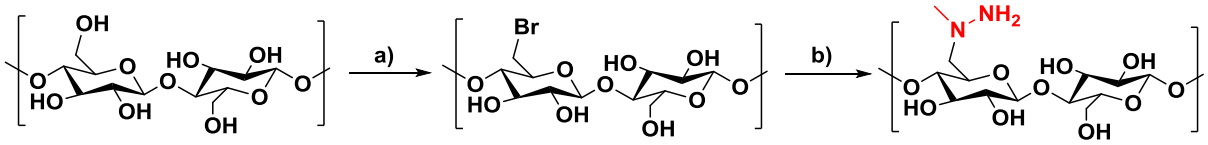
Recently, the selective decoration of cellulose nanofibrils surface with methylcoumarin and fluorescein derivatives was reported (**Figure 4.2**).[142] The obtained multicolour CNFs with potential applications in multimodality imaging, was qualitatively characterised by UV/visible and fluorescence spectroscopies. The degree of substitution was estimated by solid-state ^{13}C NMR to be **1%**.[142] Moreover, Roman and co-workers tested the potential of using functionalised CNCs as drug delivery vectors.[141,143] For this purpose, they covalently immobilised folic acid on the surface of fluorescein isothiocyanate-conjugated CNCs. By UV/visible spectroscopy the degree of substitution of one molecule of fluorescein or folic acid to **1.5×10** or **2.2×10** AGUs was determined.[141] The designed folic acid-CNC conjugates were uptaken selectively by folate receptor-positive brain tumour cells, as demonstrated by fluorescence microscopy.

4. 2 Results

4. 2. 1 Chemical modification of crystalline nanocellulose

In **Chapter 3**, we described the chemoselective coupling of aldehyde-containing proteins to hydrazine functionalised methyl glycoside. This serves as a model strategy for obtaining the selective coupling of crystalline cellulose (CNC) to fluorescent proteins.

Towards this goal, cellulose nanocrystals, provided by EMPA, were chemically modified (i.e. brominated and hydrazinated), as illustrated in **Scheme 4.1**. The halogenation conditions were adapted from those used to selectively brominate methyl glycoside. As nanocellulose is a suspension in pyridine, the temperature of the reaction was increased from 75 °C to 100 °C and the number of equivalents of the triphenylphosphine and carbon tetrabromide reagents was doubled (4.0 eq reagents to 1.0 eq CNC). The product, CNC-Br was isolated from the reaction mixture by centrifugation after being extensively washed (i.e. re-suspended in methanol, sonicated and/or vortexed) and then collected by centrifugation. The selectively brominated crystals were subsequently isolated by freeze drying the aqueous suspension and then used for the hydrazination reaction. CNC-Br crystals were sonicated in anhydrous methanol and to the suspension an excess of methyl hydrazine was added over a period of 30 h at 46 °C. The modified nanocrystals, CNC-Hydra were purified by filtration. The particles were re-suspended in methanol, sonicated and/or vortexed, and then collected by centrifugation. The collected particles were dispersed in water and isolated by freeze drying the aqueous suspension.



Scheme 4.1 Synthetic route for the selective modification of CNC. **a)** CBr_4, PPh_3, pyr, 100 °C **b)** MeN_2H_4, MeOH, 46 °C.

Following this two-step procedure the primary hydroxyl groups on the surface of CNC were selectively converted into hydrazines. After each chemical step on CNC, the morphological and crystalline integrity of the modified nanocellulose was determined by transmission electron microscopy (TEM) and X-ray diffraction (See **4.3 Discussion**). It was found that the crystalline,

structural and dimensional properties of cellulose nanocrystals were maintained over the two steps.

4. 2. 2 Qualitative analyses of the CNC - GFP conjugates

In the jellyfish *Aequorea victoria*, GFP absorbs, through its light-absorbing chromophore formed by the cyclization reaction of three residues Ser_{65}, Tyr_{66} and Gly_{67}, the blue light (470 nm) emitted by aequorin into green light (508 nm).[144] The intrinsic fluorescence of GFP enables the use of fluorescence techniques such as flow cytometry, fluorescence spectroscopy and microscopy to study the selective interaction with CNC. As shown in **Figure 4.3**, the absorption spectrum of our GFP variant shows two peaks, at 279 nm and 475 nm. When excited at 475 nm, Ald-GFP emits a maximal fluorescence at 504 nm (**Figure 4.3**, right).

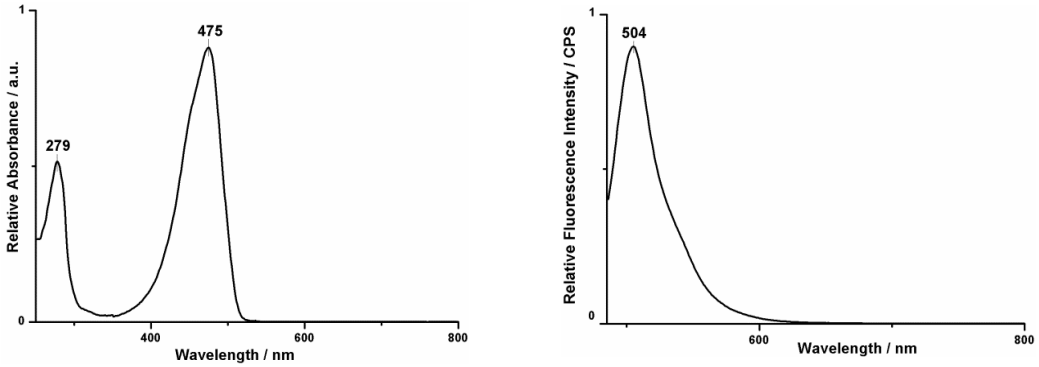


Figure 4.3 Left. UV/Vis spectrum of Ald-GFP showing the absorption peaks at 279 nm and 475 nm. **Right.** Spectrum of Ald-GFP showing the maximal intensity of the fluorescence emission at 504 nm, when excited at 475 nm.

Fluorescently labelled cellulose nanocrystals were prepared by conjugation of Ald-GFP and Loop2-GFP to CNC-Hydra. A typical protein-CNC reaction was conducted in 200 µl 50 mM phosphate, 200 mM NaCl buffer (pH 6.5) containing CNC-Hydra (5 mg) and bovine serum albumin (BSA, 17 µM). To the suspension, Ald-GFP or Loop2-GFP (50 µM) which prior to conjugation were treated with catalytic amounts of FGE-4C, were added. At the same time, the control reactions indicated in **Table 4.1** were performed. After overnight incubation, the CNC-protein conjugates were isolated by centrifugation, after repeated washing with 1M Tris/NaCl.

Table 4.1 Reactions of cellulose nanocrystals with Ald-GFP and Loop2-GFP.

Entry	Crystalline nanocellulose	Protein	Fluorescence Outcome
1	CNC-Hydra	Ald-GFP*	fluorescent
2	CNC-OH	Ald-GFP*	non-fluorescent
3	CNC-Hydra	Ald-GFP	non-fluorescent
4	CNC-OH	Ald-GFP	non-fluorescent
5	CNC-Hydra	Loop2-GFP*	fluorescent
6	CNC-OH	Loop2-GFP*	non-fluorescent
7	CNC-Hydra	Loop2-GFP	non-fluorescent
8	CNC-OH	Loop2-GFP	non-fluorescent

The proteins highlighted by * were treated with FGE-4C prior to conjugation.

Figure 4.4 (b, **e)** shows pictures of the labelled and unlabelled nanocrystals which sedimented from the aqueous suspensions of the aforementioned reactions. The CNC-Hydra particles which were conjugated with Ald-GFP* (**b**, **1**) or Loop2-GFP* (**e**, **5**) were green, whereas the conjugates isolated from the control reactions appeared to maintain the colour of the pristine material. Dilutions of the protein-particles conjugates (55.7 µg/ml) were analysed by fluorescence spectroscopy using an excitation wavelength of 475 nm. As illustrated in **Figure 4.4** (**c**, **f**), the fluorescence emission spectra corresponding to the nanoparticles CNC-Hydra conjugated to Ald-GFP* and Loop2-GFP* displayed the highest emission. At the same time, unlabelled CNC (spectra **2 - 4** and **6 - 8** in **Figure 4.2**, **c**, **f**) showed negligible emission peaks in the wavelength range of 480 nm to 650 nm.

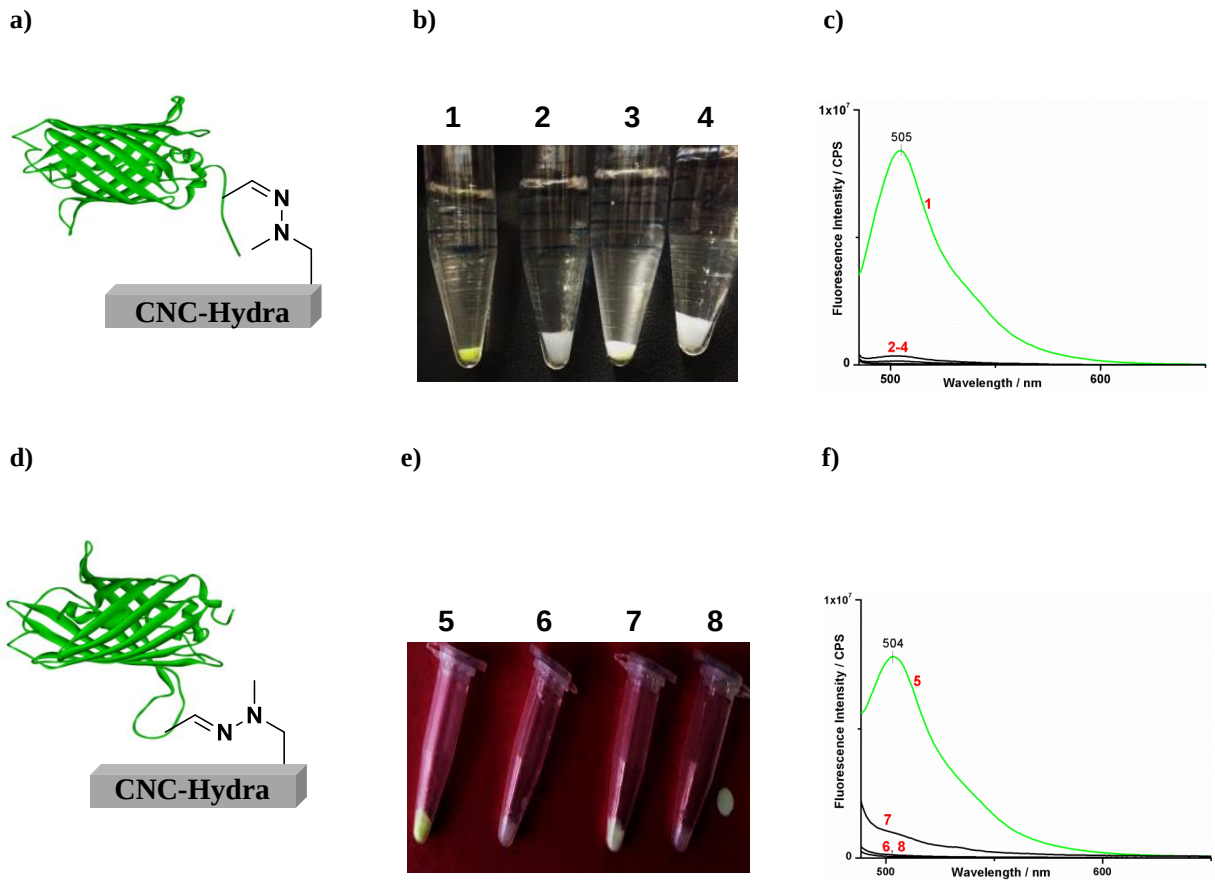


Figure 4.4 Bio-conjugates of aldehyde-tagged proteins with crystalline nanocellulose. **a)** Schematic representation showing the attachment site of Ald-GFP* to CNC-Hydra. **b)** Image of labelled and unlabelled CNC/CNC-Hydra with Ald-GFP* and Ald-GFP. **c)** Fluorescence emission spectra of aqueous suspensions of fluorescently labelled cellulose nanocrystals (**1**, 55.7 µg/ml) using an excitation wavelength of 475 nm. Fluorescence emission spectra (**2 - 4**) correspond to the particles isolated from the control reactions. **d)** Schematic representation showing the attachment site of Loop2-GFP* to CNC-Hydra. **e)** Image of labelled and unlabelled CNC with Loop2-GFP. **f)** Fluorescence emission spectra of aqueous suspensions of fluorescently labelled cellulose nanocrystals (**5**, 167.1 µg/ml) using an excitation wavelength of 475 nm. Fluorescence emission spectra (**6 - 7**) correspond to the particles isolated from the control reactions.

The selectivity of the method was further verified by additional control experiments in which we incubated Ald-GFP*/Ald-GFP with crystalline nanocellulose which was reacted only with CBr_4 in pyridine (**Table 4.2**, entry **2**) or PPh_3 (**Table 4.2**, entry **3**) in pyridine and with CNC-Br (**Table 4.2**, entry **4**). Dilutions of the isolated protein-particles conjugates (55.7 µg/ml) were analysed by fluorescence spectroscopy. As illustrated in **Figure 4.5** (spectra **4**, **5**), fluorescence emission can be seen in the control experiments where Ald-GFP* or Ald-GFP were reacted with CNC-Br. At the same time, no interactions were observed between Ald-GFP* and the other

modified particles, as these materials did not show any emission peaks in the wavelength range of 480 to 650 nm (**Figure 4.5**, spectra **1-3**).

Table 4.2 Reactions of CNC after different chemical treatments with GFP containing aldehyde or cysteine.

Entry	Sample Name	Treatment	Protein	Fluorescence Outcome
1	CNC-OH	n/a	Ald-GFP*	non-fluorescent
2	CNC-OH (CBr_4)	CBr_4	Ald-GFP*	non-fluorescent
3	CNC-OH (PPh_3)	PPh_3	Ald-GFP*	non-fluorescent
4	CNC-Br	**a)** CBr_4, PPh_3	Ald-GFP*	fluorescent
5	CNC-Br	**a)** CBr_4, PPh_3	Ald-GFP	fluorescent

The proteins highlighted by * were treated with FGE-4C prior to conjugation.

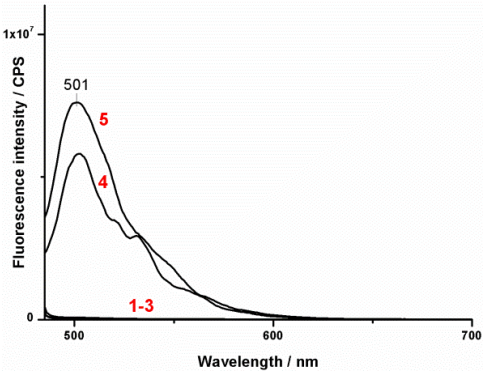


Figure 4.5 Fluorescence emission spectra of aqueous suspensions (55.7 µg/ml) of modified cellulose nanocrystals and Ald-GFP*/Ald-GFP using an excitation wavelength of 475 nm.

When CNC-Br is incubated with Ald-GFP* or Ald-GFP, protein immobilization might take place through absorption, entrapment or covalent attachment. To test whether the immobilization of Ald-GFP on CNC-Br occurred *via* covalent attachment (i.e. by displacement of bromide by the deprotonated Cys residues of GFP), we alkylated the nucleophilic thiolates on the surface of GFP to unreactive thioethers. Ald-GFP contains four Cys residues, out of which two are surface exposed. Consequently, Ald-GFP* contains three Cys residues, out of which one is surface exposed. Therefore, when Ald-GFP was reacted with an excess of iodoacetamide (5 mM), alkylation of the two exposed thiol residues was observed (**Figure 4.6**, **a**). Similarly, when Ald-

GFP* was reacted with an excess of iodoacetamide (5 mM), alkylation of the thiol surface exposed residue was detected (**Figure 4.6**, **b**).

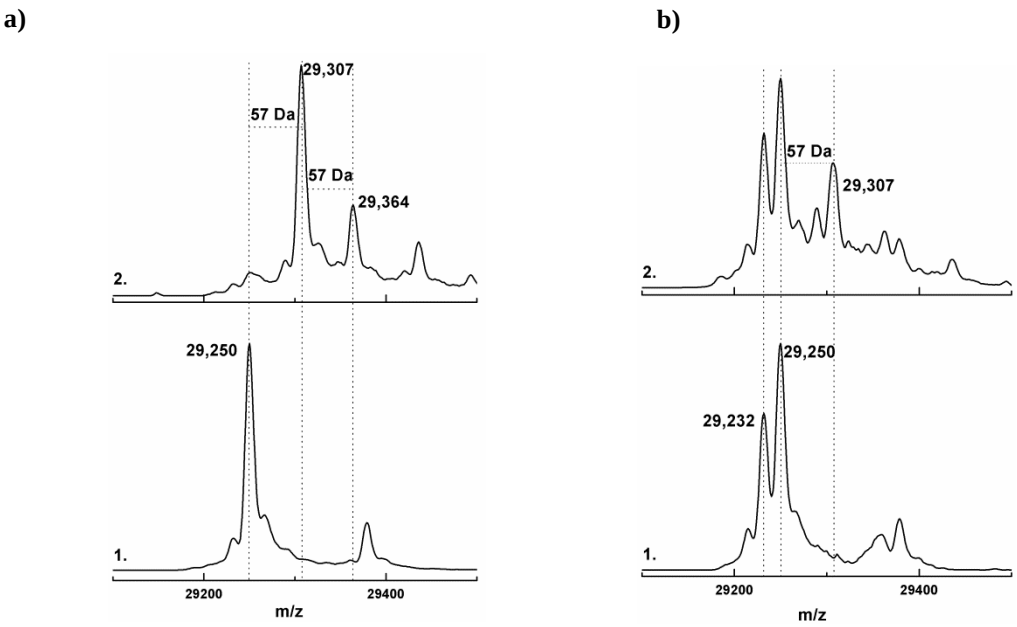


Figure 4.6 Mass spectrometry analyses of Ald-GFP alkylation. **a) 1.** Protein mass spectrometry of Ald-GFP. Ald-GFP displays an average mass of 29,250 Da (calc. 29,250 Da) **2.** Treatment of Ald-GFP with iodoacetamide (5 mM) in the presence of DTT (0.5 mM) modifies all starting protein to new species of average mass of 29,307 Da (calc. 29,307 Da) and 29,364 Da (calc. 29,364 Da, Δ_{mass} = 57 Da). **b) 1.** Protein mass spectrometry of Ald-GFP after 3 h incubation with FGE-4C. Ald-GFP displays an average mass of 29,232 Da and 29,250 Da (calc. 29,232 Da and 29,250 Da) corresponding to the aldehyde- and diol-protein variants found in thermodynamic equilibrium. **2.** Treatment of Ald-GFP with iodoacetamide (5 mM) in the presence of DTT (0.5 mM) alkylates the surface exposed thiols to a new species of average mass of 29,307 Da (calc. 29,307 Da, Δ_{mass} = 57 Da) is observed.

We then mixed the alkylated versions of GFP with CNC-Hydra and CNC-Br particles which prior to conjugation were incubated with BSA (10 mg/ml) for 20 min (**Table 4.3**, entry **1-4**). After overnight reaction, the isolated particles were analysed by fluorescence spectroscopy (**Figure 4.7**). As illustrated by spectra **3 - 4** in **Figure 4.7**, the isolated CNC-Br-GFP$_{alk}$ conjugates still emitted fluorescence intensity when excited at 475 nm. As all the exposed Cys residues in GFP were alkylated prior to incubation with CNC-Br, this demonstrates that the interaction is non-covalent and rather takes place *via* absorption or entrapment.

In support of our selective method for binding proteins, when CNC-Hydra was reacted with the alkylated versions of GFP, selective attachment was observed. That is to say, analysis by fluorescence spectroscopy revealed the highest emission at 504 nm for the hydrazone-linked

complex, and negligible emission for the non-covalent conjugate (**Figure 4.7**, spectrum **1** and **2**). Overall, the abovementioned reaction controls confirm the unique reactivity of the aldehyde functionality in aldehyde-tagged proteins towards selective conjugation to chemically modified nanocellulose.

Table 4.3 Reactions of alkylated Ald-GFP with chemically treated CNCs.

Entry	Sample Name	Chemical Treatment	Protein	Protein Modification
1	CNC-Hydra	**a)** CBr_4, PPh_3 **b)** MeN_2H_4	Ald-GFP_{alk}*	**a)** FGE-4C **b)** Iodoacetamide
2	CNC-Hydra	**a)** CBr_4, PPh_3 **b)** MeN_2H_4	Ald-GFP_{alk}	**a)** Iodoacetamide
3	CNC-Br	**a)** CBr_4, PPh_3	Ald-GFP_{alk}*	**a)** FGE-4C **b)** Iodoacetamide
4	CNC-Br	**a)** CBr_4, PPh_3	Ald-GFP_{alk}	**a)** Iodoacetamide

The proteins highlighted by * were treated with FGE-4C prior to conjugation.

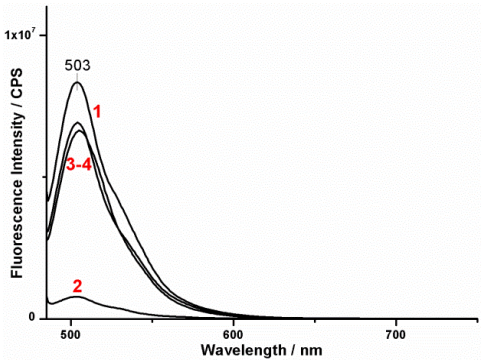


Figure 4.7 Fluorescence emission spectra of aqueous suspensions of cellulose nanocrystals (55.7 µg/ml) which were incubated with alkylated versions of Ald-GFP (**Table 4.3**) using an excitation wavelength of 475 nm.

