Exploring a unique pocket in Galectin-8N by the synthesis of D-galactoside-based benzimidazole inhibitors

Master Thesis project in organic chemistry

Author: Sid Bartold

Supervisor: Fredrik Sjövall

Examiner: Ulf Nilsson

June 2023

Lund

Table of contents

Abstract

Galectins are a family of glycan-binding proteins where galectin-8 is a part of the subgroup known as tandem-repeat type galectins. Galectin-8 is composed of two homologous CRDs, one N- and one C-terminal CRD. As it stands now, galectin-8N has been found to bind to a larger number of glycans than galectin-8C. Thus, it is currently of a higher therapeutical interest to focus research on the N-terminal CRD of galectin-8 than on the C-terminal. Studies on galectin-8N inhibitors have revealed that benzimidazole-galactoside derivates show a greater affinity and selectivity for galectin-8N by binding to unique amino acids, namely Arg45. However, recent research has shown that there are two more unique amino acids, Arg59 and Tyr141, existing in a pocket of galectin-8N. It was found that this pocket and its unique amino acids can be exploited by placing substituents on the C-4 position of the benzimidazole leading to a galectin-8N inhibitor with a higher affinity and selectivity, which was the aim of this thesis. Ten inhibitors were synthesized in six synthesis steps starting with an equatorial chlorination at C-1 of β-D-galactose pentaacetate, followed by an S_N2 reaction yielding an α-thioglycoside. A Zemplén deacetylation was then performed following a regioselective 3-O-monoalkylation with methyl bromoacetate as the reagent, which was then cyclized into a benzimidazole. For the final step, Suzuki-couplings were executed at C-4 of the benzimidazole with a *p*-carboxyphenyl substituent having the highest affinity for galectin-8N with K_d of 11 ± 1 µM of the ten synthesized inhibitors. Whilst none of the synthesized compounds managed to surpass current more potent galectin-8N inhibitors, it highlighted the importance of a substituent at the C-6 position of the benzimidazole, which was omitted in this project. However, the importance of the substituent at C-4 should not be ignored considering the synthesized compounds differ by a factor of a 100 with the *p*-carboxyphenyl substituent revealing promising results for further exploration of the pocket.

Popular Science Summary

Galectins are a group of proteins that bind strongly to galactosides, which is a galactose, one type of sugar, that is connected or bound to another organic compound, which in many cases can be another sugar. Many different galectins have been discovered inside and throughout our bodies which has led to a lot of scientific and medicinal interest in galectins, and their function. What has been found is that they play a part in a wide range of biological processes where they are implicated in many diseases, such as diabetes, cancer, HIV, arthritis, and many more. Galectin-8 (abbreviated as Gal-8), as the name suggests, is a member of this protein group that is involved in multiple of diseases with ongoing research on its role in fibrosis, corneal inflammation, organ graft rejection, various cancers such as breast cancer, and inflammatory and immune diseases. Being implicated in such a plethora of disease makes it a highly interesting target for drug discovery and development.

By blocking the protein, we can stop (inhibit) the protein from carrying out its function leading to a cure or a reduction of the disease. This can be done by developing an inhibitor, which for the case of galectins, would be a drug that mimics the galactosides, "fooling" the galectin to bind to it instead of the natural galactoside, inhibiting the function of the protein. A key issue in drug development is called selectivity and it is the "promiscuity" of the drug to want to bind to other proteins other than the intended one which is something that often times should be avoided to prevent or minimize side effects of the drug. Selectivity is currently a big issue for drugs targeting Gal-8 inhibition, where the selectivity is quite poor meaning that the synthesized inhibitor is also binding to other galectins with similar affinities. This gives rise to another problem in drug development – affinity. Affinity describes how well the drug binds to the target and is defined by the dissociation constant, K_D, expressed in molars, M.

The best current Gal-8 inhibitors show good affinity, but not yet good enough, with a poor selectivity in regard to other galectins, which brings us to the aim of this thesis. Recent research on Gal-8 revealed that it has amino acids in its binding pocket different from those in other galectins. Designing, synthesizing, and testing inhibitors that bind to these amino acids would assist in solving both the selectivity and affinity issues. This was done by a combination of first studying the interactions between the protein and the inhibitor in computer models, followed by the synthesis, and testing of the compounds. A total of ten inhibitors were made with the best showing an affinity slightly worse than current Gal-8 inhibitors, but insight was gained regarding the development of future Gal-8 inhibitors. Thus, there is still some way to go before a Gal-8 inhibitor will reach the market. However, this project shone a light, albeit a small and dim one, illuminating a part of the path to the secret of potent galectin-8 inhibitors.

Abbreviations

ALCAM	Activated leukocyte cell adhesion molecule
Arg	Arginine
Asn	Asparagine
CRD	Carbohydrate recognition domain
DCM	Dichloromethane
EDG	Electron donating group
EtOAc	Ethyl acetate
EWG	Electron withdrawing group
FAK	Focal adhesion kinase
FP	Fluorescence polarization
Gal	Galectin
Gal-8N	Galectin-8 N-terminal domain
Gln	Glutamine
Glu	Glutamic acid
His	Histidine
HIV	Human immunodeficiency virus
HPLC	High performance liquid chromatography
HRMS	High-resolution mass spectrometry
ITC	Isothermal titration calorimetry
LacNAc	N-Acetyllactosamine
LCMS	Liquid chromatography–mass spectrometry
MD	Molecular dynamics
NMR	Nuclear Magnetic Resonance
TBAB	Tetrabutylammonium bromide
TGFβ	Transforming growth factor beta
TLC	Thin-layer chromatography
Treg	Regulatory T cell
Trp	Tryptophan
Tyr	Tyrosine
VEGF-C	Vascular endothelial growth factor C