4. 2. 3 Determining the degree of substitution on CNC by fluorescence spectroscopy

In order to determine the degree of substitution (i.e. the ratio between the number of immobilised proteins and the total number of anhydroglucose units in the cellulose nanocrystals), we obtained the calibration curve of Ald-GFP in the presence of CNC-Hydra (**Figure 4.8**, left, line **1** and **2**).

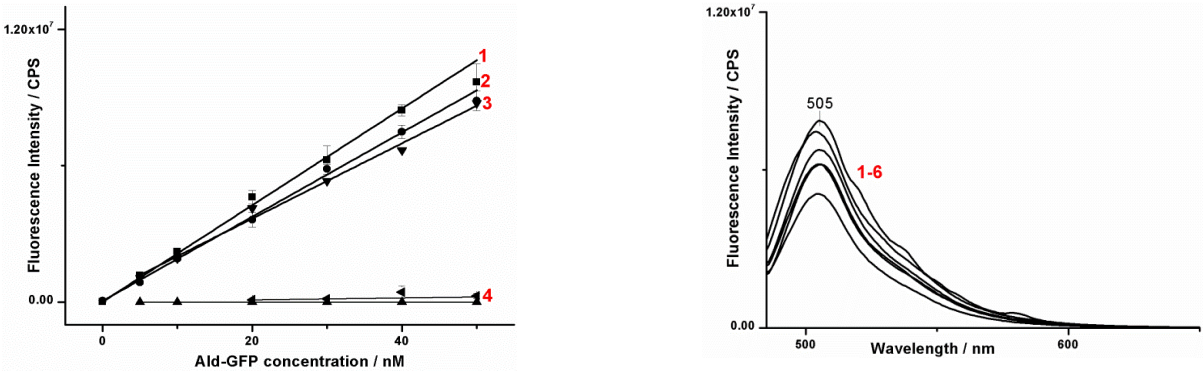


Figure 4.8 Left. Shown are the calibration curves of Ald-GFP alone or in the presence of CNC-Hydra. We plotted the fluorescence intensity versus protein concentration and fitted the data to a linear function to obtain the calibration curve. This calibration curve converts the fluorescence intensity to the concentration of Ald-GFP immobilized on CNC-Hydra. Calibration curve **1**: the exemplary calibration curve of recombinant Ald-GFP in the presence of CNC-Hydra (16.67 µg/ml). Calibration curve **2**: the exemplary calibration of recombinant Ald-GFP in the presence of CNC-Hydra (55.7 µg/ml), FI = 185,349 × [GFP] + 50,064. Calibration curve **3**: the calibration curve of recombinant Ald-GFP in aqueous medium. Calibration curve **4**: the calibration curve of recombinant Ald-GFP in the presence of pristine CNC (55.7 µg/ml). **Right.** Fluorescence emission spectra of Ald-GFP*-CNC-Hydra conjugates (55.7 µg/ml) collected from 6 different batches of CNC-Hydra, using an excitation wavelength of 475 nm.

By using the equation of the calibration curve **2** of Ald-GFP in the presence of CNC-Hydra shown in **Figure 4.8**, we converted the fluorescence intensity measured for the aqueous suspensions of the bio-conjugates (obtained from 6 different CNC-Hydra batches, **Figure 4.8**, right) to the concentration of the immobilised protein. We calculated the degree of substitution of one protein molecule to **$(1.0 \pm 0.2)\times10^4$** anhydroglucose units (AGUs).

4. 2. 4 En route to increasing the degree of substitution

We sought to improve the degree of substitution of Ald-GFP to the surface of CNC-Hydra by performing multiple bromination steps on the pristine material, as shown in **Table 4.4**. Accordingly, we synthesised three different CNC-Br samples (CNC-Br_{01}, CNC-Br_{02} and CNC-Br_{03}) which after isolation were hydrazinated, by the previously described conditions, to afford the corresponding CNC-$Hydra_{01}$, CNC-$Hydra_{02}$ and CNC-$Hydra_{03}$. Based on the measured fluorescence emission signals (**Figure 4.9**, left), the degree of substitution of one protein molecule to **$(0.9 - 1.0)\times10^4$** AGUs was calculated. As by multiple halogenation steps the yield of the modification is unchanged, these results could indicate that the hydroxyl groups on the surface of nanocellulose are topologically restricted. On the other hand, by repeated hydrazination steps on CNC-Br_{01}, we calculated the degree of substitution of one protein molecule to **$(7.3 \pm 0.9)\times10^3$** AGUs (**Figure 4.9**, right).

Table 4.4 Repeated bromination reactions performed on CNC.

Entry	Conversion (bromination)	Conversion (hydrazination)
01	CNC-OH ⟶ CNC-Br_{01}	CNC-Br_{01} ⟶ CNC-$Hydra_{01}$
02	CNC-Br_{01} ⟶ CNC-Br_{02}	CNC-Br_{02} ⟶ CNC-$Hydra_{02}$
03	CNC-Br_{02} ⟶ CNC-Br_{03}	CNC-Br_{03} ⟶ CNC-$Hydra_{03}$

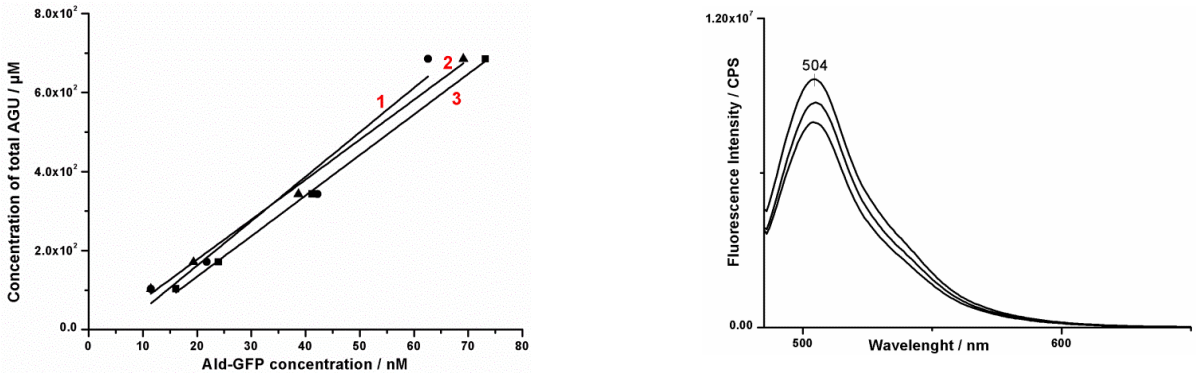
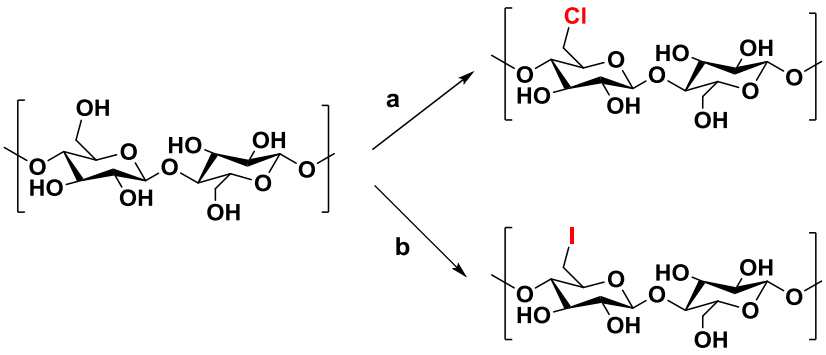


Figure 4.9 Left. Straight lines used to determine the total number of Ald-GFP immobilized on CNC-Br_{01}, CNC-Br_{02} and CNC-Br_{03}. **Right.** Fluorescence emission spectra of Ald-GFP-CNC-Hydra conjugates (resulted from repeated hydrazination on CNC-Br_{01}) (55.7 µg/ml) collected from 3 different experiments.

We investigated whether chlorination or iodination of CNC could provide better conversion yields (**Scheme 4.2**). Eyley and Thielemans employed thionyl chloride for the chlorination of the hydroxyl groups on the surface of cellulose nanocrystals.[36] Following their procedure, we synthesised CNC-Cl. After purification by successive centrifugations with dichloromethane, ethanol and deionised water, CNC-Cl particles were treated with methyl hydrazine under the abovementioned conditions. We then conjugated CNC-Hydra$_{Cl}$ to Ald-GFP*. As illustrated in **Figure 4.10** (left), the fluorescence emission spectrum (**1**) corresponding to CNC-Hydra$_{Cl}$-Ald-GFP* displayed the highest emission at 504 nm. By contrast, CNC-Cl-Ald-GFP* nanoparticles have negligible emission (spectrum **2**, 24-fold times lower than the intensity displayed by the CNC-Hydra$_{Cl}$-Ald-GFP* conjugates) in the region where CNC-Hydra$_{Cl}$-Ald-GFP conjugates emit strongly. The degree of substitution of one protein molecule to **(2.5 ± 1.6)×10^4** AGUs was calculated.



Scheme 4.2 Synthetic route for the halogenation of CNC. **a)** $SOCl_2$, pyr, toluene, 75 °C **b)** PPh_3, I_2, imidazole, THF, reflux.

For the synthesis of CNC-I, the reaction conditions were adapted from those reported for the regioselective iodination of methyl hexopyranosides.[145,146] They involve the use of a combination of triphenylphosphine, iodine and imidazole in refluxing THF. After purification by successive centrifugations with methanol and deionised water, CNC-I particles were treated with methyl hydrazine under the abovementioned conditions. The obtained particles, CNC-Hydra$_I$ were conjugated to Ald-GFP*. As illustrated in **Figure 4.10**, the fluorescence emission spectrum (**1**) corresponding to CNC-Hydra$_I$-Ald-GFP* displayed the highest emission at 504 nm. By contrast, the observed fluorescence emission in spectrum **2**, corresponding to CNC-I-Ald-GFP*, could have resulted from unspecific absorption of Ald-GFP* onto the surface of CNC-I. The degree of substitution of one protein molecule to **(1.8 ± 0.5)×10^4** AGUs was calculated.

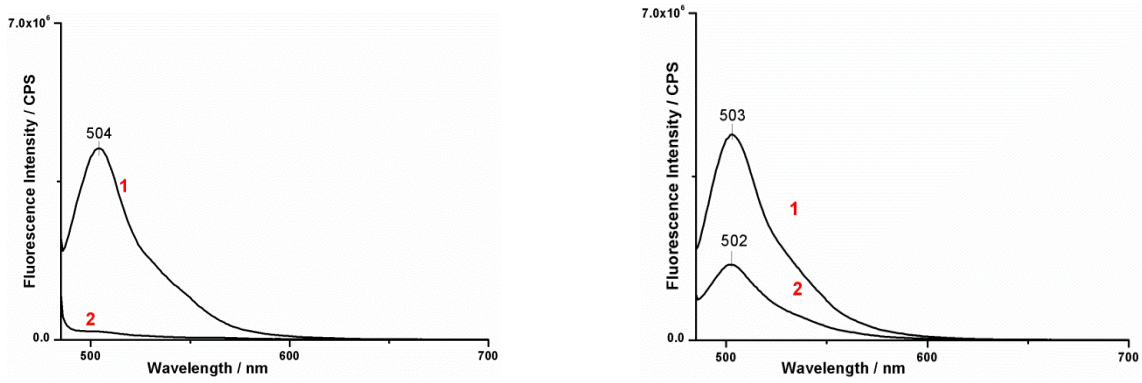
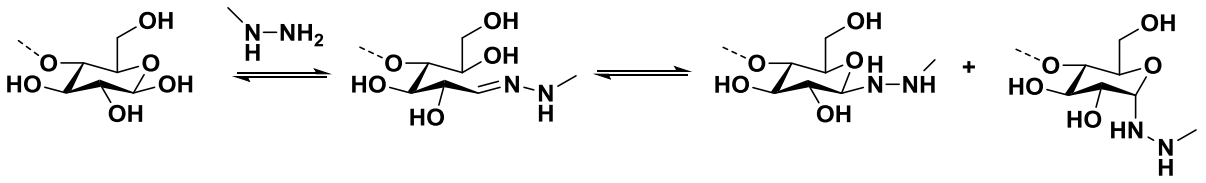


Figure 4.10 Fluorescence emission spectra of diluted suspensions (55.7 µg/ml) of labelled cellulose nanocrystals using an excitation wavelength of 475 nm. **Left.** Emission spectrum **1** corresponds to CNC-Hydra$_{Cl}$-Ald-GFP* conjugates, while emission spectrum **2** corresponds to CNC-Cl-Ald-GFP*. **Right.** Emission spectrum **1** corresponds to CNC-Hydra$_{I}$-Ald-GFP* conjugates, while emission spectrum **2** corresponds to CNC-I-Ald-GFP*.

4. 2. 5 Immobilization at the reducing-ends of CNC

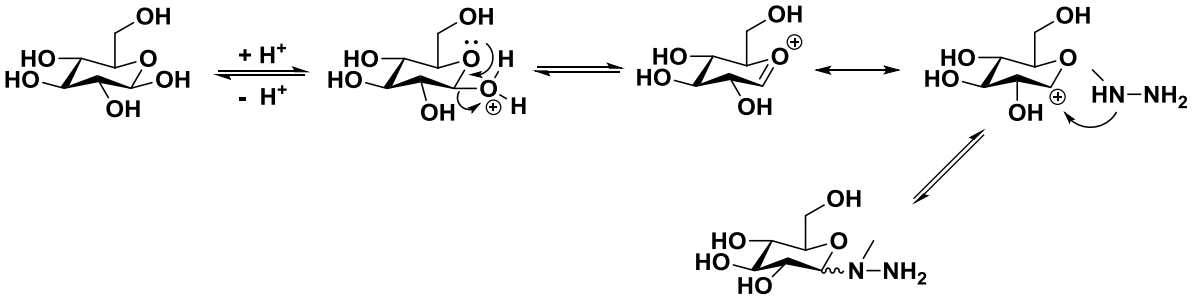
At the reducing ends of the nanocellulose chain, glucose exists primarily in the pyranose form.[131] However, depending on the position of the equilibrium small amounts of material are present in the open chain aldehyde form. When CNC is treated with methylhydrazine, the aldehyde functionality at the reducing ends reacts *via* hydrazone conjugation as illustrated in **Scheme 4.3**. Owing to the thermodynamic stability of the 6-membered rings, the open-chain hydrazone form may then re-cyclise to form a hydrazine.



Scheme 4.3 Hydrazine functionalisation at the reducing ends of CNC by treatment with methyl hydrazine.

Alternatively, the functionalization with hydrazines at the reducing ends of CNC could take place *via* a S_N1 pathway. The reaction initially proceeds by formation of a cyclic carbocation

intermediate known as the glycosyl cation[131]. The formed glycosyl cation is then attacked by methyl hydrazine to yield the hydrazine product.



Scheme 4.4 Hydrazine functionalization at the reducing ends of CNC by S_N1.

Therefore, the conjugation of Ald-GFP* with CNC-Hydra could also take place at the reducing ends by an enamine like addition (**Figure 4.11**, **a**) or hydrazone ligation (**Figure 4.11**, **b**), as described below.

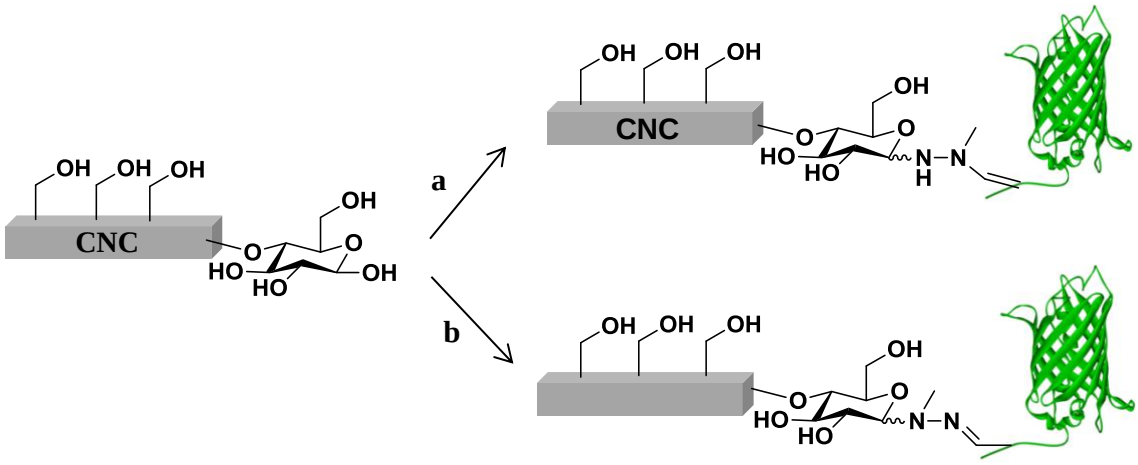


Figure 4.11 Representation of the covalent interactions of Ald-GFP* with CNC_H at the reducing ends.

In order to determine the degree of substitution at the reducing ends, we treated CNC with methyl hydrazine. The isolated nanoparticles, CNC_H were conjugated to Ald-GFP* and Ald-GFP. As shown in **Figure 4.12**, the fluorescence of the obtained conjugates was measured. The degree of substitution of one protein molecule to **4.8×10^5** AGUs was calculated.

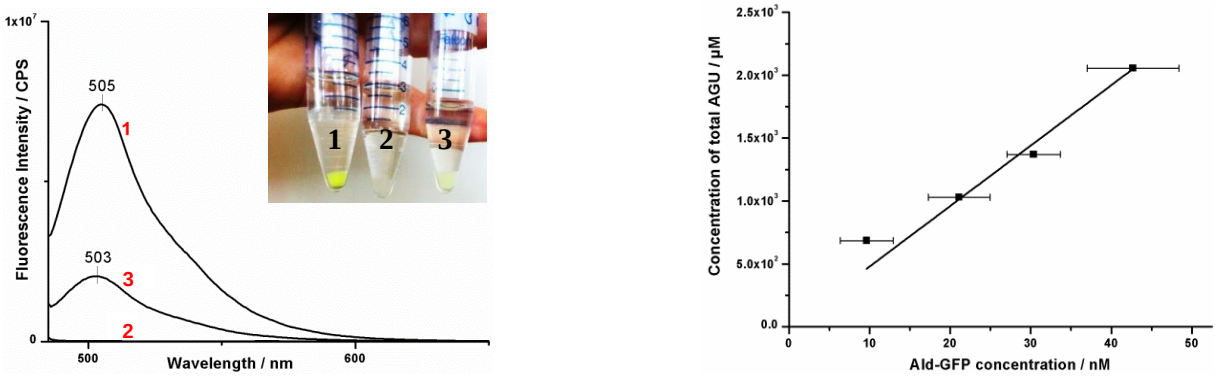
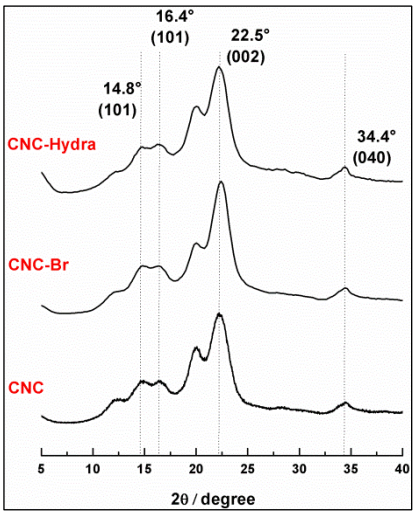


Figure 4.12 Bio-conjugates of Ald-GFP with crystalline nanocellulose. **Left**. Fluorescence emission spectra of aqueous suspensions of labelled cellulose nanocrystals (55.7 µg/ml) using an excitation wavelength of 475 nm. Image of sedimented protein-nanoparticles conjugates: **1** – CNC-Hydra-Ald-GFP*, **2** – CNC_H-Ald-GFP, **3** – CNC_H-Ald-GFP*. **Right**. Calibration curve used to determine the total number of Ald-GFP immobilized at the reducing ends of CNC_H. Error bars denote standard deviation.

4. 3 Discussion

4. 3. 1 Morphological and crystalline integrity of CNC-Br and CNC-Hydra

When performing chemical modification on cellulose nanocrystals of paramount importance are the maintenance of the original morphology (i.e. rod-like structure, nanoscale dimensions) and crystalline integrity.[147] The X-ray diffraction patterns of CNC-Br and CNC-Hydra reflect the crystalline integrity of the nanocrystals after the surface modification. As shown in **Figure 4.13**, all the crystalline characteristic peaks of the pristine material (CNC) are clearly present, which indicates the preservation of the original crystalline structure of cellulose in CNC-Br and CNC-Hydra.[148] The diffraction peaks at 2θ angles at 14.8°, 16.4°, 22.5° and 34.4° were assigned to the typical reflection planes of cellulose.[149] The crystallinity index (I_c) of CNC was calculated by Segal equation[150] $I_c = (I_{200}-I_{am})/I_{200}\times 100\%$, where I_{200} is the intensity of the peak associated with the crystalline region of cellulose ($2\theta = 22.6^{\circ}$) and I_{am} is the intensity of the baseline.[150-153]



Segal equation

$$I_c = \frac{I_{200} - I_{am}}{I_{200}} \times 100$$

Sample	Crystallinity Index, I_c (%)
CNC	71.5
CNC-Br	72.7
CNC-Hydra	65.8

Figure 4.14 Left. X-ray diffraction patterns for CNC, CNC-Br and CNC-Hydra. **Right.** Segal equation[150] was used to calculate the crystallinity index (I_c). The obtained I_c values are displayed in the table.

The morphological integrity of cellulose nanocrystals after bromination and subsequent modification with methyl hydrazine was demonstrated by transmission electron microscopy[154-160]. As shown in **Figure 4.14**, brominated and hydrazine-functionalised cellulose nanocrystals (CNC-Br and CNC-Hydra) display a rod-like morphology, with a length of 200 – 300 nm and width of

10 - 20 nm.[29] TEM and X-ray confirm that the chemical modification took place at the surface of the nanocellulose, without affecting the core crystallinity of the samples. The X-ray and TEM measurements were performed by Dr. Philippe Tingaut.

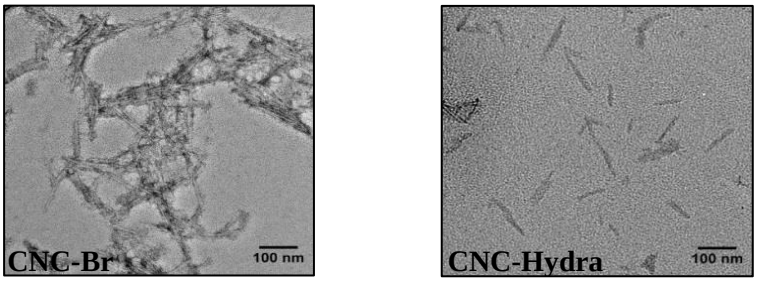


Figure 4.15 Transmission electron micrographs of CNC-Br and CNC-Hydra.

4. 3. 2 Degree of immobilization

The halogenation reactions under Appel (i.e. bromination) and mild iodination conditions are highly selective towards the conversion of the primary alcohols, in the presence of the secondary hydroxyl groups on the surface of the polymer. The TEMPO-oxidation is similarly selective towards the conversion of the primary alcohols into carboxylic acids.[13,42] However, under all the aforementioned conditions, only half of the accessible hydroxymethyl groups are available to react.[13] The other half of the primary alcohols is topologically confined or buried within the crystalline particle. This inhered topological barrier could cause difficulty when one aims to utilise the high surface area of CNC of several hundred[161] $m^2 g^{-1}$ to immobilize molecules with high-loading values. The reason beyond this reasoning is the *ab initio* low degree of substitution (DS) that has a maximum value of 0.5.

On the other hand, a high degree of substitution could disrupt the individual fibre and lead to loss of CNCs mechanical properties.[162] In addition, the number of the reported examples of damaged crystallinity of CNC (by measuring the decrease in the diameter of the rod by AFM) resulted from harsh chemical treatment[35,149] or 'peeling-off effect'[161] are by far not negligible. When one aims to maintain the intrinsic properties of CNC, the aforementioned restrictions make the chemical functionalization challenging. As a consequence, another approach was introduced by Brumer and co-workers who decorated nanocellulose with chemoenzymatically modified hemicellulose.[162]

Reaction of cellulose nanocrystals with excess thionyl chloride converts both primary and secondary hydroxyl groups into alkyl chlorides. However, it may be that substitution of chloride with methyl hydrazine at the secondary hydroxyl group is hindered.[36] As a result, only the primary hydroxyl groups on the surface of cellulose crystals can be converted into hydrazine groups. Considering that the aforementioned reactions went all through hydrazination at C_6 atom of the glucose monomers, the obtained degree of substitution values of one protein molecule to (**1.0 ± 0.2**; **1.8 ± 0.5**; **2.5 ± 1.6**)**×10^4** to AGUs following bromination, iodination and chlorination are in accord.

When recombinant proteins were covalently attached on CNC by the sortase mediated reaction (See **Chapter 1**), the amount of the selectively immobilized proteins was not specified.[42] The degree of substitution of one oxime-ligated peptide linker to **1.25×10^2** AGUs was determined by elemental analysis. Regarding the chlorination of CNC, the authors do not specify the degree of substitution following halogenation (See **Chapter 1**).[36] They do report the amount of the absorbed dye to the chemically introduced imidazolium groups (of one dye molecule to **2.0×10** AGUs) as determined by elemental analysis.[36] Recently, Zimmerman and co-workers reported the covalent immobilization on NFC of cytochrome c and papain by the EDC/NHS method.[163] The degree of substitution of one protein molecule to **1.25×10^2** AGUs was determined. Nevertheless, nanofibrillated cellulose are entangled fibres with higher surface area than CNC[163] and the EDC/NHS conjugation is non-specific.

4. 4 Experimental

4.1. Synthesis of 6-bromodeoxycellulose nanocrystals

Cellulose nanocrystals (CNC) (1.00 g, 6.17 mmol, 1.0 **eq.**) were dispersed in anhydrous pyridine (100 ml) and the obtained suspension was sonicated for 20 min until homogeneous. To the suspension at 0 °C under N_2 were added triphenylphosphine (6.47 g, 24.68 mmol, 4.0 **eq.**) and carbon tetrabromide (6.14 g, 18.51 mmol, 3.0 **eq.**). The reaction mixture was heated to 75 °C, stirred for 5 h and quenched at 0 °C by addition of methanol (10 ml). The modified nanocellulose was isolated by centrifugation at 2000 rpm and 4 °C for 30 min. The particles were re-suspended in methanol, sonicated and/or vortexed, and then collected by centrifugation at 2000 rpm and 4 °C for 30 min (×5). The collected particles were dispersed in water and freeze dried (0.8 g were recuperated).

4.2. Synthesis of 6-iododeoxycellulose nanocrystals

Cellulose nanocrystals (CNC) (0.50 g, 3.09 mmol, 1.0 **eq.**) were dispersed in anhydrous THF (50 ml) and the obtained suspension was sonicated for 20 min until homogeneous. To the suspension at 0 °C under N_2 were added triphenylphosphine (3.23 g, 12.36 mmol, 4.0 **eq.**) and imidazole (1.11 g, 16.38 mmol, 5.3 **eq.**). The reaction mixture was refluxed for 30 min before addition of a solution of I_2 (3.14 g, 12.36 mmol, 4.0 **eq.**) in THF (5 ml). The reaction was heated and refluxed for 3 h. The modified nanocellulose was isolated by centrifugation at 2000 rpm and 4 °C for 30 min. The particles were re-suspended in methanol, sonicated and/or vortexed, and then collected by centrifugation at 2000 rpm and 4 °C for 30 min (×5). The collected particles were dispersed in water and freeze dried. The obtained CNC-I particles were converted into CNC-Hydra under the conditions described below.

4.3. Synthesis of 6-chlorodeoxycellulose nanocrystals

Cellulose nanocrystals (CNC) (0.25 g, 1.55 mmol, 1.0 **eq.**) were suspended in a solution of pyridine (3 ml) and dry toluene (50 ml) under argon. A solution of thionyl chloride (5 ml) was added dropwise before subsequent heating at 65 °C for 16 hours. The modified nanocellulose was collected by filtration. The particles were re-suspended in dichloromethane, sonicated and/or vortexed, and then collected by centrifugation at 2000 rpm and 4 °C for 30 min (×10). The collected particles were dispersed in water and freeze dried. The obtained CNC-Cl particles were converted into CNC-Hydra under the conditions described below.

4.4. Synthesis of 6-methylhydrazynilcellulose nanocrystals

Surface-halogenated cellulose nanocrystals (CNC-X, X – Cl, Br, I) (0.10 g, 0.62 mmol) were sonicated in anhydrous methanol (8 ml) and to the suspension was added methyl hydrazine (3 × 2 ml) over a period of 30 h at 46 °C. The excess of methyl hydrazine was removed by filtration and the modified nanocellulose (CNC-Hydra) was isolated by centrifugation at 2000 rpm and 4 °C for 30 min. The particles were re-suspended in methanol, sonicated and/or vortexed, and then collected by centrifugation at 2000 rpm and 4 °C for 30 min (×5). The collected particles were dispersed in water and freeze dried.

4.5. Synthesis of 6-phtalimide nanocrystals

CNC-Br (0.5 g, 3.1 mmol, **1.0** eq.) was dispersed in ACN (50 ml). To the suspension were added *N*-hydroxyphtalimide (0.76 g, 4.65 mmol, **1.5** eq.), *cat.* potassium iodide (0.103 g, 0.62 mmol, **0.2** eq.) and triethylamine (6 ml) (×2). The collected particles were filtered and then collected by centrifugation, after repeated washing steps.

4.6. Synthesis of 6-aminooxy nanocrystals

CNC-aminooxy (0.2 g, 1.2 mmol, **1.0** eq.) was dispersed in dry MeOH (20 ml). To the suspension was added hydrazine hydrate. The collected particles were filtered and then collected by centrifugation, after repeated washing steps.

Chapter 5 - Immobilization of SLAC, FGE and a human antibody on crystalline nanocellulose

5. 1 Introduction

Immobilization of biocatalysts on solid supports entails advantages such as simplified purification methods coupled with the re-use of the enzymes.[164] Taking into account our selective method for linking proteins to nanocellulose, we investigated the controlled orientation of a laccase, human IgG antibody and FGE-4C on our previously described hydrazine-functionalised cellulose nanocrystals.

5. 1. 1 Small laccase from *Streptomyces coelicolor*

Laccases are the largest subgroup of the protein superfamily of multi-copper oxidases (MCOs)[165] and they were isolated from plants (*Rhus vernicifera*[166]), fungi (*Rhizoctonia solani*[167], *Neurospora crasa*[168], *Xylaria hypoxylone*[169], *Trametes versicolor*[170], *etc*) and bacteria (*Azospirillum lipoferum*[171], *Bacillus subtilis*[172], *Streptomyces griseus*[173] and *Bacillus licheniformis*[174]). Since their discovery in 1883 by Yoshida, in the Japanese lacquer tree *Rhus vernicifera*, laccases have been the centre of multiple research studies.[175] It turned out that these enzymes have a large substrate spectrum ranging from phenols, polyphenols to heterocyclic compounds.[175] The oxidation of the substrate is accompanied by the four-electron reduction of molecular oxygen to water. Although fungal laccases have higher redox potential (up to 800 mV)[176] than bacterial or plant laccases, they are difficult to be produced in bacteria.[165] Laccases consist of multiple domains. While the first and third domains contain the copper sites, the second domain plays a role in substrate-binding.[177]

In 2004, Canters and co-workers identified in the genome of *Streptomyces coelicolor* the gene encoding for a four-copper oxidase which lacks the second domain.[177] The corresponding protein, called SLAC or small laccase (**Figure 5.1**, left), was recombinantly produced in *E. coli* and its mechanism was later described. As shown in the crystal structure of SLAC, the enzyme contains four copper atoms, arranged in 3 types of centres: type 1 (T1 Cu), type 2 (T2 Cu) and type 3

(T3 Cu). While at the T1 Cu centre substrate oxidation occurs, at the T2 Cu and T3 Cu centres or TNC (trinuclear Cu cluster) oxygen binds and gets reduced to water.

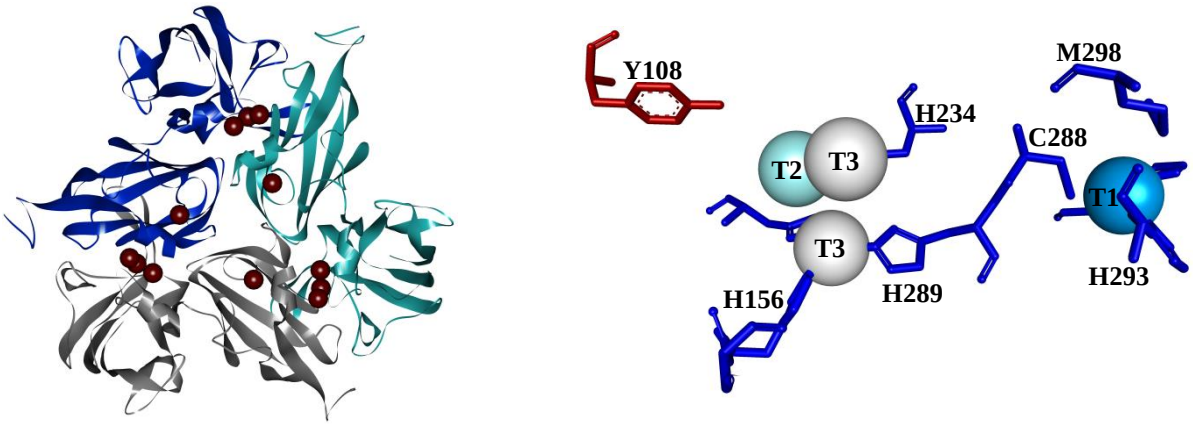
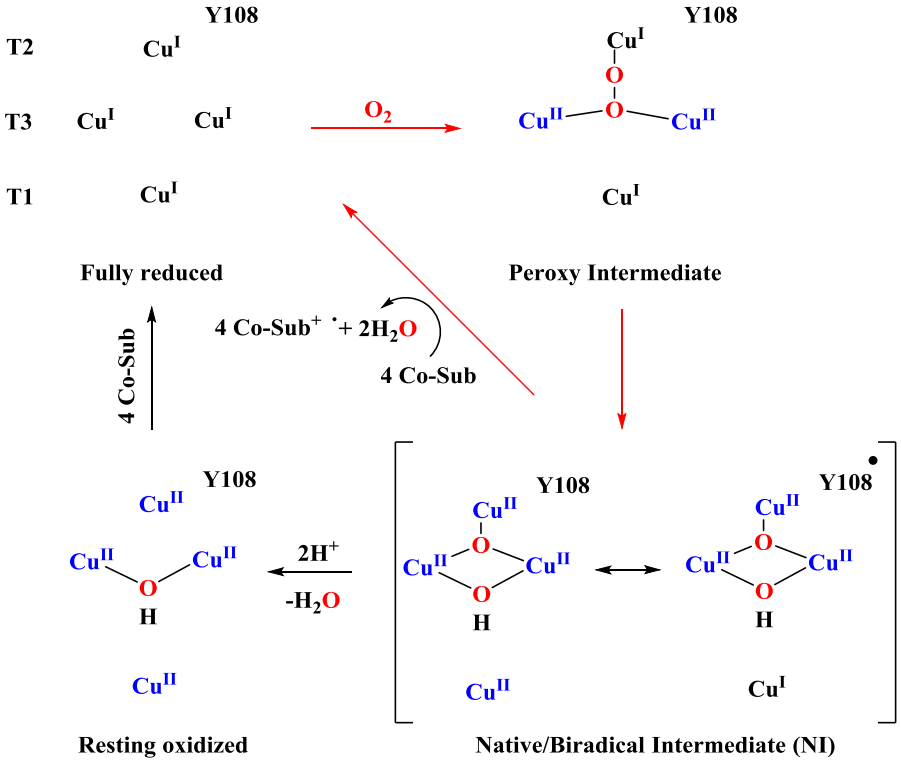


Figure 5.1 Left. Ribbon representation of the active homotrimer of the two-domain SLAC (pdb: 3CG8[178]). The copper centres are shown in brown. **Right.** A zoom into the residues near the T1 Cu (blue), T2 Cu (light blue) and T3 Cu (grey) centres.

Scheme 5.1 shows the proposed mechanism by which SLAC binds and reduces oxygen while oxidizing a co-substrate (*N,N,N′N′*-tetramethyl-*p*-phenylenediamine, TMPD).[179,180] The fully reduced state of the enzyme has four reduced copper centres (Cu^{1+}). Upon binding at the TNC centre, oxygen is irreversibly reduced and this results in the formation of a peroxy intermediate (PI). The authors proposed that the next step of the catalytic acid goes through a native/biradical intermediate. In SLAC, as indicated in the structure of the (NI), a tyrosine residue (Y108) found in close proximity (5 Å) to the T2 Cu centre was suggested to take part in the catalytic cycle.[180] By providing an electron to the T2 Cu centre, Y108 acts as a redox-active centre and favours the formation of the bi-radical intermediate. Then, the catalytic cycle could continue with reduction of the resting enzyme or the oxidised intermediate (NI).[181] The resting oxidised state contains a hydroxide bridged T3 Cu centre. The cycle is completed with donation of $4e^-$ by a co-substrate (TMPD) which reduces all Cu^{2+} to Cu^{1+}, thereby regenerating SLAC.



Scheme 5.1[180] Proposed mechanism for the reduction of oxygen to water by SLAC, where the role of Y108 is highlighted. Oxidised coppers are shown in blue, while the used co-substrate is *N,N,N′N′*-tetramethyl-*p*-phenylenediamine (TMPD).

5. 1. 2 Protein A from *Staphylococcus aureus*

Protein A is a surface protein that is covalently attached to the surface of the Gram-positive bacterium *Staphylococcus aureus*.[182] Based on the structural motifs that constitute protein A (i.e. *N*-terminal signalling region and *C*-terminal sorting motif), it was classified as a cell-wall anchored protein and belongs to the three-helical bundle family. Protein A contains five identical triple-helix domains (E-D-A-B-C) each able to bind with high affinity both Fc (the constant region of IgG involved in effector functions, **Figure 5.2**, left) and Fab (the Ig fragment responsible for antigen recognition, **Figure 5.2**, right) fragments of immunoglobulins IgG and IgM.[183] By binding to the Fc domain of human immunoglobulins, protein A blocks phagocytosis of *S. aureus* by human neutrophils.[184] By binding to the Fab domain of IgM, protein A triggers cross-linking of B cell receptors resulting in their apoptosis.[185] Moreover, protein A binds and

activates tumor necrosis factor receptor 1 in host airway epithelial cells causing inflammatory responses.[187]

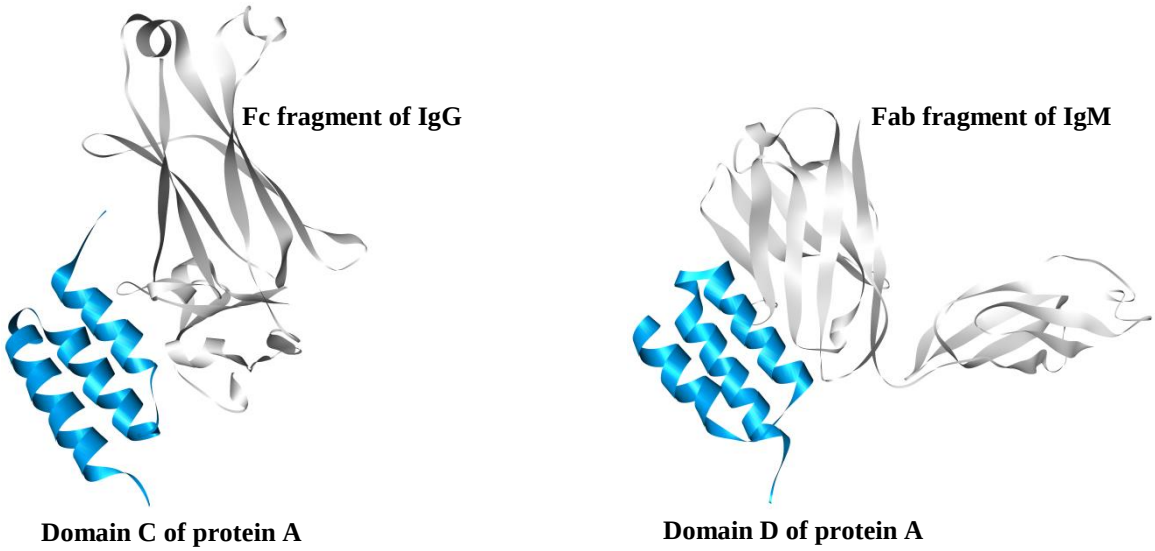


Figure 5.2 Left. Crystal structure of the fragment of the Fc region of a human IgG antibody, complexed with the domain C of protein A *S. aureus* (pdb: 4WWI[186]). **Right.** Crystal structure of the domain D of protein A *S. aureus* complexed with the Fab fragment of a human IgM antibody (pdb: 1DEE[189]). The Fc/Fab regions are shown in grey and domain C/D in blue.

Owing to its intrinsic antibody-binding affinity, protein A *S. Aureus* has been extensively used as a tool for the immobilization of antibodies. For instance, Ishihara and co-workers attached antibodies by means of protein A on amino-functionalised phospholipid polymer platform prepared on silicon substrates.[189] To investigate the role of protein A for the oriented attachment of antibodies, they applied three methods of immobilization (**Figure 5.3**). Firstly, antibodies were randomly captured on the surface of the polymer platform by adsorption. Secondly, the attachment of immunoglobulins was partially oriented *via* protein A which was absorbed onto the substrate. Thirdly, the orientation of antibodies was controlled by means of site-selectively immobilised protein A on the amino-functionalised support. As oriented antibodies displayed 100 - fold stronger affinities for antigens in comparison to randomly or partially oriented antibodies, the authors demonstrated that the site-selective immobilization of protein A defines the antibody orientation.[189] Consequently, their results highlight the importance of antibody orientation on solid supports for the development of immunoassay with high sensitivity.

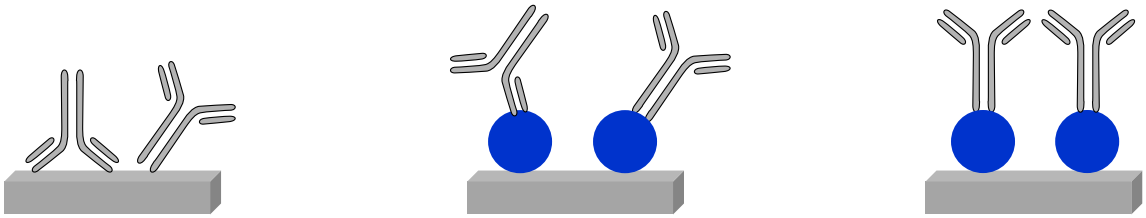


Figure 5.3[189] Models for the immobilisation of antibodies. **Left.** Randomly oriented IgG. **Centre.** Partially oriented IgG on adsorbed protein A. **Right.** Oriented immobilisation of IgG on site-selectively attached protein A.