Introduction

Galectins

Galectins overview

Galectins are a family of glycan-binding soluble proteins or lectins with a high affinity for β-galactosides and notable sequence similarity in the carbohydrate recognition domains (CRD) [1]. Currently, at least 15 members have been found in mammals [2], and they are divided into three subfamilies containing either one or two CRDs. Members of the first subfamily can be referred to as the prototype galectins and include galectin-1 (Gal-1), Gal-2, Gal-5, Gal-7, Gal-10, Gal-11, Gal-13, Gal-14, and Gal-15. Prototype galectins contain only one CRD and function either as monomers or homodimers. The second subfamily is referred to as chimeric galectins, where Gal-3 is the current sole member. It is composed of one C-terminal CRD fused to a 120-160 long amino acid non-lectin N-terminal domain. The third subfamily is known as the tandem-repeat type galectins (Gal-4, Gal-6, Gal-8, Gal-9, and Gal-12) which contain two CRDs with different specificity, connected by a flexible linker peptide [3], [4]. The structural classification of galectins is represented in **Figure 1**.

a. Prototype galectins
Gal-1, -2, -5, -7, -10, -11, -13, -14, -15

b. Chimera-type galectins
Gal-3

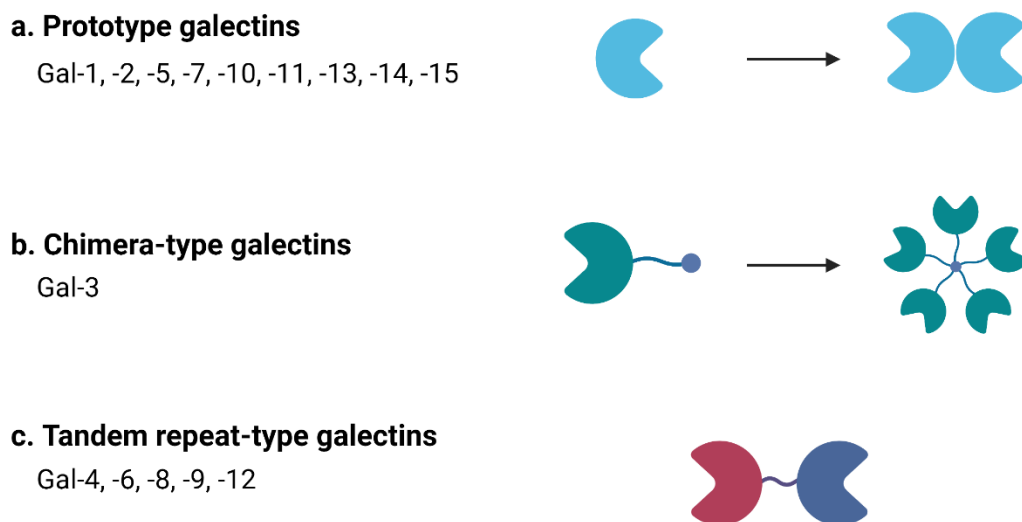**c. Tandem repeat-type galectins**
Gal-4, -6, -8, -9, -12

*Figure 1: Schematic representation of the structure of the three galectin subfamilies; **a**, prototype galectins which can dimerize via non-covalent interactions; **b**, chimera-type galectin-3 which can oligomerize via interactions with its tail domain and **c**, Tandem-repeat galectins, containing two CRDs linked by a linker region.*

Galectin CRD structure

The CRD of galectins is composed of around 135 amino acids folding as a bent β-sandwich [5], formed by two sheets, a concave six-strand sheet (**Figure 2**, labeled S1-S6) and a convex five-strand sheet (F1-F5). Due to the concave shape of the six-stranded sheet, a groove exists where the conserved carbohydrate-binding site can be found [6]. This binding site has five subsites labeled A through E with subsite C, composed of seven amino acids, being the defining β-galactoside binding site [7] (**Figure 3c**).
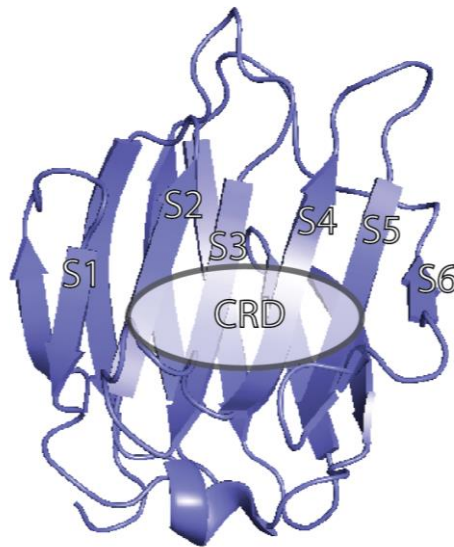
Figure 2: Alignment of galectin tertiary structure (Gal-8N, PDB code 7AEN), with β-sheets S1-S6 and the CRD represented in the concave groove.

Roles and function

Galectins are soluble proteins that are synthesized in the cytosol but may be translocated to the nucleus to interact with nuclear ligands. Galectins can also be secreted to the extracellular milieu via nonclassical pathways to interact with extracellular matrix components. In the extracellular milieu, galectins often stimulate the crosslinking of their glycosylated receptor binding partners including transmembrane protein receptors and glycolipids, leading to the formation of dynamic lattices. By forming these crosslinked microdomains, galectins greatly influence the membrane dynamics of their interacting partners, in terms of signal transduction and membrane recycling [8]–[10].

Galectins serve a multitude of biological roles including cell adhesion, since galectins bind to the β-galactosides of glycoproteins on both the cell surface and in extracellular matrices [11]; immune modulation, where they can influence cell-cell and cell-matrix interactions with

inhibitory or stimulating effects [8]; apoptosis, where galectins can have either a pro- or anti-apoptotic effect, *i.e.*, galectin-3 has been found to have an anti-apoptotic effect through the activation of a cysteine protease-dependent pathway [12] while galectin-7 has a pro-apoptotic effect by up-regulating and intensifying JFK activation, thus releasing cytochrome *c*, inducing apoptosis [13].

These all coalesce into the galectin family being implicated in a wide range of diseases such as cardiovascular disorders [14], neurodegeneration [15], cancer [16], [17], HIV [18], and many more [19]. Despite great advances having been made in recent years in the field of galectin inhibition, with several compounds entering clinical trials [20], there is a continuous need and a vested interest to develop novel, more efficient drugs targeting galectins.