Recently, Poulter and co-workers reported a regioselective strategy for covalently immobilising antibody-binding protein A, G and L on the surface of modified glass.[190] Their described strategy employs the use of protein farnesyltransferase (PFTase). PFTase catalyses the formation of a thioether bond between proteins (i.e. protein A, G and L) bearing the PFTase's recognition motif, CVIA, and farnesyl diphosphate (i.e. alkyne functionalised analogue).[191] The alkyne-containing protein A was regioselectively attached to azido-modified glass by the Cu^{+}-catalysed Huisgen cycloaddition. Protein A decorated glass served then as a support for the selective immobilization of fluorescently labelled antibodies.[190]

5. 2 Results

5. 2. 1 Immobilization of Ald-SLAC on CNC-Hydra

By cloning the gene of SLAC[177] into our general pET expression plasmid (ald-gfp), we obtained the aldehyde-tagged, Ald-SLAC. Ald-SLAC was overproduced in *E. coli* affording *ca.* 15 mg per one litre culture in high purity as indicated by SDS-PAGE analysis (Experimental Section). As shown below, incubation of Ald-SLAC with catalytic amounts of FGE-4C, followed by the labelling with the aminooxy reagent resulted in site-specific covalent modification (**Figure 5.4**).

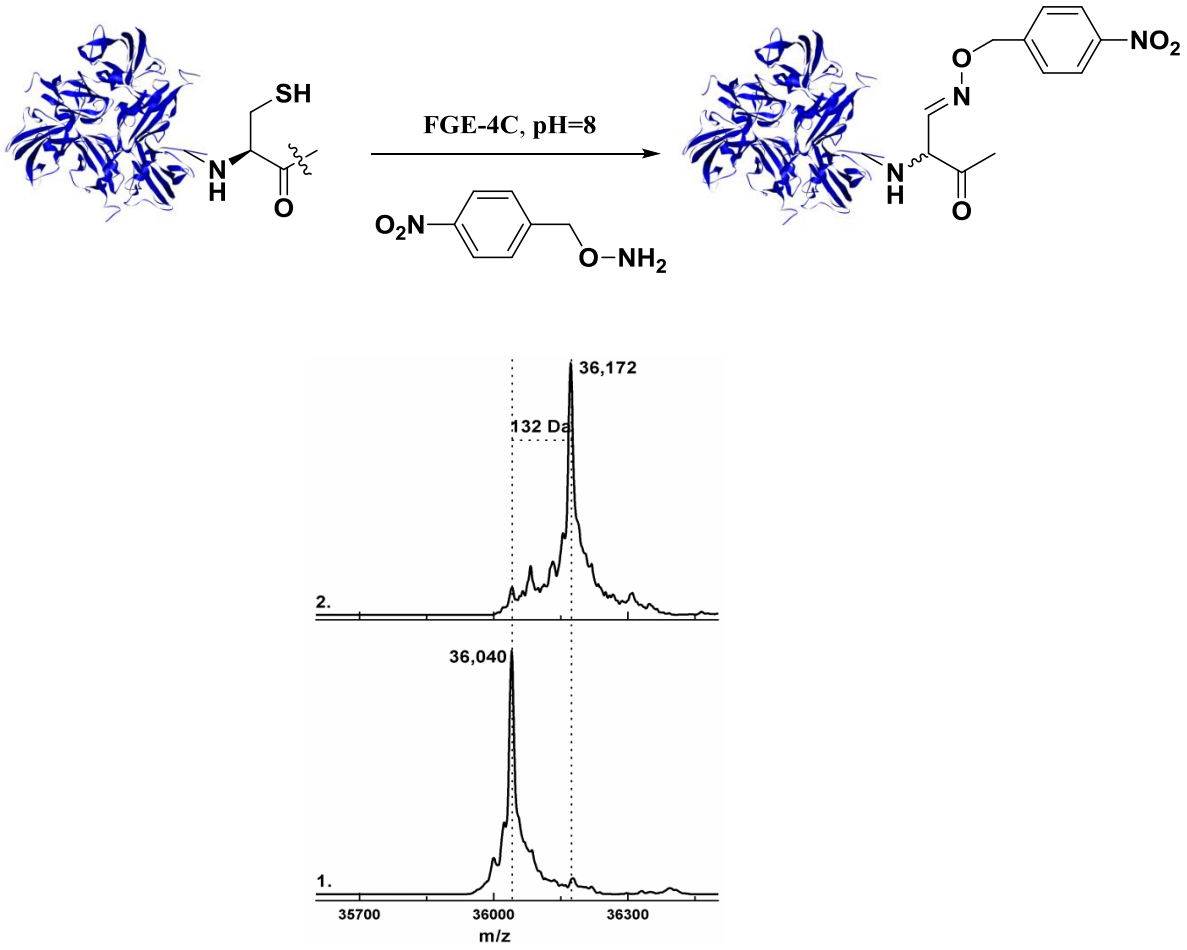


Figure 5.4 Mass spectrometry analyses of Ald-SLAC covalent modification with *O*-(4-nitrobenzyl)hydroxylamine. **1.** Protein mass spectrometry of Ald-SLAC after 10 sec incubation with FGE-4C. Ald-SLAC displays an average mass of 36,040 Da (calc. 35,999 Da) **2.** Treatment of Ald-SLAC with FGE-4C (i.e. 3 h) modifies all starting protein to new species of average mass 36,172 Da (calc. 36,131 Da). The observed Δ_{mass} = 41 Da between the calculated and observed mass could presumably arise from the carbamylation of a lysine residue or acetonitrile adduct.

We wanted to demonstrate that the surface of crystalline nanocellulose can serve as a molecularly defined tool to attach bioactive molecules. Towards this goal, we decorated the hydrazine functionalised nanocellulose, CNC-Hydra with Ald-SLAC, by the previously described protocol (**Chapter 4**). Subsequently, the activity of Ald-SLAC (in solution and suspension) was tested on a phenolic compound with a low redox potential (0.58 vs 0.43 V), 2,6-dimethoxyphenol (DMP)[192]. Accordingly, the kinetic parameters for the oxidation of DMP were calculated for the laccase when in solution or after immobilization on nanocellulose, by measuring the increase in absorbance at 468 nm (**Table 5.1**). The obtained values for the enzyme efficiency were compared to the ones reported by Toscano and co-workers (i.e. 7.35×10^2 M^{-1} s^{-1} versus 3.05×10^2 M^{-1} s^{-1}).[192] In order to determine the value of k_{cat}/K_m corresponding to the immobilised laccase, we considered the calculated degree of substitution for Ald-GFP, of one protein in 10^4 AGUs. The determined catalytic efficiency values are similar after immobilization, suggesting that the bioactivity of the laccase was preserved.

Table 5.1 – Kinetic parameters of immobilised Ald-SLAC on CNC-Hydra with DMP.

Enzyme	**k_{cat} [s^{-1}]**	**K_m [µM]**	**k_{cat}/K_m [M^{-1} s^{-1}]**
Ald-$SLAC_{sol}$	1.10	1333	8.25×10^2; 7.35×10^2
Ald-$SLAC_{imm}$	-	-	8.88×10^2
Ald-$SLAC_{imm}^{*}$	-	-	8.43×10^2
Ald-$SLAC_{re\text{-}use}$	-	-	4.71×10^2

Assay conditions: 30 °C, 50 mM BR buffer, 50 mM NaCl, pH 8.0. The oxidation of DMP was followed by measuring the increase in absorbance at 468 nm ($\Delta\varepsilon_{468nm}$ = 14800 M^{-1} cm^{-1}). The kinetic parameters (k_{cat} and K_m) were calculated based on the initial rates of the reaction at different substrate concentrations and the measurement was performed as triplicates. The obtained values are similar to the one reported by Toscano *et al.* (i.e. k_{cat} 0.87 s^{-1}, K_m 2850).[198] The k_{cat}/K_m values for Ald-$SLAC_{sol}$ were determined by hyperbola fitting and first points linear fitting. The k_{cat}/K_m values for the immobilised laccase were determined by first points linear fitting. Ald-$SLAC_{imm}^{*}$ was immobilised on an identically prepared CNC-Hydra batch. The catalytic efficiency for Ald-$SLAC_{re\text{-}use}$ was measured after 6 days of storage at -80 °C.

5. 2. 2 Immobilization of Ald-protein A on CNC-Hydra

By cloning the five *N*-terminal linked homologous domains of protein A into our general pET expression plasmid (ald-gfp), we obtained an aldehyde-tagged truncated version, Ald-protein A. The truncated variant of the 55 kDa wt protein A *S. aureus*, displays a mass of 36 kDa and lacks the polymorphic Xr and Xc regions (that have potential roles in pathogenesis)[193] and the sortase cleavage motif LPETG. Ald-protein A was overproduced in *E. coli* affording *ca.* 14 mg per one litre culture in high purity as indicated by SDS-PAGE analysis (Experimental Section). As shown below, incubation of Ald-protein A with catalytic amounts of FGE-4C, followed by the labelling with the aminooxy reagent resulted in site-specific covalent modification.

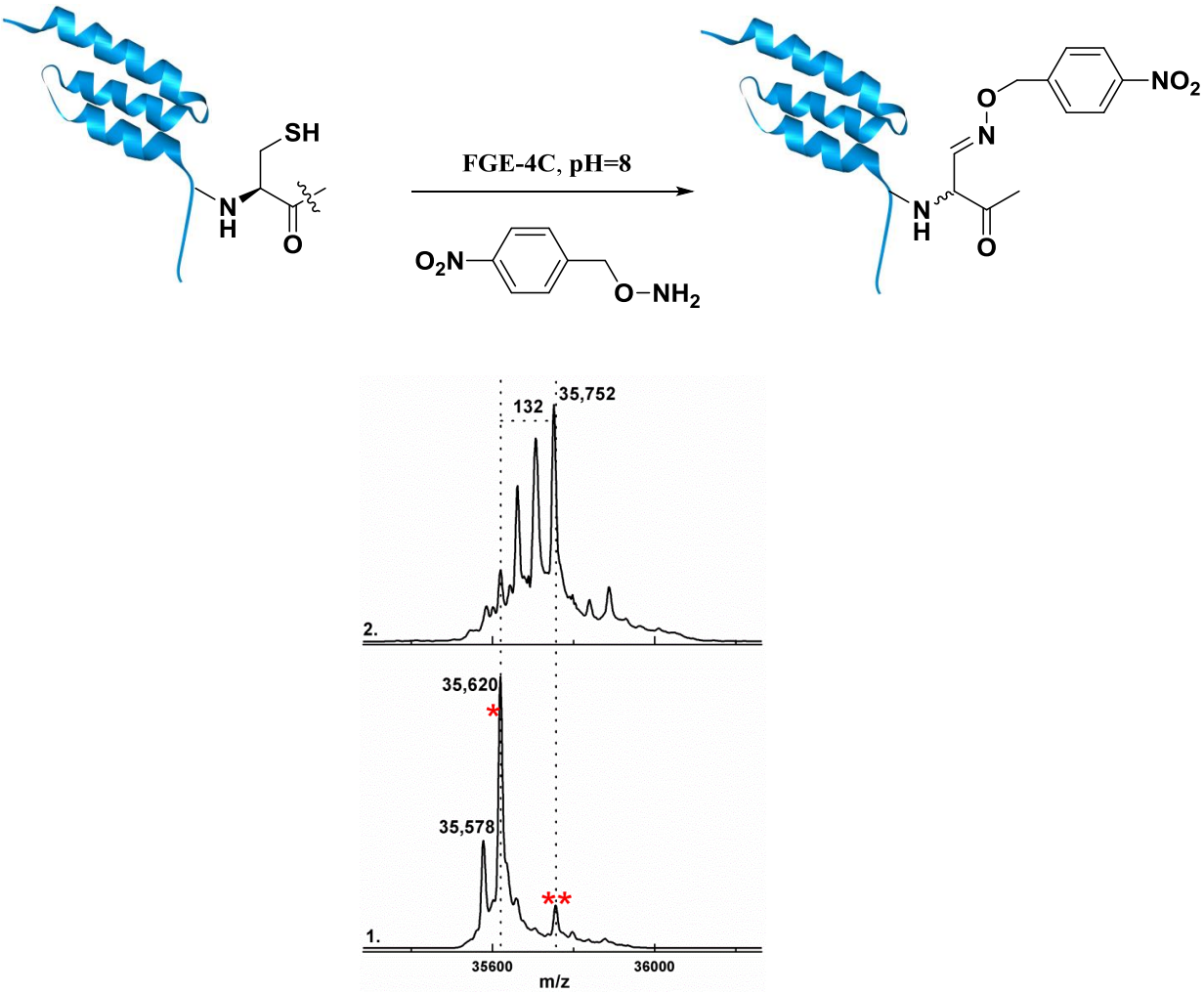


Figure 5.5 Mass spectrometry analyses of Ald-protein A covalent modification with *O*-(4-nitrobenzyl)hydroxylamine. **1.** Protein mass spectrometry of Ald-protein A after 10 sec incubation with FGE-4C. Ald-protein A displays an average mass of 35,578 Da and 35,620 Da (calc. 35,579 Da) **2.** Treatment of Ald-protein A with FGE-4C (i.e. 3 h) modifies the starting protein to new species of average mass 35,707 Da and

35,752 Da (calc. 35,710 Da). Peaks highlighted by * (Δ_{mass} = 42 Da) could probably arise from the carbamylation of a lysine residue. Peaks highlighted by ** (Δ_{mass} = 178 Da) could arise from gluconoylation.

Following the strategy shown in **Figure 5.6**, we then explored the highly-oriented immobilization of human antibody IgG onto cellulose nanocrystals. For this purpose, we decorated the surface of hydrazine functionalised nanocellulose, CNC-Hydra with the aldehyde-containing protein A, by the previously described protocol (**Chapter 4**).

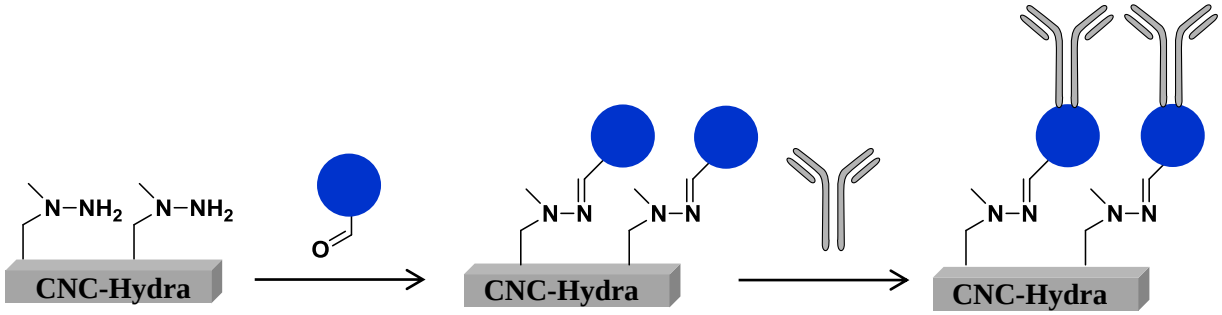


Figure 5.6 Schematic representation describing the immobilization of human IgG onto nanocellulose. Hydrazine functionalised cellulose nanocrystals were decorated with aldehyde-containing protein A. Then, the assembled bio-conjugates were used to selectively capture human IgG antibody from a BSA concentrated solution.

We tested the ability of our nanoparticle-protein sensor to detect antibodies, by dispersing them in a solution (20 ml) containing human antibody IgG (0.1 mg) and BSA (100 mg) in 50 mM sodium phosphate buffer at pH 7.0. As shown in **Figure 5.7**, transfer of IgG to the CNC-Hydra-Ald-protein A was observed qualitatively by gel electrophoresis (lane **2**, **4** and **6**). In support of our method for selectively immobilizing antibodies on the surface of cellulose nanocrystals, BSA was included as a non-binding internal control. The amount of IgG was kept constant, whereas the concentration was decreased in a buffer containing excess BSA to show the capability of protein A modified CNC to selectively capture the antibody. Analysis by SDS-PAGE, revealed selective capturing of IgG by CNC-Hydra-Ald-protein A, but not of BSA, even in the presence of 1000 - fold excess of the protein (lane **2**). Moreover, as the same amount of nanoparticles was loaded on lane **2**, **4** and **6**, this experiment demonstrates the sensitivity of the probe (i.e. the amount of immobilised IgG is similar irrespective of the concentration of IgG).

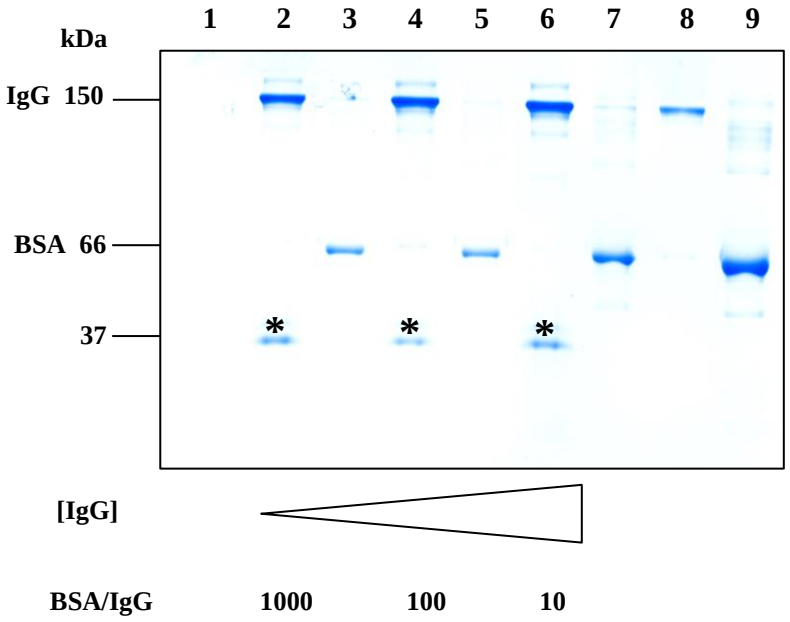


Figure 5.7 Image of SDS-PAGE showing the selective immobilization of human IgG on protein A decorated nanocellulose. As shown in lane **1**, no protein bands are visible corresponding to pristine CNC-Ald-protein A conjugates which were treated with IgG in presence of 10 - fold BSA. Lane **2** shows the band corresponding to the immobilized IgG by CNC-Hydra-Ald-proteinA conjugate (1.6 mg) from 20 ml reaction volume containing IgG (5 µg/ml) in the presence of 1000 - fold BSA. Lane **3**, **5**, **7** display the band corresponding to a BSA/IgG mixture in a 10 : 1 ratio. Lane **4** shows the band corresponding to the immobilized IgG by CNC-Hydra-Ald-protein A conjugate (1.6 mg) from 2 ml reaction volume containing IgG (50 µg/ml) in the presence of 100 - fold BSA. Lane **6** shows the band corresponding to the immobilized IgG by CNC-Hydra-Ald-proteinA conjugate (1.6 mg) from 0.2 ml reaction volume containing IgG (500 µg/ml) in the presence of 10 - fold BSA. Lane **8** shows the band corresponding to human IgG antibody, and lane **9** the band corresponding to the BSA protein. The bands highlighted by ***** correspond to displaced unfolded Ald-protein A from the surface of CNC-Hydra.

5. 2. 3 Immobilization of Ald-FGE-4C on CNC-Hydra

In order to expand the *in vitro* selective modification of aldehyde-tagged proteins by catalysts which are immobilised on nanocellulose, we constructed an FGE variant which contains the aldehyde-tag at its *N*-terminus (**Figure 5.8**). To demonstrate that Ald-FGE-4C underwent Cys to FGly conversion in *E. coli* during production, we reacted the purified protein with *O*-(4-nitrobenzyl)hydroxylamine. As shown in **Figure 5.8**, following 2 h incubation with the aminooxy reagent, the aldehyde-containing protein was labelled *via* oxime ligation.

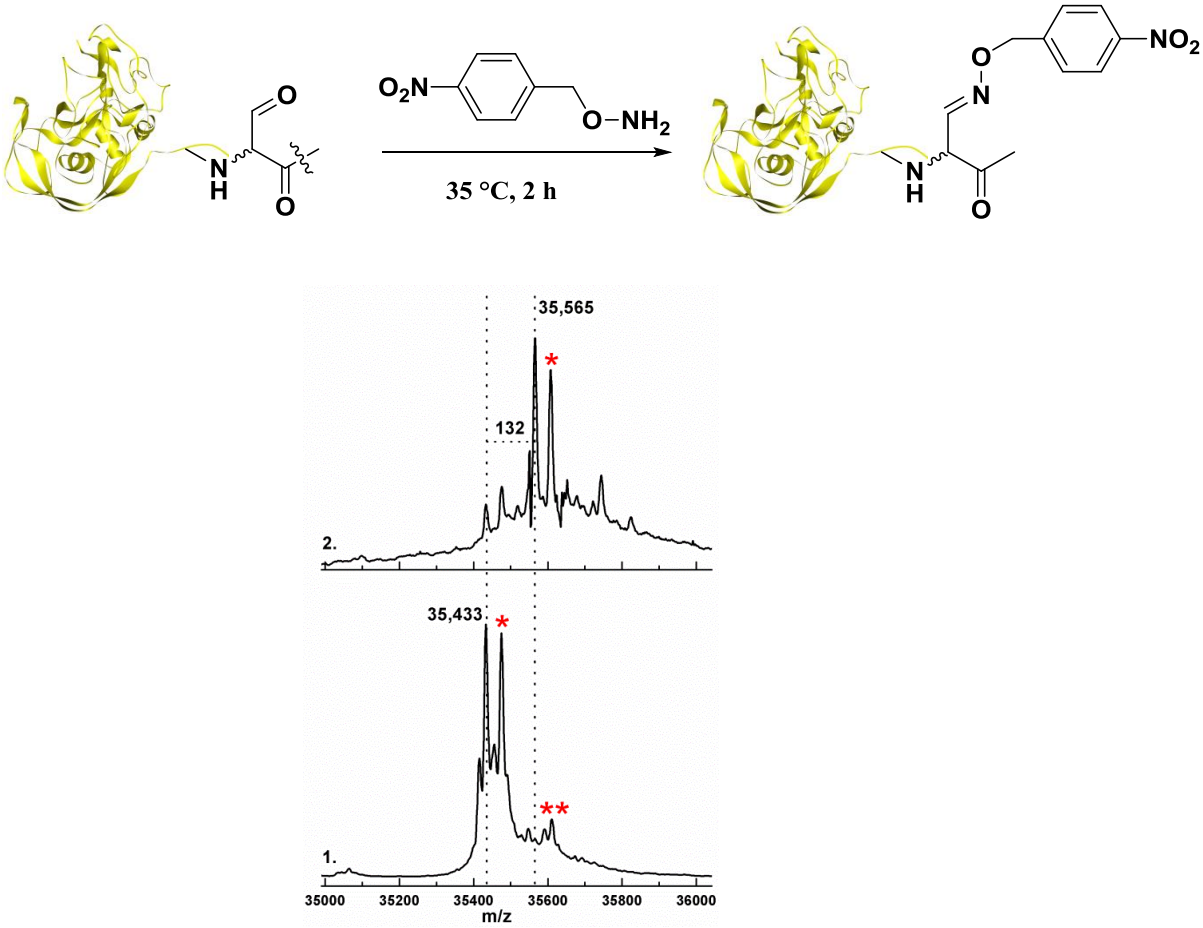


Figure 5.8 Top. Labeling reaction of Ald-FGE-4C (pdb: 1Y1E). **Bottom.** Mass spectrometry analyses of Ald-FGE-4C covalent modification with *O*-(4-nitrobenzyl)hydroxylamine. **1.** Protein mass spectrometry of Ald-FGE-4C after purification. Ald-FGE-4C displays an average mass of 35,433 Da and 35,475 Da (calc. 35,434 Da) **2.** Incubation of Ald-FGE-4C with *O*-(4-nitrobenzyl)hydroxylamine (i.e. 3 h) modifies all starting protein to new species of average mass 35,565 Da (calc. 35,566 Da). Peaks highlighted by * (Δ_{mass} = 42 Da) could probably arise from the carbamylation of a lysine residue. Peaks highlighted by ** (Δ_{mass} = 178 Da) could probably arise from gluconoylation.

Upon showing that Ald-FGE-4C contains the aldehyde functionality, we covalently attached the protein onto the CNC-Hydra nanoparticles *via* hydrazone ligation. We then tested the activity of the immobilised catalyst on Ald-GFP. Full conversion of the Ald-GFP substrate was observed by HRMS monitoring (**Figure 5.9**). Moreover, the hydrazone-linked Ald-FGE-4C is stable for >5h in the presence of thiol-, aldehyde- or diol- containing protein substrates. This was concluded based on the extremely weak fluorescence intensity of the CNC-Hydra-Ald-FGE-4C particles after removal of the oxidised Ald-GFP (**Figure 5.10**, left). This also confirms that all the hydrazine functionalities available on CNC-Hydra were fully occupied by Ald-FGE-4C.

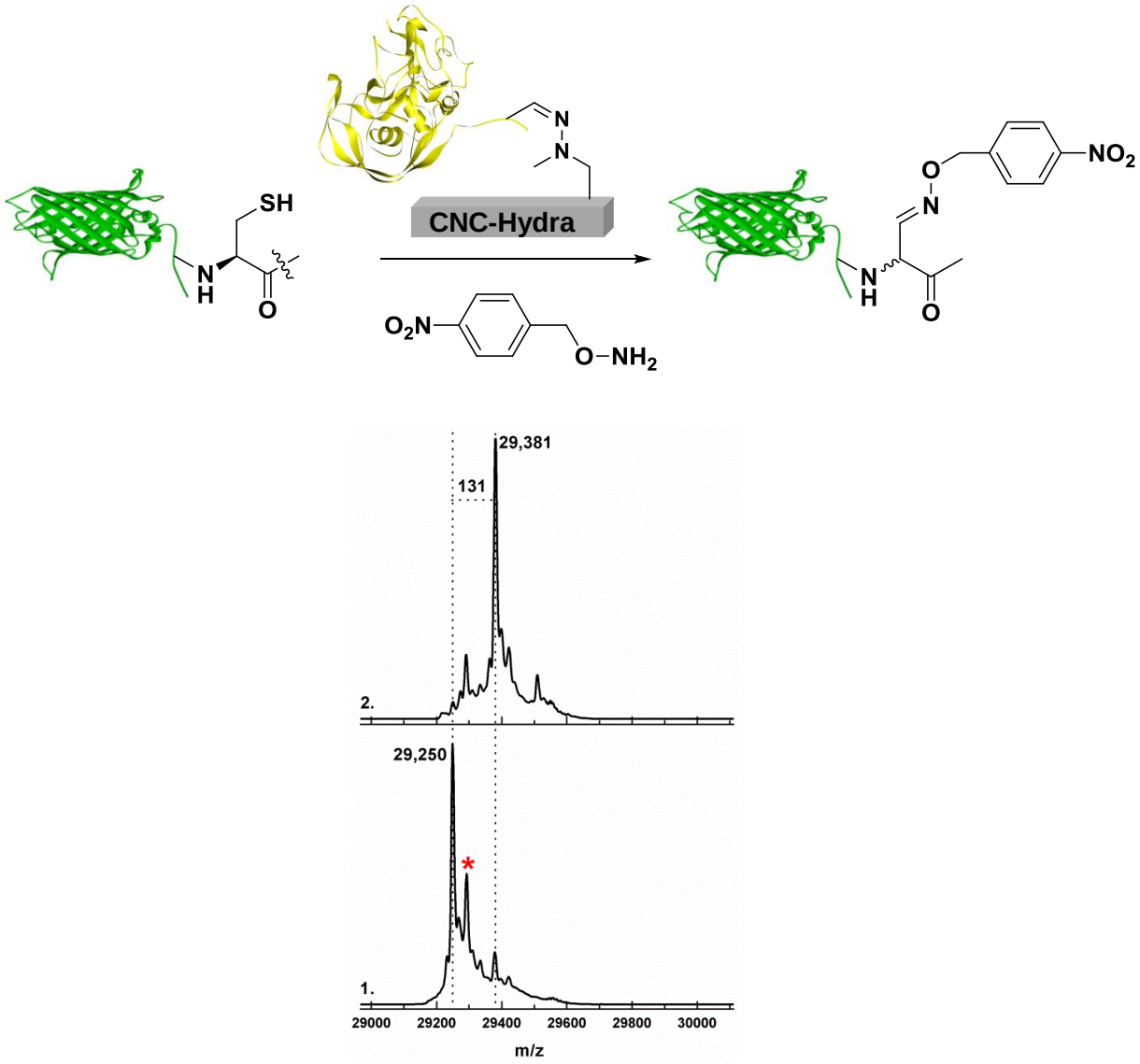


Figure 5.9 Top. Site-selective oxidation of Ald-GFP by nanocellulose immobilised CNC-Hydra. **Bottom.** Mass spectrometry analyses of Ald-GFP covalent modification with *O*-(4-nitrobenzyl)hydroxylamine. **1.** After 10 s incubation with the immobilized Ald-FGE-4C, followed by labelling with *O*-(4-nitrobenzyl)hydroxylamine (i.e. 3 h) Ald-GFP displays an average mass of 29,250 Da (calc. 29,250 Da) **2.** After 5 h incubation with the immobilized Ald-FGE-4C followed by labelling with *O*-(4-nitrobenzyl)hydroxylamine (i.e. 3 h) all starting

protein is modified to new species of average mass 29,381 Da (calc. 29,382 Da). Peaks highlighted by * (Δ_{mass} = 42 Da) could probably arise from acetylation of a lysine residue or acetonitrile adduct.[35]

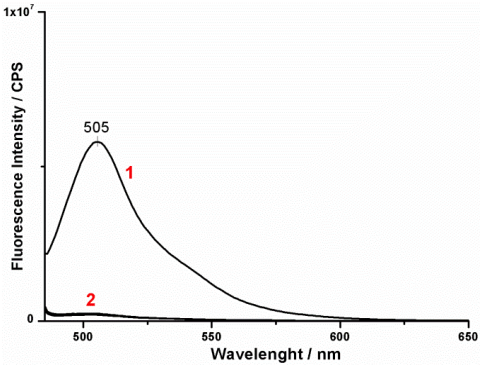
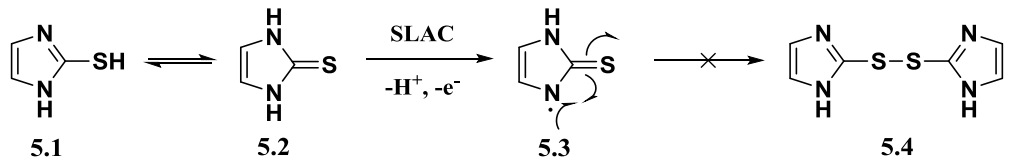


Figure 5.10 Fluorescence emission spectra (**2**) of aqueous suspensions (55.7 µg/ml) of CNC-Hydra-Ald-FGE-4C after 5 h incubation with Ald-GFP. Fluorescence emission spectrum (**1**) corresponds to aqueous suspensions (55.7 µg/ml) of CNC-Hydra-Ald-GFP, included as a reference.

5. 3 Discussion

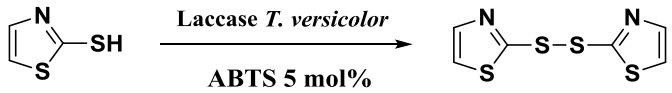
Catalytic functionality was introduced on the surface of nanocellulose by attachment of a bacterial laccase, to which the *N*-terminal aldehyde tag was genetically fused. The immobilised small laccase shows similar catalytic efficiency on 2,6-dimethoxyphenol[193] to its non-immobilised version, when considering a degree of substitution of one protein in 10^4 AGUs. Recently, Hong and co-workers reported the immobilization of a fungal laccase from *Trametes versicolor* on native bacterial nanocellulose.[194] The crystalline cellulose produced by *Gluconacetobacter xylinus* served as a support for adsorption or cross-linkage of laccase *via* glutaraldehyde. The activity of the laccase was tested after immobilization and it turned out that the enzyme retained 69% of its original activity even after seven cycles.[194] More recently, Zillea and co-workers tested the biocompatibility of laccase-bacterial nanocellulose bioconjugates for biomedical purposes as wound dressings.[195] The laccase from *Myceliophthora thermophila* adsorbed onto the surface of bacterial nanocellulose displays antimicrobial effects. The material exerts its bactericidal effect by penetrating the cell walls of both Gram-positive and Gram-negative bacteria.[195]

As SLAC is a highly stable and easily produced laccase, it generated interest for its use as a cathode in biofuel cells[196] or within the biotechnology sector. For this purpose, the exploration of SLAC's activity on other phenolic or heterocyclic compounds could expand its substrate spectrum. In one such attempt, we investigated the laccase-catalysed oxidative dimerization of 2-mercaptoimidazole (**5.1**) which is in equilibrium with its thiourea tautomer (**5.2**), as shown in **Scheme 5.2**. The activity of SLAC for the oxidative dimerization of 2-mercaptoimidazole was monitored as a function of time by reverse phase HPLC (Experimental Section). However, no change in the UV trace at 270 nm was observed after addition of laccase, which preliminarily implies that SLAC does not have the oxidation potential to catalyse this dimerization.



Scheme 5.2 Dimerization of 2-mercaptoimidazole catalysed by SLAC.

This comes in agreement to poor yields (i.e. 26 %) or no conversion reported by Abdel-Mohsen and co-workers who employed laccases from *Trametes versicolor* and *Agaricus bisporus* to undergo dimerization of heterocyclic thiols such as thiazole-2-thiol (**5.5**).[197] However, according to their experiments, it was noticed that when performing the laccase-catalysed reaction in the presence of radical mediators such as TEMPO or ABTS the yields of the reactions were improved to 59 % (**Scheme 5.3**).



5.5 **5.6**

Scheme 5.3 Dimerization of thiazole-2-thiol laccase.

We described a general approach for the oriented attachment of antibodies on the surface of modified crystalline nanocellulose. For this purpose, we employed a truncated form of antibody-binding protein A (Ald-protein A) which was successfully attached to CNC-Hydra. The obtained conjugates served as selective nanosensors for the immobilization of human IgG antibody from a concentrated solution of BSA. When analysing the selectivity of the method, dissociation of the hydrazone linked Ald-protein A from the surface of CNC-Hydra was qualitatively detected by SDS-PAGE analysis. This observation points to the reversibility of the hydrazone conjugation under slightly acidic conditions. The advantage of surface linked protein A relies in the stability of the formed covalent bond under the acidic conditions (pH 2.5) required for the elution of antibodies. To achieve such goals while using FGE as a protein engineering tool, one could explore the conjugation of aldehyde-tagged protein A to amine-functionalised cellulose nanocrystals, following the strategy employed by Edgar and Fox for the modification of cellulose (**Figure 5.11**, **a)**.[198] Alternatively, to avoid the use of additional reduction steps, the Pictet-Spengler ligation developed by Bertozzi and co-workers for the aldehyde-tagged proteins could be used instead (**Figure 5.11**, **b**).[199] Moreover, the surface of CNC could be decorated with aminooxy functionality *via* the reaction scheme proposed in **Figure 5.11**, **c**.[200,201]

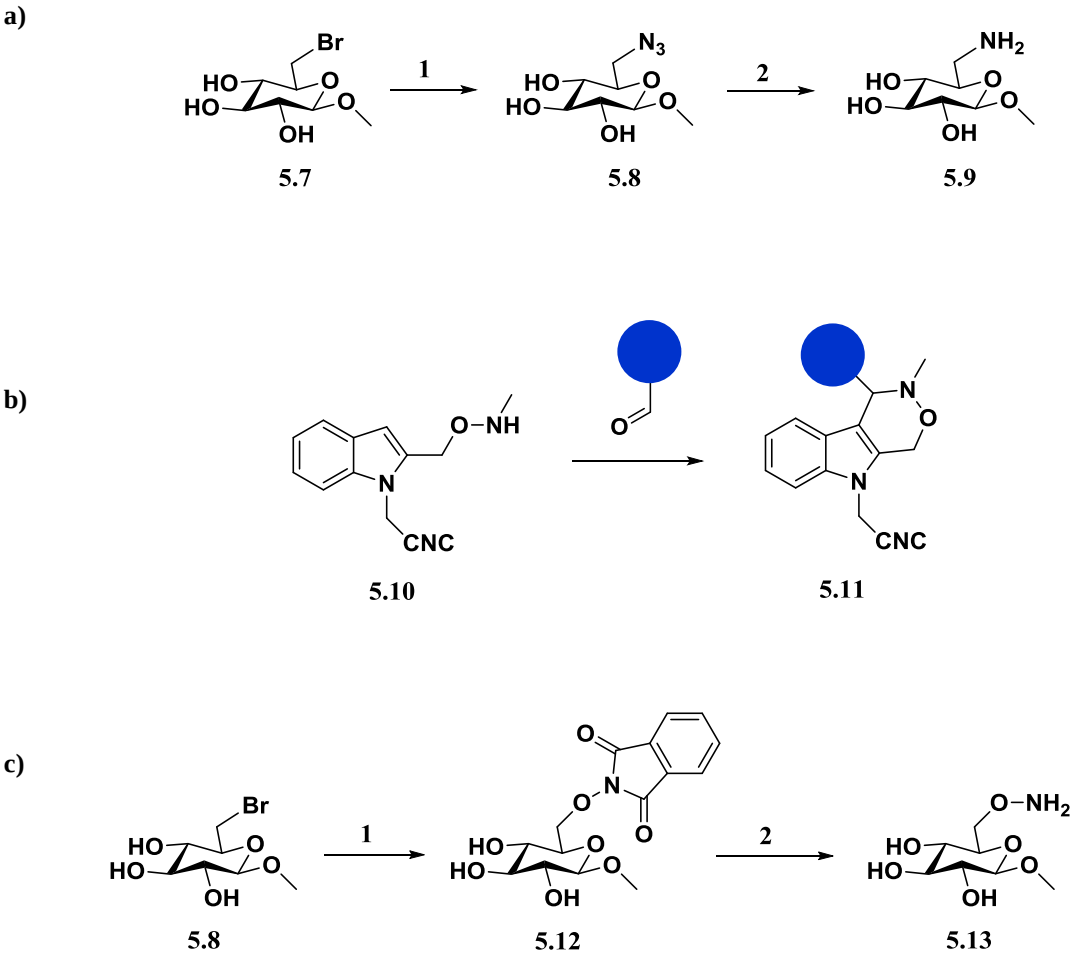


Figure 5.11 Reaction models for the functionalisation of nanocellulose.[198-201] **a) 1.** NaN_3, DMF, 30 h, 70 °C. **2.** PPh_3, MeOH, rt, 24 h. **b)** Pictet-Spengler ligation. **c) 1.** *N*-hydroxyphtalimide, NEt_3, KI, 80 °C **2.** NH_2-NH_2.

Recently, Jia and co-workers utilised FGE for the site-specific immobilization of aldehyde-containing protein A on hydrazine-modified agarose matrices.[202] Accordingly, the protein A resin was employed for the purification of antibodies from human serum. After elution of the bound proteins on the resin with citric buffer (pH 2.5), it was found that the eluted material contained the desired human IgG in a 90% purity.[202]

The multifunctionality of Ald-FGE-4C was explored by linkage of the active catalyst on nanocellulose. The scheme of attaching Ald-FGE-4C convinces through its simplicity. Considering the importance of FGE in the field of biotechnology for generation of antibody-drug conjugates, the advantages of its selective immobilization on cellulose nanocrystals could be further explored.

5. 4 Experimental

Table 5.2 Oligonucleotides used in these experiments.

Primer	Sequence (5'-> 3')
Ald-slac sense	ACGATATCATATGTCGGCACGTACCGCTC
Ald-slac antisense	ATATACTCGAGTTTGATTTCCAGATGTTCGTGTTC
Ald-protA sense	AAAACCGCATATGGCGCAACACGATGAAGCTC
Ald-protA antisense	GCACCAAAAGAGGAAGACAATCGTACGCGCTGCGTTATTGCCTAATCTCGAGGGCTTTT
Ald-fge4C sense	ATATCATATGCCGTCGTTTGACTTTG
Ald-fge4C antisense	ATATCTCGAGTTACAGCGGATC

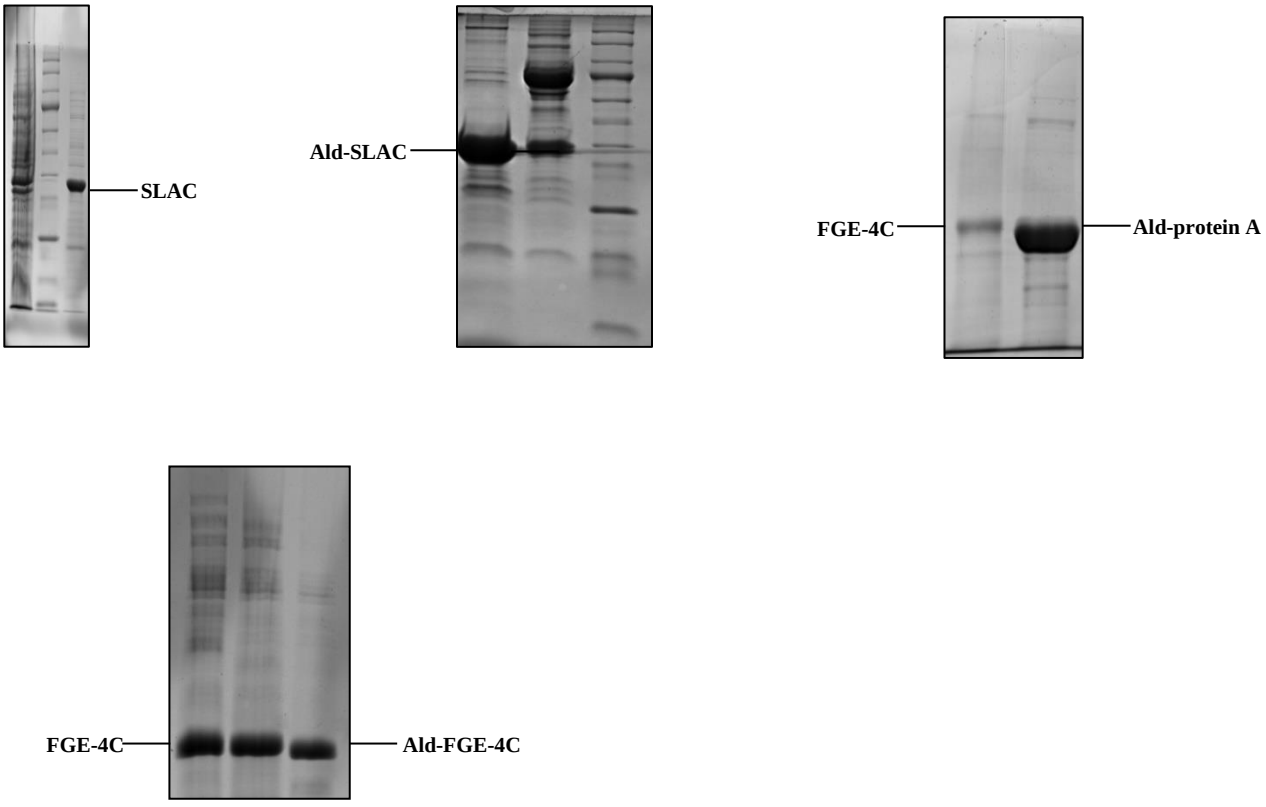
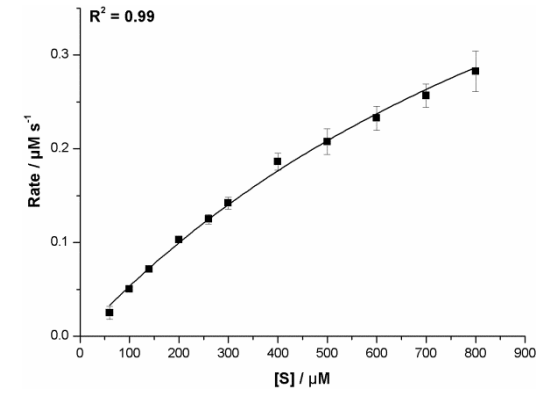
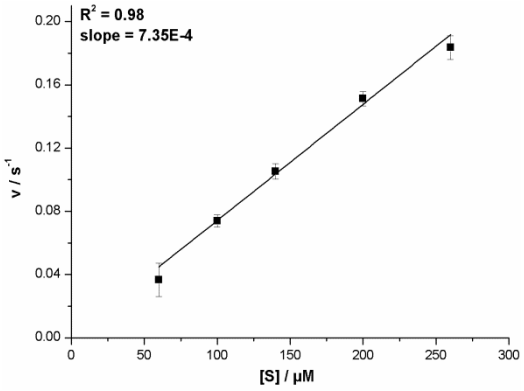


Figure 5.12 SDS-PAGE of SLAC, Ald-SLAC, Ald-protein A, Ald-FGE-4C





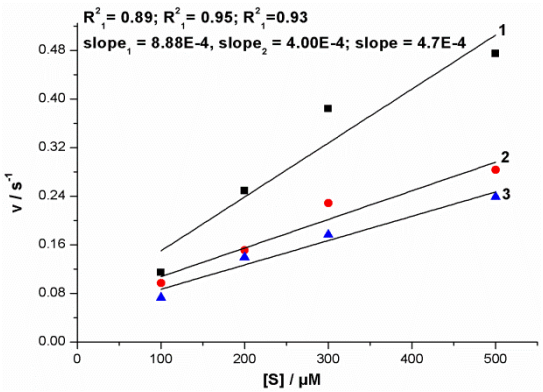


Figure 5.13 Top. Michaelis-Menten plots correlating substrate concentration with rate/[E] for Ald-SLAC on DMP substrate. The catalytic efficiency was determined by linear fitting of first points. **Bottom.** Plots correlating substrate concentration with rate/[E] for the immobilised Ald-SLAC on DMP substrate. Suspension of Ald-Laccase-CNC-Hydra (linear fitting **1**), Ald-Laccase-CNC-Hydra after 6 days storage at -80 °C (linear fitting **2**), Ald-Laccase-CNC-Hydra after overnight, rt (linear fitting **3**). The catalytic efficiency was determined by linear fitting of first points.