Galectin-8

Distribution and function
The biodistribution of Gal-8 is broad and includes organs such as the liver, the kidney, the cardiac muscle, the lung, and the brain [21]. While Gal-8 is widely expressed in normal tissues, its expression is markedly affected in certain pathologies, such as in inflamed tissues like the cornea [22] or in cancer including prostate, lung, kidney, bladder, and breast cancer [21], and has been proposed as prognostic [23], [24] and diagnostic markers [25] in various cancer types.

Due to its glycan-binding capacity, Gal-8 is involved in fine-tuning a vast range of cellular processes. Among others, Gal-8 ligand partners include glycosylated receptors, adhesion molecules, and intracellular proteins [1], [26]. Consequently, Gal-8-mediated effects span from extracellular cell adhesion regulation to intracellular autophagolysosomal pathway modulation.

In the extracellular milieu, Gal-8 binds to matrix proteins such as integrins and fibronectin and is involved in cell signaling via e.g., the activation of focal adhesion kinase (FAK) [27] and cytoskeleton rearranging, affecting cell adhesion and migration. By binding to vascular endothelial growth factor-C (VEGF-C), Gal-8 regulates its growth factor signaling ability, promoting pathological lymph vessel formation, and influencing pathological processes such as cancer, organ graft rejection, or corneal inflammation [22]. By knocking down the expression of Gal-8 or its endogenous ligand activated leukocyte cell adhesion molecule (ALCAM), in breast cancer cells in vitro and in vivo, cell migration and proliferation were decreased, suggesting Gal-8 as a promising target for breast cancer [28]. Importantly, Gal-8 is

also broadly involved in immunomodulatory processes. Through the activation of TGFβ signaling, Gal-8 triggers the differentiation of highly suppressive Treg cells, leading to the worsening of inflammatory and autoimmune diseases [29]. Interestingly, a pivotal role for Gal-8 in regulating autophagy has recently been established. Through the binding to bacterial glycans in damaged endosomes, Gal-8 triggers protective autophagy, avoiding bacterial infections [30]. Similarly, Gal-8-mediated mechanisms may take part in protective autophagy against viral infections [31], [32]. Additionally, by its ability to bind damaged endosomes, Gal-8 has been explored as a robust research tool to assess endosomal damage of various therapeutics including siRNA [33].

Structure

Gal-8 is a 35 kDa tandem-repeat type galectin composed of two homologous CRDs, *i.e.*, one N-terminal and one C-terminal CRD, with approximately 140 amino acids each, and connected by a unique ~30 amino acid long peptide link [34]. The two CRDs share a 38% homology [35] where each domain has been shown to exhibit affinity for different glycans. The CRD of the N-terminal domain (Gal-8N) exhibits a strong and unique binding capacity to 3'-O-sulfated or 3'-O-sialylated glycans [36] while the C-terminal CRD recognizes blood group antigens and poly-N-acetyllactosamine (LacNac) glycans [37]. Hence, gal-8N has been proposed to bind to a broader range of β-D-galactopyranosides and with higher affinity than the C-terminal domain.

Besides their similarities, certain major differences in the CRD of Gal-8N can be leveraged to specifically target it. In addition to the shared amino acids in both CRDs i.e. His65, Asn67, Arg69, Asn79, Trp86, and Glu89, Gal-8N presents specific amino acids, namely Arg45, Gln47, Arg59, and Tyr141, with Arg59 being especially important for the previously mentioned Gal-8N preferential binding to 3'-O-sulfated or 3'-O-sialylated glycans [36].

Structure-functional analysis has shown that both the C- and N-terminal CRDs are needed for Gal-8 functionality [38], Hence, the inhibition of one of them is sufficient to abrogate its biological functions. The aforementioned Gal-8N characteristics motivated us to focus on the design and synthesis of inhibitors specifically targeting Gal-8N.
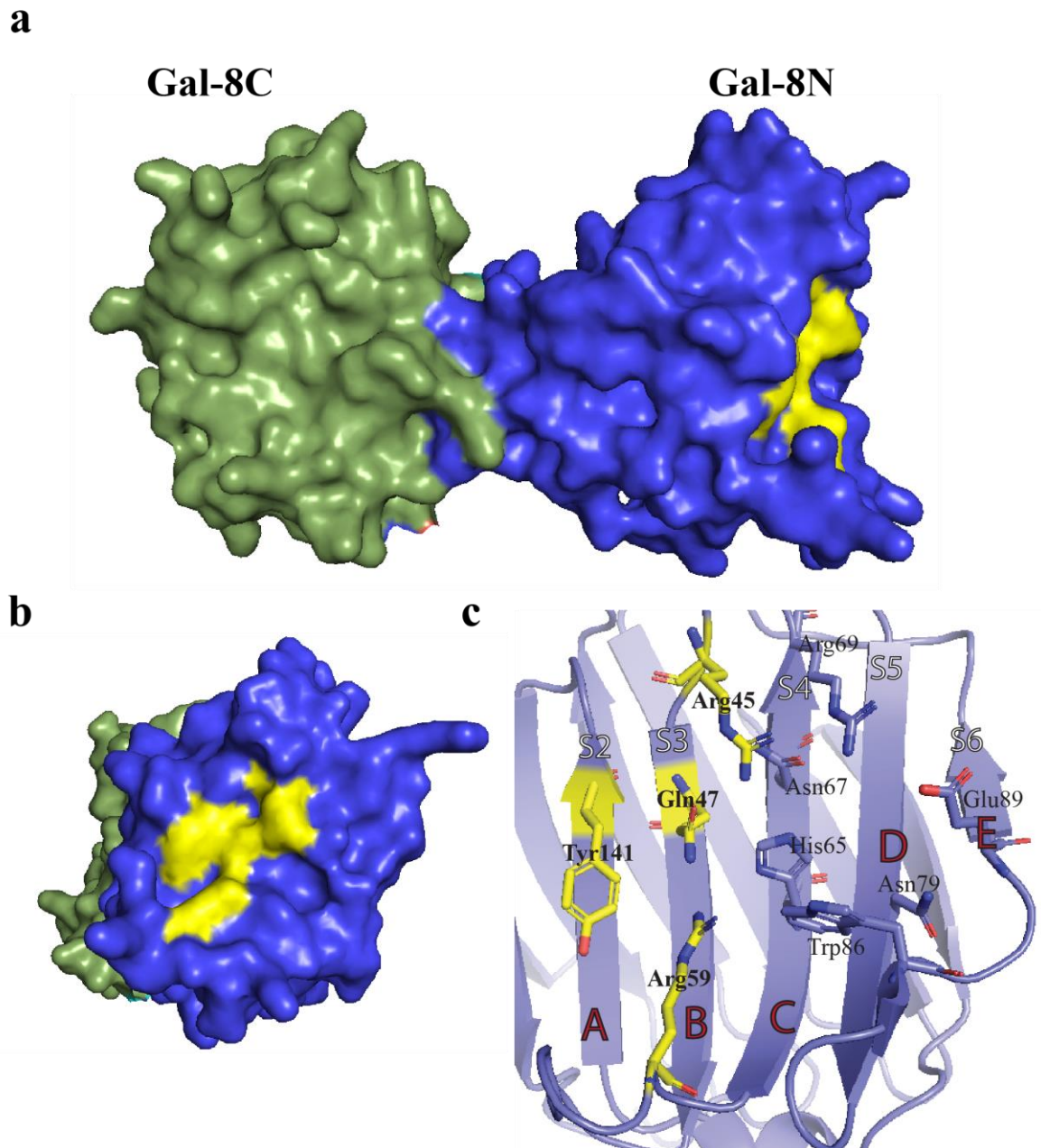
Figure 3. a, *Galectin-8 electron density map, with the C-terminal domain (green) and N-terminal domain (Gal-8N, purple) joined by the linker region (Gal-8, PDB code 3VKL).* ***b***, *Unique amino acids of Gal-8N are colored in yellow, highlighting a groove in the CRD and the pocket between Thr141 and Arg59.* ***c***, *Alignment of galectin tertiary structure (Gal-8N, PDB code 7AEN), indicating β-sheets S1-S6 and CRD binding subsites A-E. CRD residues are highlighted and represented with sticks (i.e., His65, Asn67, Arg69, Asn79, Trp86 and Glu89) (carbon in lilac, oxygen in red, nitrogen in blue), with unique Gal-8N residues colored in yellow (i.e. Arg45, Gln47, Arg59 and Tyr141).*