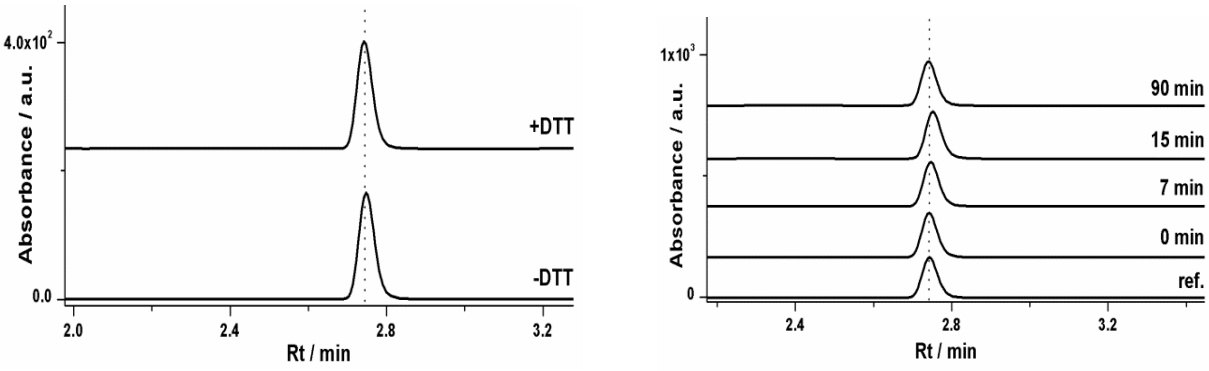


Figure 5.14 Left. HPLC UV trace at 270 nm of 2-mercaptoimidazole treated with DTT. **Right.** HPLC UV trace at 270 nm of reaction monitor of 2-mercaptoimidazole with SLAC as a function of time.

References

1. Barthelat, F. *Bioinsp. Biomim.* **2010**, *5*, 1-8 'Nacre from mollusk shells: a model for high-performance structural materials'

2. Rho, J.; Kuhn-Spearing, L.; Zioupos, P. *Med. Eng. Phys.* **1998**, *20*, 92–102 'Mechanical properties and the hierarchical structure of bone'

3. Oh, D. X.; Cha, Y. J.; Nguyen, H.; Je, H. H.; Jho, Y. S.; Hwang, D. S.; Yoon, D. K. *Sci. Rep.* **2016**, *6*, 1-6 'Chiral nematic self-assembly of minimally surface damaged chitin nanofibrils and its load bearing functions'

4. Keckes, J.; Burgert, I.; Frühmann, K.; Müller, M.; Kölln, K.; Hamilton, M.; Burghammer, M.; Roth, S. V.; Stanzl-Tschegg, S.; Fratzl, P. *Nat. Mat.* **2003**, *2*, 810 – 813 'Cell-wall recovery after irreversible deformation of wood'

5. Barthelat, F.; Yin, Z.; Buehler, M. J. *Nat. Rev. Mat.* **2016**, *1*, 1-16 'Structure and mechanics of interfaces in biological materials'

6. Yahyazadehfar, M.; Arola, D. *Acta Biomater.* **2015**, *19*, 33-45 'The role of organic proteins on the crack growth resistance of human enamel'

7. Moon, R. J.; Martini, A.; Nairn, J.; Simonsen, J.; Youngblood, J. *Chem. Soc. Rev.* **2011**, *40*, 3941-3994 'Cellulose nanomaterials review: structure, properties and nanocomposites'

8. Somerville, C. *Annu. Rev. Cell Dev. Biol.* **2006**, *22*, 53-78 'Cellulose synthesis in higher plants'

9. Sørensen, I.; Domozych, D.; Willats, W. G. T. *Plant Physiol.* **2010**, *153*, 366–372 'How have plant cell walls evolved?'

10. Cosgrove, D. J. *Nat. Rev.* **2005**, *6*, 850-861 'Growth of the plant cell wall'

11. Cosgrove, D. J. *Nat. Rev.* **2000**, *407*, 321-326 'Loosening of plant cell walls by expansins'

12. Fry, S. C. *Curr. Biol.* **1994**, *4*, 815-817 'Unzipped by expansins'

13. Habibi, Y. *Chem. Soc. Rev.* **2014**, *43*, 1519 'Key advances in the chemical modification of nanocelluloses'

14. Arioli, T.; Peng, L.; Betzner, A. S.; Burn, J.; Wittke, W.; Herth, W.; Camilleri, C.; Hofte, H.; Plazinski, J.; Birch, R.; Cork, A.; Glover, J.; Redmond, J.; Williamson, R. E. *Science* **1998**, *279*, 717–720 'Molecular analysis of cellulose biosynthesis in *Arabidopsis*'

15. Ross, P.; Mayer, R.; Benziman, M. *Micobiol. Rev.* **1991**, *55*, 35-58 'Cellulose biosynthesis and function in bacteria'

16. Römling, U.; Galperin, M. G. *Trends Microbiol.* **2015**, *23*, 545–557 'Bacterial cellulose biosynthesis: diversity of operons, subunits, products and functions'

17. Morgan, J. L.; Strumillo, J.; Zimmer, J. *Nature* **2013**, *493*, 181-186 'Crystallographic snapshot of cellulose synthesis and membrane translocation'

18. Morgan, J. L. W.; McNamara, J. T.; Fischer, M.; Rich, J.; Chen, H-M.; Withers, S. G.; Zimmer, J. *Nature* **2016**, *531*, 329-334 'Observing cellulose biosynthesis and membrane translocation *in crystallo*'

19. Lairson, L. L., Henrissat, B., Davies, G. J.; Withers, S. G. *Annu. Rev. Biochem.* **2008**, *77*, 521–555 'Glycosyltransferases: structures, functions, and mechanisms'

20. Cramer, F. B.; Hockett, R. C.; Purves, C. B. *J. Am. Chem. Soc.* **1939**, *61*, 3463–3464 'The unesterified secondary hydroxyls in acetone soluble cellulose acetate'

21. Tokiwa, Y.; Calabia, B. P.; Ugwu, C. U.; Aiba, S. *Int. J. Mol. Sci.* **2009**, *10*, 3722–3742 'Biodegradability of plastics'

22. Alley, R. *Catalogue of the Tate Gallery's collection of modern art other than works British artists*, Tate Gallery and Sotheby Parke-Bernet, London **1981**, 239-40

23. Law, K. L.; Morét-Ferguson, S.; Maximenko, N. A.; Proskurowski, G.; Peacock, E. E.; Hafner, J.; Reddy, C. M. *Science* **2010**, *329*, 1185-1188 'Plastic accumulation in the North Atlantic subtropical gyre'

24. Ranby, B.G. *Acta Chim. Scand.* **1949**, *3*, 649-650 'Aqueous colloidal solutions of cellulose micelles'

25. Herzog, R. O.; Jancke, W. *Ber.* **1920**, *53*, 2162-2164 'Über den physikalischen Aufbau einiger hochmolekularer organischer Verbindungen'

26. Ribi, E. *Nature* **1951**, *168*, 1082-1083 'Submicroscopic structure of fibres and their formation'

27. Habibi, Y.; Lucia, L. A.; Rojas, O. J. *Chem. Rev.* **2010**, *110*, 3479-3500 'Cellulose nanocrystals: chemistry, self-assembly and applications'

28. Mariano, M.; Kissi, N. E.; Dufresne, A. *Polym. Phys.* **2014**, *52*, 791-806 'Cellulose nanocrystals and related nanocomposites: review of some properties and challenges'

29. Usov, I.; Nystro, G.; Adamcik, J.; Handschin, S.; Schutz, C.; Fall, A.; Bergstrom, L; Mezzenga, R. *Nat. Comm.* **2015**, *6*, 7564 'Morphology of nanocellulose by TEM'

30. Valo, H.; Arola, S.; Laaksonen, P.; Torkkeli, M.; Peltonen, L.; Linder, M. B.; Serimaa, R.; Kuga, S.; Hirvonen, J.; Laaksonen, T. *Eur. J. Pharm. Sci.* **2013**, *50*, 69-77 'Drug release from nanoparticles embedded in four different nanofibrillar cellulose aerogels'

31. Taheri, A.; Mohammadi, M. *Chem. Biol. Drug. Des.* **2015**, *86*, 102–106 'The use of cellulose nanocrystals for potential application in topical delivery of hydroquinone'

32. Lam, E.; Male, K. B.; Chong, j. H.; Leung, A. C. W.; Luong, J. H. T. *Trends Biotechnol.* **2012**, *30*, 283-290 'Applications of functionalized and nanoparticle-modified nanocrystalline cellulose'

33. Eyley, S.; Thielemans, W. *Nanoscale* **2014**, *6*, 7764-7779 'Surface modification of cellulose nanocrystals'

34. De Nooy, A. E. J.; Besemer, A. C.; van Bekkum, H. *Recl. TraV. Chim. Pays-Bas* **1994**, *113*, 165 'Highly selective tempo mediated oxidation of primary alcohol groups in polysaccharides'

35. Sassi, J.; Chanzy, H. Cellulose **1995**, *2*, 111-127 'Ultrastructural aspects of the acetylation of cellulose'

36. Eyley, S.; Thielemans, W. *Chem. Commun.* **2011**, *47*, 4177–4179 'Imidazolium grafted cellulose nanocrystals for ion exchange applications'

37. Bahl, O. P.; Hullar, T. L.; Smith, F. *J. Org. Chem.* **1964**, *29*, 1076–1078 'Reduction of the products of periodate oxidation of carbohydrates. XIII. Determination of sugars in polysaccharides oxidized by periodate'

38. Sergiu Coseri, S.; Biliuta, G.; Simionescu, B. C.; Stana-Kleinschek, K.; Ribitsch, V.; Harabagiu, V. *Carbohydr. Polym.* **2013**, *93*, 207–215 'Oxidized cellulose - survey of the most recent achievements'

39. Hasani, M.; Cranston, E. D.; Westmana, G.; Gray, D. G. *Soft Matter* **2008**, *4*, 2238–2244 'Cationic surface functionalization of cellulose nanocrystals'

40. Zhang, Z.; Sèbe, G.; Rentsch, D.; Zimmermann, T.; Tingaut, P. *Chem. Mater.* **2014**, *26*, 2659–2668 'Ultralightweight and flexible silylated nanocellulose sponges for the selective removal of oil from water'

41. Hassan, M. L.; Moorefield, C. M.; Elbatal, H. S.; Newkome, G. R.; Modarelli, D. A.; Romano, N. C. *Mater. Sci. Eng. B* **2012**, *177*, 350-358 'Fluorescent cellulose nanocrystals *via* supramolecular assembly of terpyridine-modified cellulose nanocrystals and terpyridine-modified perylene'

42. Uth, C.; Zielonka, S.; Horner, S.; Rasche, N.; Plog, A.; Orelma, H.; Avrutina, O.; Zhang, K.; Kolmar, H. *Angew. Chem. Int. Ed.* **2014**, *53*, 1-7 'A chemoenzymatic approach to protein immobilization onto crystalline cellulose nanoscaffolds'

43. Huang, J.; Li, C.; Gray, D. G. *ACS Sustain. Chem. Eng.* **2013**, *1*, 1160−1164 'Cellulose nanocrystals incorporating fluorescent methylcoumarin groups'

44. Jay W. Grate, J. W. Kai-For Mo, K.; Yongsoon Shin, Y.; Andreas Vasdekis, A.; Marvin G. Warner, M. G.; Kelly, R. T.; Orr, G.; Hu, D.; Dehoff, K. J.; Brockman, F. J.; Wilkins, M. J. *Bioconjugate Chem.* **2015**, *26*, 593−601 'Alexa Fluor-labeled fluorescent cellulose nanocrystals for bioimaging solid cellulose in spatially structured microenvironments'

45. Yang, X.; Cranston, E. D. *Chem. Mater.* **2014**, *26*, 6016-6025 'Chemically cross-linked cellulose nanocrystal aerogels with shape recovery and superabsorbent properties'

46. Yang, X.; Shi, K.; Zhitomirsky, I; Cranston, E. D. *Adv. Mater.* **2015**, *27*, 6104-6109 'Cellulose nanocrystal aerogels as universal 3D lightweight substrates for supercapacitor materials'

47. Ivanova, A.; Fattakhova-Rohlfing, D.; Kayaalp, B. E.; Jiri Rathouský, J.; T. *J. Am. Chem. Soc.* **2014**, *136*, 5930–5937 'Tailoring the morphology of mesoporous titania thin tilms through biotemplating with nanocrystalline cellulose'

48. Dong, S.; Roman, M. *J. Am. Chem. Soc.* **2007**, *129*, 13810-13811 'Fluorescently labeled cellulose nanocrystals for bioimaging applications'

49. Mahmoud, K. A.; Mena, J. A.; Male, K. B.; Hrapovic, S.; Kamen, A.; Luong, J. H. T. *ACS Appl. Mater. Interfaces* **2010**, *2*, 2924-2932 'Effect of surface charge on the cellular uptake and cytotoxicity of fluorescent labeled cellulose nanocrystals'

50. Colombo, L.; Zoia, L.; Violatto, M. B.; Previdi, S.; Talamini, L.; Sitia, L.; Nicotra, F.; Orlandi, M.; Salmona, M.; Recordati, C.; Bigini, P.; La Ferla, B. *Biomacromolecules* **2015**, *16*, 2862–2871 'Organ distribution and bone tropism of cellulose nanocrystals in living mice'

51. Malho, J. Arola, S.; Laaksonen, P.; Szilvay, G. R.; Ikkala, O.; Linder, M. B. *Angew. Chem. Int. Ed.* **2015**, *54*, 12025 –12028 'Modular architecture of protein binding units for designing properties of cellulose nanomaterials'

52. Malho, J.; Ouellet-Plamondon, C.; Ruggeberg, M.; Laaksonen, P.; Ikkala, O.; Burgert, I.; Linder, M. B. *Biomacromolecules* **2015**, *16*, 311-318 'Enhanced plastic deformations of nanofibrillated cellulose film by adsorbed moisture and protein-mediated interactions'

53. Linder, M.; R. Szilvay, G. R.; Nakari-Setälä, T.; Söderlund, H.; Penttilä, M. *Protein Sci.* **2002**, *11*, 2257–2266 'Surface adhesion of fusion proteins containing the hydrophobins HFBI and HFBII from *Trichoderma reesei*'

54. Happs, R.M.; Guan, X.; Resch, M.G.; Davis, M.F.; Beckham, G.T.; Tan, Z.; Crowley, M.F. *Febs J.* **2015**, *282*, 4341-4356 '*O*-glycosylation effects on family 1 carbohydrate-binding module solution structures'

55. Hakanpaa, J. M.; Szilvay, G. R.; Kaljunen, H.; Maksimainen, M.; Linder, M.; Rouvinen, J. *Protein Sci.* 2006, *15*, 2129-2140 'Two crystal structures of *Trichoderma reesei* hydrophobin HFBI. The structure of a protein amphiphile with and without detergent interaction'

56. Szilvay, G. R., Paananen, A.; Laurikainen, K.; Vuorimaa, E.; Lemmetyinen, H.; Peltonen, J.; Linder, M. B *Biochemistry* **2007**, *46*, 2345–2354 'Self-assembled hydrophobin protein films at the air−water interface: structural analysis and molecular engineering'

57. Malho, J.; Arola, S.; Laaksonen, P.; Szilvay, G. R.; Ikkala, O.; Linder, M. B. *Angew. Chem. Int. Ed.* **2015**, *54*, 12025–12028 'Modular architecture of protein binding units for designing properties of cellulose nanomaterials'

58. Fritz, C.; Jeuck, B.; Salas, C.; Gonzalez, R.; Jameel, H.; Rojas, O. J. *Adv. Polym. Sci.* **2015** 'Nanocellulose and proteins: exploiting their interactions for production, immobilization and synthesis of biocompatible materials'

59. Arola, S.; Tammelin, T.; Setala, H.; Tullila, A.; Linder, M. B. *Biomacromolecules* **2012**, *13*, 594–603 'Immobilization–stabilization of proteins on nanofibrillated cellulose derivatives and their bioactive film formation'

60. Rao, S. V.; Anderson, K. W.; Bachas, L. G. *Mikrochim. Acta* **1998**, *128*, 127-143 'Oriented immobilization of proteins'

61. Murphy, C. J.; Gole, A. M.; Stone, J. W.; Sisco, P. N.; Alkilany, A. M.; Goldsmith, E. C.; Baxter, S. C. *Acc. Chem. Res.* **2008**, *41*, 1721–1730 'Gold nanoparticles in biology: beyond toxicity to cellular imaging'

62. Mahmoud, K. A.; Male, K. B.; Hrapovic, S.; Luong, J. H. T. *ACS Appl. Mater. Interfaces* **2009**, *7*, 1383-1386 'Cellulose nanocrystal/gold nanoparticle composite as a matrix for enzyme immobilization'

63. Abad, J. M.; Mertens, S. F. L.; Pita, M.; Fernandez, V. M.; Schiffrin, D. J. *J. Am. Chem. Soc.* **2005**, *127*, 5689-5694 'Functionalization of thioctic acid-capped gold nanoparticles for specific immobilization of histidine-tagged proteins'

64. Incani, V.; Danumah, C.; Boluk, Y. *Cellulose* **2013**, *20*, 191–200 'Nanocomposites of nanocrystalline cellulose for enzyme immobilization'

65. Karaaslan, M. A.; Gao, G.; Kadla, J. F. *Cellulose* **2013**, *20*, 2655–2665 'Nanocrystalline cellulose/*β*-casein conjugated nanoparticles prepared by click chemistry'

66. Kolb, H. C.; Finn, M. G.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2001**, *40*, 2004-2021 'Click chemistry: diverse chemical function from a few good reactions'

67. Wong, L. S.; Khan, F.; Micklefield, J. *Chem. Rev.* **2009**, *109*, 4025–4053 'Selective covalent protein immobilization: strategies and applications'

68. Mao, H.; Hart, S. A.; Schink, A.; Pollok, B. A. *J. Am. Chem. Soc.* **2004**, *126*, 2670–2671 'Sortase-mediated protein ligation: a new method for protein engineering'

69. Guimaraes, C. P.; Witte, M. D.; Theile, C. S.; Bozkurt, G.; Kundrat, L.; Annet E M Blom**, A. E. M.;** Ploegh, H. L. *Nat. Protoc.* **2013**, *8*, 1787-1799 'Site-specific *C*-terminal and internal loop labeling of proteins using sortase-mediated reactions'

70. Guo, X.; Wang, Q.; Swarts, B. M.; Guo, Z. *J. Am. Chem. Soc.* **2009**, *131*, 9878–9879 'Sortase-catalyzed peptide−glycosylphosphatidylinositol analogue ligation'

71. Samantaray, S.; Marathe, U.; Dasgupta, S.; Nandicoori, V. K.; Roy, R. P. *J. Am. Chem. Soc*. **2008**, *130*, 2132–2133 'Peptide−sugar ligation catalyzed by transpeptidase sortase: a facile approach to neoglycoconjugate synthesis'

72. Beerli, R. R.; Hell, T.; Merkel, A. S.; Grawunder, U. *PLoS One.* **2015**, *10*, 1-17 'Sortase enzyme-mediated generation of site-specifically conjugated antibody drug conjugates with high *in vitro* and *in vivo* potency'

73. Ilangovan, U.; Ton-That, H.; Iwahara, J.; Schneewind, O.; Clubb, R.T. *Proc. Natl. Acad. Sci.* **2001**, *98*, 6056-6061 'Structure of sortase, the transpeptidase that anchors proteins to the cell wall of *Staphylococcus aureus*'

74. Rush, J. S.; R. Bertozzi, C. R. *J. Am. Chem. Soc.* **2008**, *130*, 12240–12241 'New aldehyde tag sequences identified by screening formylglycine generating enzymes *in vitro* and *in vivo*'

75. Carlson, B.L.; Ballister, E. R.; Skordalakes, E.; King, D. S.; Breidenbach, M. A.; Gilmore, S. A.; Berger, J. M.; Bertozzi, C. R. *J. Biol. Chem.* **2008**, *283*, 20117-20125 'Function and structure of a prokaryotic formylglycine-generating enzyme'

76. Rommerskirch, W.; von Figura, K. *Proc. Natl. Acad. Sci.* **1992**, *89*, 2561-2565 'Multiple sulfatase deficiency: catalytically inactive sulfatases are expressed from retrovirally introduced sulfatase cDNAs'

77. Schlotawa, L.; Radhakrishnan, K.; Baumgartner, M.; Schmid, R.; Schmidt, B.; Dierks, T.; Gärtner, J. *Eur. J. Hum. Genet.* **2013**, *21*, 1020–1023 'Rapid degradation of an active formylglycine generating enzyme variant leads to a late infantile severe form of multiple sulfatase deficiency'

78. Dierks, T.; Dickmanns, A.; Preusser-Kunze, A.; Schmidt, B.; Mariappan, M.; von Figura, K.; Ficner, R.; Rudolph, M. G. *Cell* **2005**, *121*, 541-52 'Molecular basis for multiple sulfatase deficiency and mechanism for formylglycine generation of the human formylglycine-generating enzyme'

79. Carrico, I. S.;Carlson, B. L.; R Bertozzi, C. R. *Nat. Chem. Biol.* **2007**, *3*, 321 – 322 'Introducing genetically encoded aldehydes into proteins'

80. Hudak, J. E.; Barfield, R. M.; de Hart, G. W.; Grob, P.; Nogales, E.; Bertozzi, C. R.; Rabuka, D. *Angew. Chem. Int. Ed.* **2012**, *51*, 4161 –4165 'Synthesis of heterobifunctional protein fusions using copper-free click chemistry and the aldehyde tag'

81. Rabuka, D.; Rush, J. S.; deHart, G. W.; Wu, P.; Bertozzi, C. R. *Nat. Protoc.* **2012**, *7*, 1052-1067 'Site-specific chemical protein conjugation using genetically encoded aldehyde tags'

82. Hudak, J. E.;Yu, H. H.; Bertozzi, C. R. *J. Am. Chem. Soc.* **2011**, *133*, 16127–16135 'Protein glycoengineering enabled by the versatile synthesis of aminooxy glycans and the genetically encoded aldehyde tag'

83. York, D.; Baker, J.; Holder, P. G.; Jones, L. C.; Drake, P. M.; Barfield, R. M.; Bleck, G. T.; Rabuka, D. *BMC Biotechnol.* **2016**, *16*, 1-11 'Generating aldehyde-tagged antibodies with high titers and high formylglycine yields by supplementing culture media with copper(II)'

84. Spiro, R. G. *Glycobiology* **2002**, *12*, 43-56 'Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds'

85. Haines, N.; Irvine, K. D.; *Nat. Rev.* **2003**, *4*, 786-797 'Glycosylation regulates Notch signalling'

86. Bisello, A.; Greenberg, Z.; Behar, V.; Rosenblatt, M.; Suva, L. J.; Chorev, M.; *Biochemistry* **1996**, *35*, 15890-15895 'Role of glycosylation in expression and function of the human parathyroid hormone/parathyroid hormone-related protein receptor'

87. Christiansen, M. N.; Chik, J.; Lee, L.; Anugraham, M.; Abrahams, J. L.; Packer, N. H. *Proteomics* **2014**, *14*, 525 'Cell surface protein glycosylation in cancer'

88. Ubersax, J. A.; Ferrell, J. E. *Nat. Rev.* **2007**, *8*, 530-541'Mechanisms of specificity in protein phosphorylation'

89. Kouzarides, T. *EMBO J.* **2000**, *19*, 1176-1179 'Acetylation: a regulatory modification to rival phosphorylation?'

90. Aletta, J. M.; Cimato, T. R.; Ettinger, M. J. *Trends Biochem. Sci.* **1998**, *3*, 89-91 'Protein methylation : a signal event in post-translational modification'

91. Walsh, C. T. 'Posttranslational modification of proteins: expanding nature's inventory' **2006**, Roberts and Company Publishers, Colorado

92. Walsh, G.; Jefferis, R. *Nat. Biotechnol.* **2006**, *24*, 1241-1252 'Post-translational modifications in the context of therapeutic proteins'

93. Wang, L.; Amin, M. H. *Chem. Biol.* **2014**, *21*, 51-66 'Chemical and chemoenzymatic synthesis of glycoproteins for deciphering functions'

94. Kasteren, S. I.; Kramer, H. B.; Gamblin, D. P.; Davis, B. G. *Nat. Protoc.* **2007**, *2*, 3185-3194 'Site-selective glycosylation of proteins: creating synthetic glycoproteins'

95. Dierks, T.; Schmidt, B.; Borissenko, L.V.; Peng, J.; Preusser, A.; Mariappan, M.; von Figura, K. *Cell* **2003**, *113*, 435-444 'Multiple sulfatase deficiency is caused by mutations in the gene encoding the human C_{α}-formylglycine generating enzyme'

96. Liu, J.; Hanne, J.; Britton, B. M.; Shoffner, M.; Albers, A. E.; Bennett, J.; Zatezalo, R.; Barfield, R.; Rabuka, D.; Lee, J-B.; Fishel, R. *Sci. Rep.* **2015**, *5*, 16883 'An efficient site-specific method for irreversible covalent labeling of proteins with a fluorophore'

97. Liang, S. I.; McFarland, J. M.; Rabuka, D.; Gartner, Z. J. *J. Am. Chem. Soc.* **2014**, *136*, 10850−10853 'A modular approach for assembling aldehyde-tagged proteins on DNA scaffolds'

98. Rodríguez-Zavala, J. S.; Allali-Hassani, A.; Weiner, H. *Protein Sci.* **2006**, *15*, 1387–1396 'Characterization of *E. coli* tetrameric aldehyde dehydrogenases with atypical properties compared to other aldehyde dehydrogenases'

99. Knop, M.; Engi, P.; Lemnaru, R.; Seebeck, F. P. *ChemBioChem* **2015**, *16*, 2147–2150 '*In vitro* reconstitution of formylglycine-generating enzymes requires copper(I)'

100. Meining, W.; Mortl, S.; Fischer, M.; Cushman, M.; Bacher, A.; Ladenstein, R. *J. Mol. Biol.* **2000**, *299*, 181-197 'The atomic structure of pentameric lumazine synthase from *Saccharomyces cerevisiae* at 1.85-Å resolution reveals the binding mode of a phosphonate intermediate analogue'

101. Mougous, J. D.; Green, R. E.; Williams, S. J.; Brenner, S. E.; Bertozzi, C. R. *Chem. Biol.* **2002**, *9*, 767-76 'Sulfotransferases and sulfatases in mycobacteria'

102. Barondeau, D. P.; Kassmann, C. J.; Tainer, J. A.; Getzoff, E. D. *J. Am. Chem. Soc.* **2007**, *129*, 3118-3126 'The case of the missing ring: radical cleavage of a carbon-carbon bond and implications for GFP chromophore biosynthesis'

103. Seebeck, F. P.; Woycechowsky, K. J.; Zhuang, W.; Rabe, J. P.; Hilvert, D. *J. Am. Chem. Soc.* **2006**, *128*, 4516–4517 'A simple tagging system for protein encapsulation'

104. Wörsdörfer, B.; Woycechowsky, K. J.; Hilvert, D. *Science* **2011**, *331*, 589-92 'Directed evolution of a protein container'

105. Min, J.; Kim, S.; Lee, J.; Kang, S. ***RSC Adv.*** **2014**, ***4***, 48596-48600 'Lumazine synthase protein cage nanoparticles as modular delivery platforms for targeted drug delivery'

106. Zhang, X.; Meining, W.; Fischer, M.; Bacher, A.; Ladenstein, R. *J. Mol. Biol.* **2001**, *306*, 1099-1114 'X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile *Aquifex aeolicus* at 1.6-Å resolution: determinants of thermostability revealed from structural comparisons'

107. Vieille, C.; Zeikus, G. J. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 1-43 'Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability'

108. Mukhi, N.; Dhindwal, S.; Uppal, S.; Kapoor, A.; Arya, R.; Kumar, P.; Kaur, J.; Kundu, S. *Biochemistry* **2016**, *55*, 1724–1740 'Structural and functional significance of the *N*- and *C*-terminal appendages in *Arabidopsis* truncated hemoglobin'

109. Wong, L-W.; Tae, H-S.; Cromer, B. A. *ACS Chem. Neurosci.* **2014**, *5*, 1266–1277 'Role of the ρ1 $GABA_C$ receptor *N*-terminus in assembly, trafficking and function'

110. Yang, F.; Moss, L. G.; Phillips Jr., G. N. *Nat. Biotechnol.* **1996**, *14*, 1246-1251 'The molecular structure of green fluorescent protein'

111. Abedi, M. R.; Caponigro, G.; Kamb, A. *Nucleic Acid Res.* **1998**, *26*, 623-630 'Green fluorescent protein as a scaffold for intracellular presentation of peptides'

112. Roeser, D.; Preusser-Kunze, A.; Schmidt, B.; Gasow, K.; Wittmann, J. G.; Dierks, T.; von Figura, K.; Rudolph, M. G. *Proc. Natl. Acad. Sci.* **2006**, *103*, 81-86 'A general binding mechanism for all human sulfatases by the formylglycine-generating enzyme'

113. Kollipara, L.; Zahedi, R. P. *Proteomics* **2013**, *13*, 941–944 'Protein carbamylation: *in vivo* modification or *in vitro* artefact?'

114. Geoghegan, K. F.; Dixon, H. B.; Rosner, P. J.; Hoth, L. R.; Lanzetti, A. J.; Borzilleri, K. A.; Marr, E. S.; Pezzullo, L. H.; Martin, L. B.; LeMotte, P. K.; McColl, A. S.; Kamath, A.

V.; Stroh, J. G. *Anal. Biochem.* **1999**, *267*, 169-84 ‘Spontaneous alpha-*N*-phosphogluconoylation of a "His tag" in *Escherichia coli*: the cause of extra mass of 258 or 178 Da in fusion proteins’

115. Boltes, I.; Czapinska, H.; Kahnert, A.; Von Buelow, R.; Dirks, T.; Schmidt, B.; von Figura, K.; Kertesz, M. A.; Uson, I. *Structure* **2001**, *9*, 483-491 ‘1.3 Å Structure of arylsulfatase from *Pseudomonas Aeruginosa* establishes the catalytic mechanism of sulfate ester cleavage in the sulfatase family’

116. Smith, E. L.; Giddens, J. P.; Iavarone, A. T.; Godula, K.; Wang, L-X.; Bertozzi, C. R. *Bioconjugate Chem.* **2014**, *25*, 788–795 ‘Chemoenzymatic Fc glycosylation *via* engineered aldehyde tags’

117. Holder, P. G.; Jones, L. C.; Drake, P. M.; Barfield, R. M.; Banas, S.; de Hart, G. W.; Baker, J.; Rabuka, D. *J. Biol. Chem.* **2015**, 290, 15730-15745 ‘Reconstitution of formylglycine-generating enzyme with copper(II) for aldehyde tag conversion’

118. Drake, P. M.; Albers, A. E.; Baker, J.; Stefanie Banas, S.; Barfield, R. M.; Bhat, A. S.; de Hart, G. W.; Garofalo, A. W.; Holder, P.; Jones, L. C.; Kudirka, R.; McFarland, J.; Zmolek, W.; Rabuka, D. *Bioconjugate Chem.* **2014**, *25*, 1331–1341 ‘Aldehyde tag coupled with HIPS chemistry enables the production of ADCs conjugated site-specifically to different antibody regions with distinct *in vivo* efficacy and PK outcomes’

119. Hang, C. H.; Bertozzi, C. R. *Acc. Chem. Res.* **2001**, *34*, 727-736 ‘Chemoselective approaches to glycoprotein assembly’

120. Lang, K.; Chin, J. W. *ACS Chem. Biol.* **2014**, *9*, 16−20 ‘Bioorthogonal reactions for labeling proteins’

121. Nigst, T. A.; Antipova, A.; Mayr, H. *J. Org. Chem.* **2012**, *77*, 8142−8155 ‘Nucleophilic reactivities of hydrazines and amines: the futile search for the α-effect in hydrazine reactivities’

122. Chenevier, P.; Bourel-Bonnet, L.; Roux, D. *J. Am. Chem. Soc.* **2003**, *125*, 16261–16270 ‘Chemical characterization of α-oxohydrazone ligation on colloids: toward grafting molecular addresses onto biological vectors’

123. Egli, S.; Nussbaumer, M. G.; Balasubramanian, V.; Chami, M. ; Bruns, N. ; Palivan, C.; Meier, W. *J. Am. Chem. Soc.* **2011**, *133*, 4476–4483 ‘Biocompatible functionalization of polymersome surfaces: a new approach to surface immobilization and cell targeting using polymersomes’

124. Blanco-Canosa, J. B.; Medintz, I. L; Farrell, D.; Mattoussi, H.; Dawson, P. E. *J. Am. Chem. Soc.* **2010**, *132*, 10027–10033 ‘Rapid covalent ligation of fluorescent peptides to water solubilised quantum dots’

125. Agarwal, P.; Kudirka, R.; Albers, A. E.; Barfield, R. M.; de Hart, G. W.; Drake, P. M.; Jones, L. C.; Rabuka, D. *Bioconjugate Chem.* **2013**, *24*, 846–851 ‘Hydrazino-Pictet-Spengler ligation as a biocompatible method for the generation of stable protein conjugates’

126. Ragnarsson, U. *Chem. Soc. Rev.* **2001**, *30*, 205-213 'Synthetic methodology for alkyl substituted hydrazines'

127. Brosse, N.; Pinto, M.; Jamart-Gregoire, B. *J. Org. Chem.* **2000**, *65*, 4370-4374 'New synthesis of 1,1-substituted hydrazines by alkylation of *N*-acyl- or *N*-alkyloxy-carbonyl aminophthalimide using the Mitsunobu protocol'

128. Jiang, L.; Lu, X.; Zhang, H.; Jiang, Y.; Ma, D. *J. Org. Chem.* **2009**, *74*, 4542–4546 'CuI/4-hydro-L-proline as a more effective catalytic system for coupling of aryl bromides with N-Boc hydrazine and aqueous ammonia'

129. Isaad, J. *Tetrahedron* **2013**, *69*, 2239-2250 'Highly water soluble dyes based on pyrazolone derivatives of carbohydrates'

130. Rodriguez, E. C.; Lisa A. Marcaurelle, L. A.; Bertozzi, C. R. *J. Org. Chem.* **1998**, *63*, 7134-7135 'Aminooxy-, hydrazide-, and thiosemicarbazide-functionalized saccharides: versatile reagents for glycoconjugate synthesis'

131. Davies, B. G.; Fairbanks, A. J. 'Carbohydrate chemistry: OUP chemistry primer no. 99', **2002**, Oxford University Press, Oxford

132. Gao, L.; Hollingsworth, R. I. *Tetrahedron* **2005**, *61*, 3805-3811 'Trihydroxy-2-thiaquinolizidine derivatives as a new class of bicyclic glycosidase inhibitors'

133. Nakane, M.; Hutchinson, C. R.; Gollman, H. *Tetrahedron Lett.* **1980**, *21*, 1213-1216 'A convenient and general synthesis of 5-vinylhexofuranosides from 6-halo-6-deoxypyranosides'

134. Shimizu, Y.; Noshita, M.; Mukai, Y.; Morimoto, H.; Ohshima, T. ***Chem. Commun.* 2014**, ***50***, 12623-12625 'Cleavage of unactivated amide bonds by ammonium salt-accelerated hydrazinolysis'

135. Kudirka, R.; Barfield, R. M.; McFarland, J.; Albers, A. E.; de Hart, G. W.; Drake, P. M.; Holder, P. G.; Banas, S.; Jones, L. C.; Garofalo, A. W.; Rabuka, D. *Chem. Biol.* **2015**, *22*, 293–298 'Generating site-specifically modified proteins *via* a versatile and stable nucleophilic carbon ligation'

136. Yoshikawa, N.; Tan, L.; McWilliams, J. C.; Ramasamy, D.; Sheppard, R. *Org. Lett.* **2010**, *12*, 276–279 'Catalytic enantioselective hydrogenation of *N*-alkoxycarbonyl hydrazones: a practical synthesis of chiral hydrazines'

137. Dirksen, A.; Hackeng, T. M.; Dawson, P. E. *Angew. Chem. Int. Ed.* **2006**, *45*, 7581 –7584 'Nucleophilic catalysis of oxime ligation'

138. Abitbol, T.; Palermo, A.; Moran-Mirabal, J. M.; Cranston, E. D. *Biomacromolecules* **2013**, *14*, 3278–3284 'Fluorescent labeling and characterization of cellulose nanocrystals with varying charge contents'

139. De, M.; Rana, S.; Akpinar, H.; Miranda, O. R.; Arvizo, R. R.; Bunz, U. H. F.; Rotello, V. M. *Nat. Chem.* **2009**, *1*, 461-465 'Sensing of proteins in human serum using conjugates of nanoparticles and green fluorescent protein'

140. Navarro, J. R. G.; Bergstrom, L. *RSC Adv.* **2014**, *4*, 60757 'Labelling of *N*-hydroxysuccinimide-modified rhodamine B on cellulose nanofibrils by the amidation reaction'

141. Dong, S.; Cho, H. J.; Lee, Y. W.; Roman, M. *Biomacromolecules* **2014**, *15*, 1560−1567 'Synthesis and cellular uptake of folic acid-conjugated cellulose nanocrystals for cancer targeting'

142. Navarro, J. R. G.; Conzatti, G.; Yu, Y.; Fall, A. B.; Mathew, R.; Edén, M.; Bergström, L. *Biomacromolecules* **2015**, *16*, 1293−1300 'Multicolor fluorescent labeling of cellulose nanofibrils by click chemistry'

143. Geng, Y.; Paul Dalhaimer, P.; Cai, S.; Tsai, R.; Tewari, M.; Minko, T.; Discher, D. E. *Nat. Nano.* **2007**, *2*, 249-255 'Shape effects of filaments versus spherical particles in flow and drug delivery'

144. van Thor, J. J.; Gensch, T.; Hellingwerf, K. J.; Johnson, L. N. *Nat. Struct. Biol.* **2001**, *9*, 37–41 'Phototransformation of green fluorescent protein with UV and visible light leads to decarboxylation of glutamate 222'

145. Skaanderup, P. R.; Poulsen, C. R.; Hyldtoft, L.; Jørgensen, M. R.; Madsen, R. *Synthesis* **2002**, *12*, 1721–1727 'Regioselective conversion of primary alcohols into iodides in unprotected methyl furanosides and pyranosides regioselctive conversion of primary alcohols in methyl glycosides into iodides'

146. Eskandari, R.; Kuntz, D. A.; Rose, D. R.; Pinto, M. B. *Org. Lett.* **2010**, *12*, 1632-1635 'Potent glucosidase inhibitors: de-*O*-sulfonated ponkoranol and its stereoisomer'

147. Dufresne, A. *Mater. Today* **2003**, *16*, 220-227 'Nanocellulose: a new ageless biomaterial'

148. Neto, W. P. F.; Jean-Luc Putaux, J.; Mariano, M.; Ogawa, Y.; Otaguro, H.; Pasquinia, D.; Dufresne, A. *RSC Advances* **2016**, *1-9* 'Comprehensive morphological and structural investigation of cellulose I and II nanocrystals prepared by sulphuric acid hydrolysis'

149. Lin, N.; Dufresne, A. *Macromolecules* **2013**, *46*, 5570−5583 'Physical and/or chemical compatibilization of extruded cellulose nanocrystal reinforced polystyrene nanocomposites'

150. Segal, L.; Creely, J. J.; Martin, A. A.; Conrad, C. M. *Text. Res. J.* **1959**, *29*, 786–794 'An empirical method for estimating the degree of crystallinity of native cellulose using X-ray diffractometer'

151. Morais, J. P. S.; Rosa, M.; Filho, M. M.; Nascimento, L. D.; Nascimento, D. M.; Cassales, A. B. *Carbohydr. Polym.* **2013**, *91*, 229-235 'Extraction and characterization of nanocellulose structures from raw cotton linter'

152. Maiti, S.; Jayaramudub, J.; Das, K.; Reddy, S. M.; Sadikub, R.; Rayc, S. S.; Liua, D. *Carbohydr. Polym.* **2013**, *98*, 562–567 'Preparation and characterization of nanocellulose with new shape from different precursor'