Methods

Molecular Dynamics

Computer simulations using molecular dynamics (MD) is a versatile tool in the toolbox of a chemist and a biologist, allowing one to view the movement of atoms and molecules, and their interaction in biomolecular processes. Protein folding, ligand binding, and conformational change can be simulated to get an idea of *e.g.*, which inhibitor to synthesize in drug design. MD is a mathematical model using Newton's law of motion to predict the movement and position of each atom due to the force exerted on each atom with the force calculated using a force field model. There are obvious limitations to this model however, the force field model often uses Coulombic interactions between the atoms, leading to an approximation of the movement of the atoms. Another limitation is that the formation and breaking of covalent bonds is typically not simulated [39].

Fluorescence Polarization

Fluorescence polarization (FP) is a popular and versatile method to evaluate *e.g.*, binding interactions between hormone-receptor, antigen-antibody, and ligand-protein. For the FP-method used in this thesis, a fluorophore (fluorescein being a popular one, and the one used here) probe is excited with plane-polarized light and the degree of polarization is measured by two photomultipliers, giving the value of FP [40]. The degree of polarization of the fluorophore is inversely proportional to its molecular rotation, hence, when the probe is bound to the protein it will rotate slower compared to when it is free. Thus, the K_D of the inhibitor can be calculated through the percent bound/free species being linearly proportional to FP [41]. Isothermal titration calorimetry (ITC) is another method to study binding data between ligand-protein by measuring the difference of heat when binding occurs. Comparing FP and ITC, FP is considerably cheaper to use, utilizing cheaper equipment and requiring less of both reagents and protein. Another advantage of FP is the ease of use and interpretating of the data since all the concentrations of the components are known [42].

Results and Discussion

Design

While affinity is important in drug development, selectivity is key. The unique characteristics of galectin-8N CRD make it a more promising target for inhibition compared to Gal-8C. Importantly, recent research has revealed another unique feature of Gal-8N. Namely, a unique perpendicular pocket where the amino acid Tyr141 has been found binding to ligands along with Arg59 yielding highly selective Gal-8N inhibitors [43], [44]. This, together with the synthesized compound **1** (**Figure 4**) [45] forms the basis of this project.
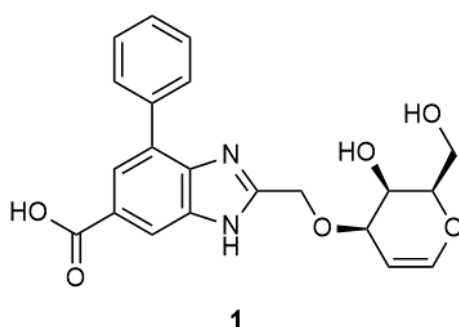
Figure 4. Galectin-8N ligand. $K_d = 5 \pm 0.5\ \mu M$

Thus, the aim of this master's thesis is to synthesize and test the Gal-8N inhibition of different D-galactoside-based benzimidazoles with various R-groups on the C-4 position of the benzimidazole (**Figure 5**) to explore the pocket of Gal-8N.

While it has been found that the carboxylic acid at similar positions as in compound **1** imparts a greater affinity and selectivity to Gal-8N through a water-mediated hydrogen bond to Gly142 [46], it is not an optimal moiety for drugs due to its high polarity [47]. Hence the carboxylic acid on the C-6 position of the benzimidazole in compound **1** has been omitted in compound **2**.
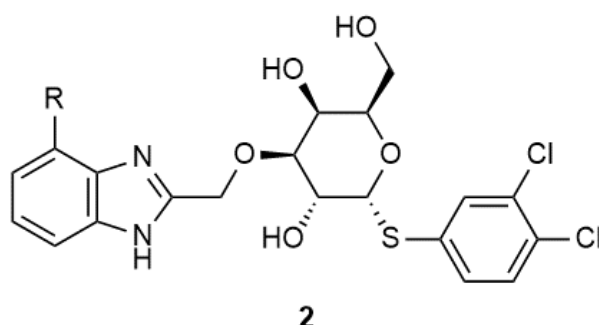
Figure 5. Benzimidazole derivatives designed in this thesis (R=substituents).