153. Edwards, J. V.; Fontenot, K. R.; Haldane, D.; Prevost, N. T.; Condon, B. D.; Grimm, C. *Cellulose* **2016**, **23**, 1283–1295 'Human neutrophil elastase peptide sensors conjugated to cellulosic and nanocellulosic materials: part I, synthesis and characterization of fluorescent analogs'

154. Xu, X.; Liu, F.; Jiang, L.; Zhu, J. Y.; Haagenson, D.; Wiesenborn, D. P. *ACS Appl. Mater. Interfaces* **2013**, *5*, 2999−3009 'Cellulose nanocrystals vs. cellulose nanofibrils: a comparative study on their microstructures and effects as polymer reinforcing agents'

155. Klemm, D.; Kramer, F.; Moritz, S.; Lindström, T.; Ankerfors, M.; Gray, D.; Dorris, A. *Angew. Chem. Int. Ed.* **2011**, *50*, 5438-5466 'Nanocelluloses: a new family of nature-based materials'

156. Shi, Z.; Tang, J.; Chen, L.; Yan, C.; Tanvir, S.; Anderson, W. A.; Berryb, R. M.; Tam, K. C. *J. Mater. Chem. B* **2015**, *3*, 603 'Enhanced colloidal stability and antibacterial performance of silver nanoparticles/cellulose nanocrystal hybrids'

157. Habibi, Y.; Goffin, A.; Schiltz, N.; Duquesne, E.; Dubois, P.; Dufresne, A. *J. Mater. Chem.* **2008**, *18*, 5002–5010 'Bionanocomposites based on poly(3-caprolactone)-grafted cellulose nanocrystals by ring-opening polymerization'

158. Azzam, F.; Heux, L.; Putaux, J.; Jean, B. *Biomacromolecules* **2010**, *11*, 3652–3659 'Preparation by grafting onto, characterization, and properties of thermally responsive polymer-decorated cellulose nanocrystals'

159. Araki , J.; Mishima, S. *Molecules* **2015**, *20*, 169-184 'Steric stabilization of "charge-free" cellulose nanowhiskers by grafting of poly(ethylene glycol)'

160. Lin, N.; Huangb, J.; Dufresne, A. *Nanoscale* **2012**, *4*, 3274 'Preparation, properties and applications of polysaccharide nanocrystals in advanced functional nanomaterials: a review'

161. Huang, J.; Chang, P. R.; Lin, N.; Dufresne, A. 'Polysaccharide-based nanocrystals: chemistry and applications',Wiley **2014**

162. Brumer, H.; Zhou, Q.; Baumann, M. J.; Carlsson, K.; Teeri, T. T. *J. Am. Chem. Soc*. **2004**, 126, 5715–5721 'Activation of crystalline cellulose surfaces through the chemoenzymatic modification of xyloglucan'

163. Weishaupt, R.; Siqueira, G.; Schubert, M.; Tingaut, P.; Maniura-Weber, K.; Zimmermann, T.; Thöny-Meyer, L.; Faccio, G.; Ihssen, J. *Biomacromolecules* **2015**, *16*, 3640−3650 'TEMPO-oxidized nanofibrillated cellulose as a high density carrier for bioactive molecules'

164. Gericke, M.; Trygg, J.; Fardim, P. *Chem. Rev.* **2013**, *113*, 4812−4836 'Functional cellulose beads: preparation, characterization, and applications'

165. Ihssen, J.; Reiss, R.; Luchsinger, R.; Thöny-Meyer, L.; Richter, M. *Sci. Rep.* **2015**, *5*, 1-13 'Biochemical properties and yields of diverse bacterial laccase-like multicopper oxidases expressed in *Escherichia coli*'

166. Nakamura, T. *Adv. Exp. Med. Biol.* **1976**, *74*, 408-423 'Oxidation and reduction of copper ions in catalytic reactions of *Rhus* laccase'

167. Wahleithner J. A.; Xu, F.; Brown, K. M.; Brown, S. H.; Golightly, E. J.; Halkier, T.; Kauppinen, S.; Pederson, A.; Schneider, P. *Curr. Genet.* **1996**, *29*, 395-403 'The identification and characterization of four laccases from the plant pathogenic fungus *Rhizoctonia solani*'

168. Froehner, S. C.; Eriksson, K. *J. Bacteriol.* **1974**, *120*, 458–465 'Purification and properties of *Neurospora crassa* laccase'

169. Pointing, S. B.; Pelling, A. L.; Smith, G. J.; Hyde, K. D.; Reddy, C. A. *Mycol. Res.* **2005**, *109*, 115-24 'Screening of basidiomycetes and xylariaceous fungi for lignin peroxidase and laccase gene-specific sequences'

170. Piontek, K.; Antorini, M.; Choinowski, T. *J. Biol. Chem.* **2002**, *277*, 37663-37669 'Crystal structure of a laccase from the fungus *Trametes versicolor* at 1.90-Å resolution containing a full complement of coppers'

171. Givaudan, A.; Effosse, A.; Faure, D.; Potier, P.; Bouillant, M.; Bally, R. *FEMS Microbiol. Lett.* **1993**, *108*, 205-210 'Polyphenol oxidase in *Azospirillum lipoferum* isolated from rice rhizosphere: Evidence for laccase activity in non-motile strains of *Azospirillum lipoferum*'

172. Martins, L. O.; Soares, C. M.; Pereira, M. M.; Teixeira, M.; Costa, T.; Jones, G. H.; Henriques, A. O. *J. Biol. Chem.* **2002**, *277*, 18849-18859 'Molecular and biochemical characterization of a highly stable bacterial laccase that occurs as a structural component of the *Bacillus subtilis* endospore coat'

173. Endo, K.; Hayashi, Y.; Hibi, T.; Hosono, K.; Beppu, T.; Ueda, K. *J. Biochem.* **2003**, *133*, 671–677 'Enzymological characterization of EpoA, a laccase-like phenol oxidase produced by *Streptomyces griseus*'

174. Koschorreck, K.; Richter, S. M.; Ene, A. B.; Roduner, E.; Schmid, R. D.; Urlacher, V. B. *Appl. Microbiol. Biotechnol.* **2008**, *79*, 217-24 'Cloning and characterization of a new laccase from *Bacillus licheniformis* catalyzing dimerization of phenolic acids'

175. Thurston, C. F *Microbiology* **1994**, *140*, 19-26 'The structure and function of fungal laccases'

176. Li, K.; Xu, F.; Eriksson, K. L. *Appl. Environ. Microbiol.* **1999**, *65*, 2654–2660 'Comparison of fungal laccases and redox mediators in oxidation of a nonphenolic lignin model compound'

177. Machczynski, M. C.; Vijgenboom, E.; Samyn, B.; Canters, G. W. *Protein Sci.* 2004, *13*, 2388–2397 'Characterization of SLAC: a small laccase from *Streptomyces coelicolor* with unprecedented activity'

178. Skalova, T.; Dohnalek, J.; Ostergaard, L. H.; Ostergaard, P. R.; Kolenko, P.; Duskova, J.; Stepankova, A.; Hasek, J. *J. Mol. Biol.* **2009**, *385*, 1165-1178 'The structure of the small laccase from *Streptomyces coelicolor* reveals a link between laccases and nitrite reductases'

179. Tepper, A. W. J. W.; Milikisyants, S.; Sottini, S.; Vijgenboom, E.; Groenen, E. J. J.; Canters, G. W. *J. Am. Chem. Soc.* **2009**, *131*, 11680–11682 'Identification of a radical intermediate in the enzymatic reduction of oxygen by a small laccase'

180. Gupta, A.; Nederlof, I.; Sottini, S.; Tepper, A. W. J. W.; Groenen, E. J. J.; Thomassen, E. A. J.; Canters, G. W. *J. Am. Chem. Soc.* **2012**, *134*, 18213−18216 'Involvement of Tyr108 in the enzyme mechanism of the small laccase from *Streptomyces coelicolor*'

181. Jones, S. M., Solomon, E. I. *Cell Mol. Life. Sci.* **2015**, *72*, 869–883 'Electron transfer and reaction mechanism of laccases'

182. Jansson, B.; Uhlén, M.; Nygren, P. A. *FEMS Immunol. Med. Microbiol.* **1998**, *20*, 69-78 'All individual domains of staphylococcal protein A show Fab binding'

183. Foster, T. J.; Geoghegan, J. A.; Ganesh, V. K.; Höök, M. *Nat. Rev. Microbiol.* **2014**, 12, 49-62 'Adhesion, invasion and evasion: the many functions of the surface proteins of *Staphylococcus aureus*'

184. Falugi, F.; Kim, H. K.; Missiakas, D. M.; Schneewind, O. *mBio* **2013**, *4*, 1-9 'Role of protein A in the evasion of host adaptive immune responses by *Staphylococcus aureus*'

185. Forsgren, A.; Nordström, K. *Ann. Acad. Sci.* **1974**, *236*, 252-266 'Protein A from *Staphylococcus aureus*: the biological significance of its reaction with IgG'

186. Deis, L. N.; Wu, Q.; Wang, Y.; Qi, Y.; Daniels, K. G.; Zhou, P.; Oas, T. G. *Proc. Natl. Acad. Sci.* **2015**, *112*, 9028-9033 'Suppression of conformational heterogeneity at a protein-protein interface'

187. Gómez, M. I.; Lee, A.; Reddy, B.; Muir, A.; Soong, G.; Pitt, A.; Cheung, A.; Prince, A. *Nat. Med.* **2004**, *10*, 842 – 848 '*Staphylococcus aureus* protein A induces airway epithelial inflammatory responses by activating TNFR1'

188. Graille, M.; Stura, E. A.; Corper, A. L.; Sutton, B. J.; Taussig, M. J.; Charbonnier, J. B.; Silverman, G. J. *Proc. Natl. Acad. Sci.* **2000**, *97*, 5399-5404 'Crystal structure of a *Staphylococcus aureus* protein A domain complexed with the Fab fragment of a human IgM antibody: structural basis for recognition of B-cell receptors and superantigen activity'

189. Tajima, N.; Takai, M.; Ishihara, K. *Anal. Chem.* **2011**, *83*, 1969–1976 'Significance of antibody orientation unraveled: well-oriented antibodies recorded high binding affinity'

190. Seo, J.; Lee, S.; Poulter, C. D. *J. Am. Chem. Soc.* **2013**, *135*, 8973−8980 'Regioselective covalent immobilization of recombinant antibody-binding proteins A, G, and L for construction of antibody arrays'

191. Labadie, G. R.; Viswanathan, R.; Poulter, C. D. *J. Org. Chem.* **2007**, *72*, 9291-9297 'Farnesyl diphosphate analogues with ω-bioorthogonal azide and alkyne functional groups for protein farnesyl transferase-catalyzed ligation reactions'

192. Toscano, M. D.; De Maria, L.; Lobedanz, S.; Østergaard, L. H. *ChemBioChem* **2013**, *14*, 1209 – 1211 'Optimization of a small laccase by active-site redesign'

193. Kobayashi, N.; Urasawa, S.; Uehara, N.; Watanabe, N. *Epidemiol. Infect.* **1999**, *122*, 241-249 'Analysis of genomic diversity within the Xr-region of the protein A gene in clinical isolates of *Staphylococcus aureus*'

194. Chen, L.; Zou, M.; Hong, F. F. *Front. Microbiol.* **2015**, *6*, 1-8 'Evaluation of fungal laccase immobilized on natural nanostructured bacterial cellulose'

195. Sampaioa, L. M. P.; Padrão, J.; Fariaa, J.; Silva, J. P.; Silvac, C. J.; Dourado, F.; Zillea, A. *Carbohydr. Polym.* **2016**, *145*, 1–12 'Laccase immobilization on bacterial nanocellulose membranes: antimicrobial, kinetic and stability properties'

196. Wheeldon, I. R.; Gallaway, J. W.; Barton, S. C.; Banta, S. *Proc. Natl. Acad. Sci.* **2008**, *105*, 15275-15280 'Bioelectrocatalytic hydrogels from electronconducting metallopolypeptides coassembled with bifunctional enzymatic building blocks'

197. Abdel-Mohsen, H. T.; Sudheendran, K.; Conrad, J.; Beifuss, U. *Green Chem.* **2013**, *15*, 1490-1495 'Synthesis of disulfides by laccase-catalyzed oxidative coupling of heterocyclic thiols'

198. Fox, S. C.; Edgar, K. J. *Biomacromolecules* **2012**, *13*, 992−1001 'Staudinger reduction chemistry of cellulose: synthesis of selectively O-acylated 6-amino-6-deoxy-cellulose'

199. Agarwal, P.; Weijden, J.; Sletten, E. M.; Rabukab, D.; Bertozzi, C. R. *Proc. Natl. Acad. Sci.* **2013**, *110*, 46-51 'A Pictet-Spengler ligation for protein chemical modification'

200. Welch, J. T.; Seper, K. W. *J. Org. Chem.* **1988**, *53*, 2991-2999 'Synthesis, regioselective deprotonation, and stereoselective alkylation of fluoro ketimines'

201. Sampaioa, L. M. P.; Padrão, J.; Fariaa, J.; Silva, J. P.; Silvac, C. J.; Dourado, F.; Zillea, A. *Carbohydr. Polym.* **2016**, *145*, 1–12 'Laccase immobilization on bacterial nanocellulose membranes: antimicrobial, kinetic and stability properties'

202. Zang, B.; Ren, J.; Xu, L.; Jia, L. *J. Chrom. B* **2016**, *1008*, 132-138 'Direct site-specific immobilization of protein A *via* aldehyde-hydrazide conjugation'

Protein sequences discussed in this thesis which could be produced in *E. coli*

Ald-GFP

GTPLCGPSRAHMGSGASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTL
VTTLCYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGN
ILGHKLEYNYNSHNVYIMADKQKNGIKVNFKTRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKD
PNEKRDHMALLEFVTAAGITHGMDELYKLEHHHHHH

Calculated mass: 29,270 Da; found: 29,250 Da

Loop1-GFP

GSSHHHHHHSSGLVPRGSHIHMGSGASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTT
GKLPVPWPTLVTTLCYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIEL
KGIDFKEDGNILGHKLEYNYNSHNVYIMADKGTPLCGPSRNGIKVNFKTRHNIEDGSVQLADHYQQNTPIGDGPV
LLPDNHYLSTQSALSKDPNEKRDHMALLEFVTAAGITHGMDELY

Calculated mass: 29,895 Da; found: 29,874 Da

Loop2-GFP

GSSHHHHHHSSGLVPRGSHIHMGSGASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTT
GKLPVPWPTLVTTLCYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIEL
KGIDFKEDGNILGHKLEYNYNSHNVYIMADKAGKGTPLCGPSRAGNGIKVNFKTRHNIEDGSVQLADHYQQNTPI
GDGPVLLPDNHYLSTQSALSKDPNEKRDHMALLEFVTAAGITHGMDELY

Calculated mass: 30,280 Da; found: 30,259 Da

Ald-YLS

GTPLCGPSRAHMAVKGLGKPDQVYDGSKIRVGIIHARWNRVIIDALVKGAIERMVSLGVEEKNIIIETVPGSYEL
PWGTKRFVDRQAKLGKPLDVVIPIGVLIKGSTMHFEYISDSTTHALMNLQEKVDMPVIFGLLTCMTEEQALARAG
IDEAHSMHNHGEDWGAAAVEMAVKFGKNAFLEHHHHHH

Calculated mass: 20,740 Da; found: 20,737 Da

Ald-ALS

GTPLCGPSRAHMEIYEGKLTAEGLRFGIVASRFNHALVDRLVEGAIDCIVRHGGREEDITLVRVPGSWEIPVAAG
ELARKEDIDAVIAIGVLIRGATPHFDYIASEVSKGLANLSLELRKPITFGVITADTLEQAIERAGTKHGNKGWEA
ALSAIEMANLFKSLRLEHHHHHH

Calculated mass: 18,859 Da; found: 18,845 Da and 18,889 Da

SLAC

GSSHHHHHHSSGLVPRGSHIHMSARTAPAGGEVRHLKMYAEKLADGQMGYGFEKGKASVPGPLIEVNEGDTLHIE
FTNTMDVRASLHVHGLDYEISSDGTAMNKSDVEPGGTRTYTWRTHKPGRRDDGTWRPGSAGYWHYHDHVVGTEHG
TGGIRNGLYGPVIVRRKGDVLPDATHTIVFNDMTINNRKPHTGPDFEATVGDRVEIVMITHGEYYHTFHMHGHRW
ADNRTGILTGPDDPSRVIDNKITGPADSFGFQIIAGEGVGAGAWMYHCHVQSHSDMGMVGLFLVKKPDGTIPGYE
PHEHGGATAKSGESGEPTGGAAAHEHEHLEI

Calculated mass: 36,011 Da; found: 36,010 Da

Ald-SLAC

GTPLCGPSRAHMSARTAPAGGEVRHLKMYAEKLADGQMGYGFEKGKASVPGPLIEVNEGDTLHIEFTNTMDVRAS
LHVHGLDYEISSDGTAMNKSDVEPGGTRTYTWRTHKPGRRDDGTWRPGSAGYWHYHDHVVGTEHGTGGIRNGLYG
PVIVRRKGDVLPDATHTIVFNDMTINNRKPHTGPDFEATVGDRVEIVMITHGEYYHTFHMHGHRWADNRTGILTG
PDDPSRVIDNKITGPADSFGFQIIAGEGVGAGAWMYHCHVQSHSDMGMVGLFLVKKPDGTIPGYEPHEHGGATAK
SGESGEPTGGAAAHEHEHLEIKLEHHHHHH

Calculated mass: 35,999 Da; found: 36,000 Da and 36,040 Da

Ald-protein A

GTPLCGPSRAHMAQHDEAQQNAFYQVLNMPNLNADQRNGFIQSLKDDPSQSANVLGEAQKLNDSQAPKADAQQNN
FNKDQQSAFYEILNMPNLNEAQRNGFIQSLKDDPSQSTNVLGEAKKLNESQAPKADNNFNKEQQNAFYEILNMPN
LNEEQRNGFIQSLKDDPSQSANLLSEAKKLNESQAPKADNKFNKEQQNAFYEILHLPNLNEEQRNGFIQSLKDDP
SQSANLLAEAKKLNDAQAPKADNKFNKEQQNAFYEILHLPNLTEEQRNGFIQSLKDDPSVSKEILAEAKKLNDAQ
APKEEDNLEHHHHHH

Calculated mass: 35,579 Da; found: 35,620 Da

Ald-FGE-4C

GTPLCGPSRAHMPSFDFDIPRRSPQEIAKGMVAIPGGTFRMGGEDPDAFPEDGEGPVRTVRLSPFLIDRYAVSNR
QFAAFVKATGYVTDAERYGWSFVFHAHVAPGTPVMDAVVPEAPWWVAVPGAYWKAPEGPGSSITDRPNHPVVHVS
WNDAVAYATWAGKRLPTEAEWEMAARGGLDQARYPWGNELTPRGRHRANIWQGTFPVHDTGEDGYTGTAPVNAFA
PNGYGLYNVAGNVWEWAADWWSADWHATESPATRIDPRGPETGTARVTKGGSFLCHESYCNRYRVAARTSNTPDS
SAAHTGFRAAADPLKLEHHHHHH

Calculated mass: 35,434 Da; found: 35,433 Da

Aknowledgement

I would like to thank to all the people who helped me during my PhD thesis with scientific advice and friendship. I am thankful to NRP66 for financing the project.

I am grateful to my PhD supervisor, Prof. Florian Seebeck for helpful advice and the freedom to read and develop my project. I benefited greatly from his knowledge in protein chemistry and creativity.

I would like to thank Prof. Dennis Gillingham for accepting the co-examination of my PhD thesis and for the nice discussions. Many thanks to Prof. Christof Sparr for co-chairing the PhD defence and providing enjoyable teaching labs.

Furthermore, I would like to thank Dr. Philippe Tingaut for providing the nanocellulose, performing the analytical analysis of the modified crystals and useful advice and motivation.

I am very grateful to Dr. Valentin Köhler for general advices and for always being there to help. I am also very thankful to Matthias Knop for sharing his enzyme, teaching me from the beginning how to work with proteins and good scientific advice.

I would like to express my gratitude to previous supervisors and teachers who supported me throughout my education Prof. Richard Brown, Dr. Samuel Watkins, Dr. Thomas Logothesis, Dr. Jeremy Hinks, Dr. Eugen Stulz and Nicoleta-Dana Ionita.

I am very thankful to all my lab mates for being such a good team and for the great moments we had. Special thanks to Dr. Jian Gao for help with peptides, Dr. Gabriel Mashabela for his kindness and motivation and Dr. Sebastien Coyne for very good tips in molecular biology and bringing scientific spirit in the lab.

Many thanks to Sascha for his support when I was going through difficult moments.

To Maurus for bringing colour and the sound of violin into my life.

To mom, dad and Diana for they unlimited love and support.

Curriculum Vitae
Roxana Lemnaru

Nationality: Romanian
Date of birth: 20.04.1989

Education

08/2012 – Present	PhD thesis in Chemistry University of Basel, Switzerland
2012	MChem in Chemistry University of Southampton, United Kingdom
2008	High-school in mathematics and computer science CNVA, Romania

Research experience

08/2012 – Present	PhD thesis: 'A chemoenzymatic strategy for protein-nanocellulose conjugates' Laboratory of Prof. F. P. Seebeck, University of Basel
2011	Master thesis: 'En route to tryptamine and tryptophols' Malaria project, Novartis
2010	Bachelor thesis: 'Synthesis of flavonoids with bactericidal activity' Laboratory of Prof. R. Brown, University of Southampton

Conferences

2016	International symposium in chemical biology, Geneva Winning prize - travel grant award
2015	International nanoscience student conference, Basel Best talk award – 2nd place (out of 26)
2012	MChem symposium, Southampton Best poster award – 2nd place (out of 41)