MD Simulations

A phenyl substituent at the C-4 position of benzimidazole conferred the lowest K_D (**Figure 4**) [45], thus building upon that, and extending the phenyl group with further substituents was hypothesized to improve affinity of novel compounds even further. Several MD simulations were performed, with different substituents at the C-4 position of benzimidazole, using Desmond implemented in Schrödinger Maestro (2021-1) with 1 kcal/mol constraints placed on the O-4 atom of the galactose and on the stranded backbone atoms with a simulation time ranging from 10 to 200 ns. In **Figure 6**, the *p*-carboxyphenyl substituent establishes noteworthy interactions, namely the π-cation interaction between Arg59 and the phenyl group of the substituent. Likewise, the same phenyl group establishes a π-π interaction with Tyr141, along with a water-mediated hydrogen bond between the hydroxy group of the *p*-carboxyphenyl and Ser55.
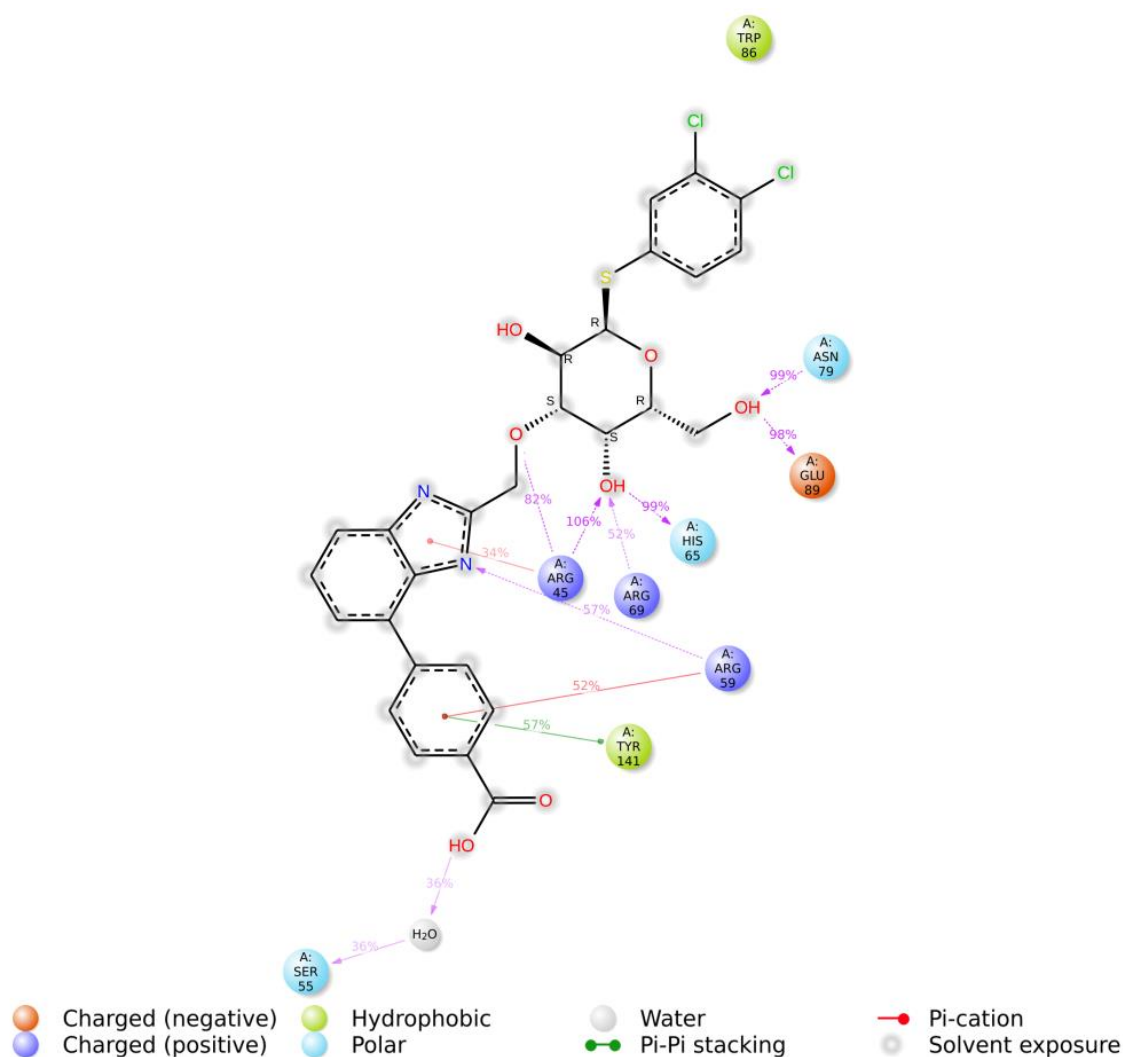
Figure 6. *10 ns simulation time of Ligand-Protein Contacts between D-galactoside benzimidazole derivate with a p-carboxyphenyl substituent at the C-4 of the benzimidazole and Gal-8N. Interactions that occur more than 30.0% of the simulation time are shown.*

The predicted interaction with Ser55 was interesting to further elaborate on, which eventually led to the pyrazole substituent seen in **Figure 7,** forming almost twice as frequent the Arg59 contacts compared to the ligand in **Figure 6** and a direct hydrogen bond between Ser55 and the nitrogen atom of the pyrazole. However, Arg45 diminishes its π-cation interaction with the benzimidazole bicyclic ring.
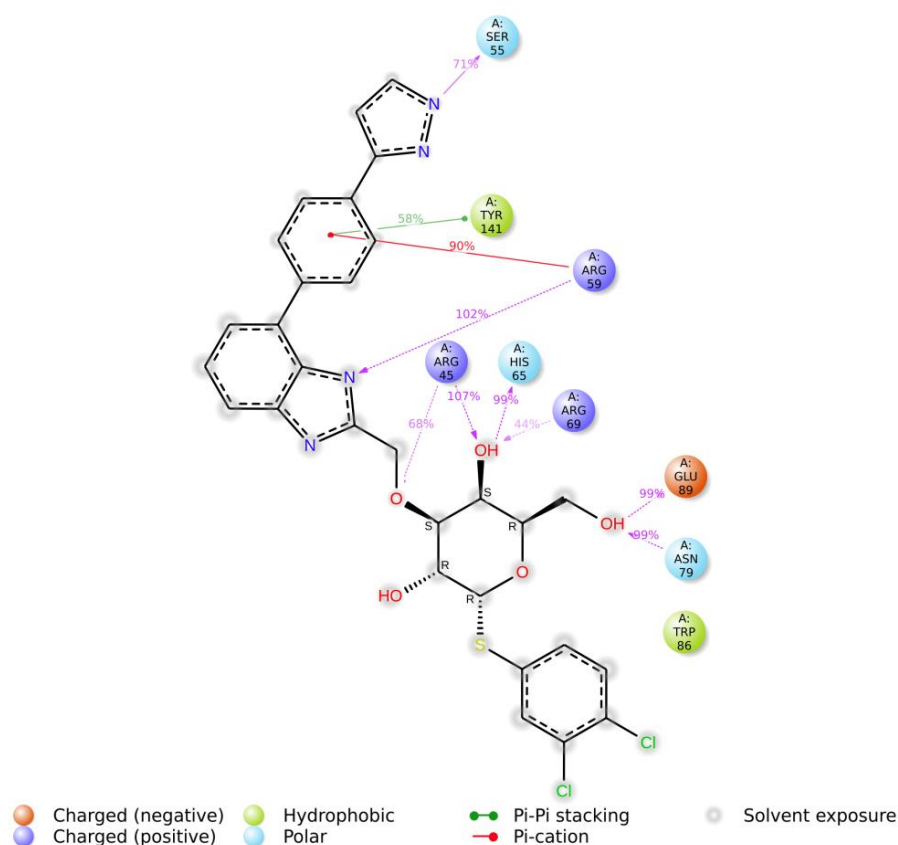

Figure 7. *Ligand-Protein Contacts between D-galactoside benzimidazole derivative with a 3-phenyl-1H-pyrazole substituent at the C-4 position of the benzimidazole and Gal-8N. 20ns simulation time. Interactions that occur more than 30.0% of the simulation time are shown.*

We hypothesized that placing the substituents in a para position on the phenyl at benzimidazole C-4 would better match the CRD pocket shape. To test this, we compared the meta and para positions of several substituents, with a hydroxyl substituent being exemplified in **Figure 8**.

The results revealed that the para position may be superior for inhibition and increased selectivity of Gal-8N, given that fewer Arg59 interactions, and no Asp49, Gln51, or Tyr141 interactions are found for the meta-position. This, however, should be taken with a grain of salt and must be empirically proven by synthesizing the compounds and testing them, since *in silico* simulations may of course not accurately reflect the real interactions and behavior between the inhibitor and the protein, since the computer model cannot simulate the complex biological

system (yet), and does not consider all parameters and interactions accurately, *e.g.*, fluctuating pH, ligand solubility, open system, atom trajectory, etc.
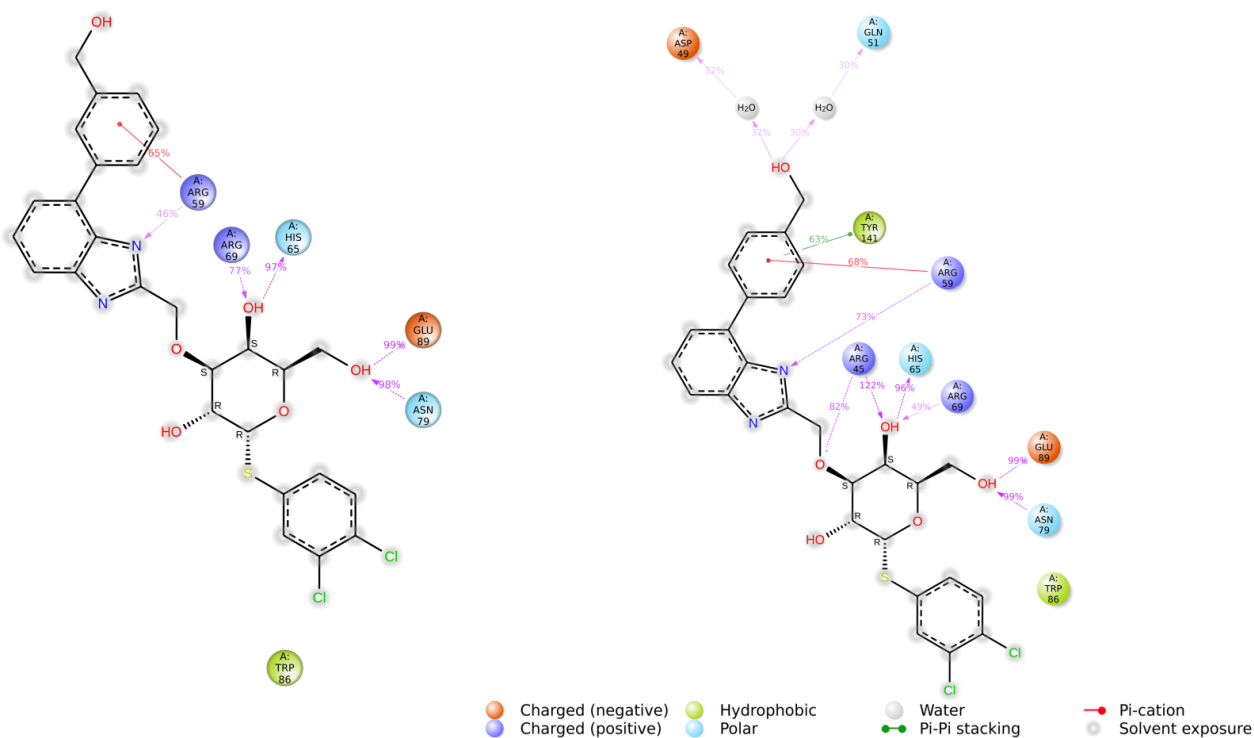

*Figure 8. Comparison of ligand-protein contacts between D-galactoside benzimidazole derivatives with hydroxymethylphenyl substituents at the C-4 position of the benzimidazole and Gal-8N. **Left**: m-hydroxymethylphenyl substituent. **Right**: p-hydroxymethyphenyl substituent. 20 ns simulation time. Interactions that occur more than 30.0% of the simulation time are shown.*

Synthesis

The synthetic route as seen in **Scheme 1** was designed and optimized by Dr. Mujtaba Hassan and MSc Radvile Juškaitė in the Ulf Nilsson group at Lund University [45]. By first forming a glycosidic bond between the β-D-galactose pentaacetate and the 3,4-dichlorobenzenethiol, with the Suzuki-coupling as the last step, the usage of protecting groups is not necessary, and a wide library of compounds with substituents at the C-4 position of the benzimidazole can quickly be established.
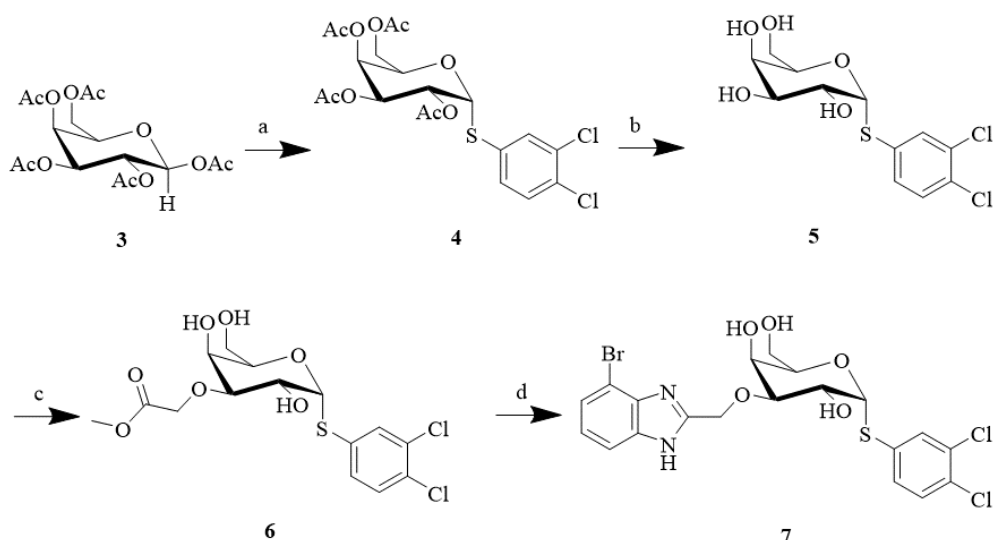
Scheme 1 *Synthetic route towards D-galactoside-based benzimidazoles. (a) 1. PCl₅, BF₃O(Et)₂, dry DCM, rt, 10 min. 2. NaH, dry DMF, 3,4-dichlorobenzenethiol, 50 °C, overnight. (27.5%). (b) dry methanol, sodium methoxide, rt, overnight. (66%) (c) 10:1 ACN/DMF, methyl bromoacetate, dibutyltin chloride, TBAB, potassium carbonate, 80 °C, overnight. (46-66%) (d) 3-bromo-1,2-diamonobenzene, H₂O, 100 °C reflux, 24h. (48.9%). The final step is shown in **Scheme 2**.*

Step a

β-D-galactose pentaacetate **3** was reacted with phosphorous pentachloride and boron trifluoride diethyl etherate yielding equatorial chlorination at the C-1 position of the carbohydrate. The C-1 equatorial β-configuration is kinetically favored due to neighboring group participation with the acetoxy group at the C-2 position. However, the α-anomer configuration is thermodynamically favored, hence it is important to quench the reaction as soon as the starting material is consumed to prevent the chloride from anomerising to the axial α-position.

The anomeric effect is where the axial α-configuration is thermodynamically favored over the equatorial β-configuration for electronegative substituents on the anomeric carbon (adjacent to a heteroatom in a six-membered ring system). This is a stereoelectronic effect where the ring is stabilized from donation of the oxygen's free axial lone pairs into the C-Cl σ* [48].
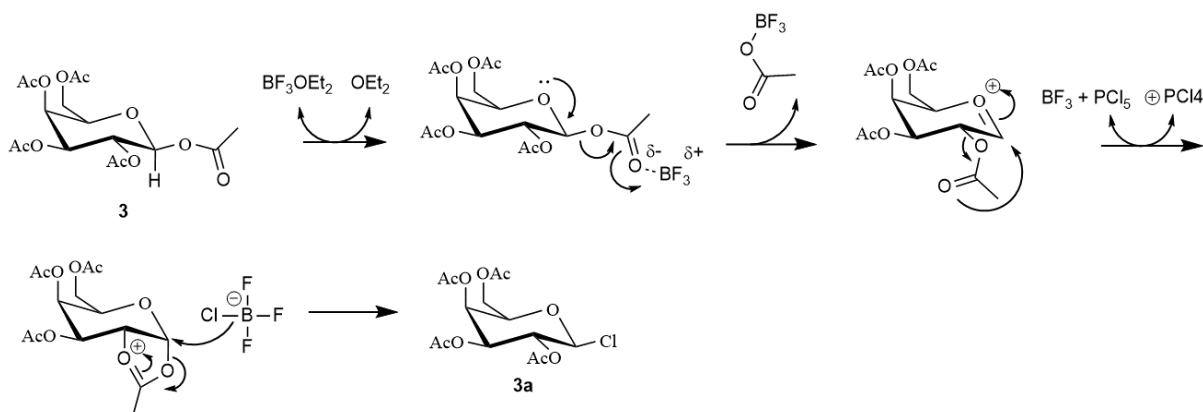
3,4-Dichlorobenzenethiol and sodium hydride were then mixed to deprotonate [49] the thiol before adding the chlorinated sugar **3a** where an Sₙ2 reaction occurs. The deprotonated thiol attacks the chloride-containing carbon yielding the α-thioglycoside **4** (**Scheme 3**). Note that the configurational change from the equatorial to the axial position is due to the Sₙ2 reaction [50].
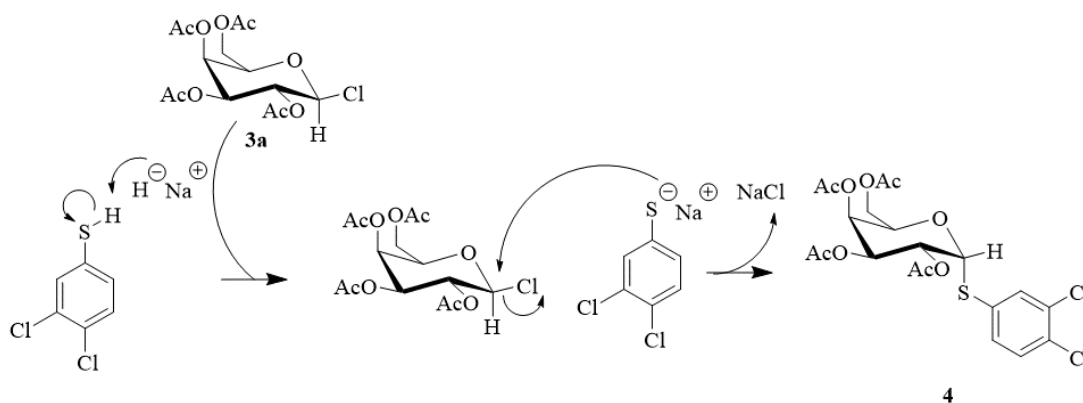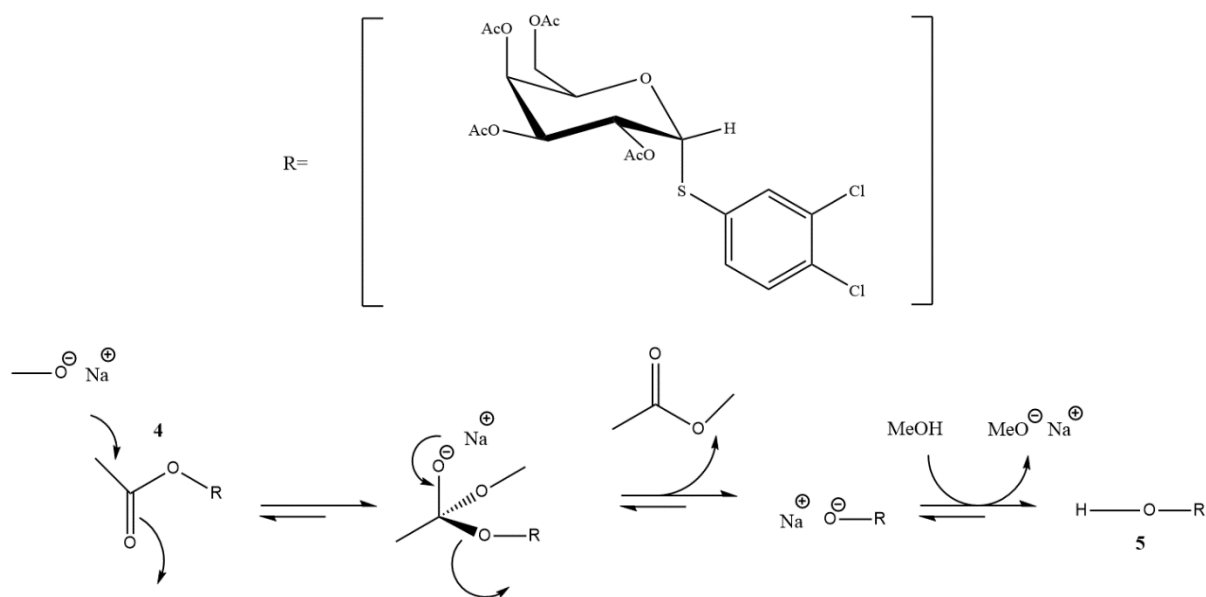

Scheme 3 Reaction mechanism for the Sₙ2 reaction giving an α-thioglycoside.

The yield of the synthesis by Pal *et al*. was reported to be 68% [46] is substantially higher than the achieved yield of 27.5%. For the first step, the TLC was checked ten minutes after the reaction started where some starting material was remaining and quenched after 17 minutes. Hence, it is possible that not all the starting material was consumed. Comparatively, the reference synthesis was allowed to react for 30 minutes before being quenched.

In addition, product **4** was reported to be a white solid and was purified by column chromatography twice compared to this experiment, where it was purified by column chromatography only once and yielded a yellow solid.

Step b

For the next step, compound **4** was dissolved in dry methanol with the addition of sodium methoxide to initiate a Zemplén deacetylation [51]

Scheme 4 *Reaction mechanism for Zemplén deacetylation.*

Only a catalytic amount of sodium methoxide is needed since the sodium methoxide is regenerated following the deacetylation as can be seen in **scheme 4**. The reaction was run overnight with a yield of 66% after purification by column chromatography. Interestingly, when dissolving the α-thioglycoside **4** in methanol, a white precipitate formed, which in hindsight could have been filtered out before adding the sodium methoxide, but at the time it was not known if it could have been the glycoside. LCMS-analysis of the white precipitate later revealed no traces of starting material **4**, hence it could safely have been filtered out.

Step c

Regioselectivity and protection is key when moieties are added to carbohydrates to prevent cross-reactions. Organotin reagents in stoichiometric amount used to be a popular method [52] to achieve this. However, due to the health and environmental hazards associated with organotins [53], new methods have been developed to minimize their usage. For **step c** a method [54] using dibutyltin dichloride and TBAB in catalytic amounts was used to achieve regioselective mono-alkylation. First, the addition of dibutyltin dichloride to compound **5** forms a dibutylstannylene acetal with the vicinal hydroxyl groups at positions C-3 and C-4, followed by bromide from TBAB coordinating with tin forming an equilibrium between the tetracoordinate and pentacoordinate tin atom where the selective cleavage of Sn-O [52]leads to a reactive oxygen anion coordinated to the TBA ion. A nucleophilic attack then proceeds from the oxygen to the methyl bromoacetate. Another unreacted compound **5** is then deprotonated by potassium carbonate and coordinating to the same tin atom where species

exchange between the tetracoordinate and pentacoordinate tin gives compound **6**, and regeneration of the dibutylstannylene acetal [55].
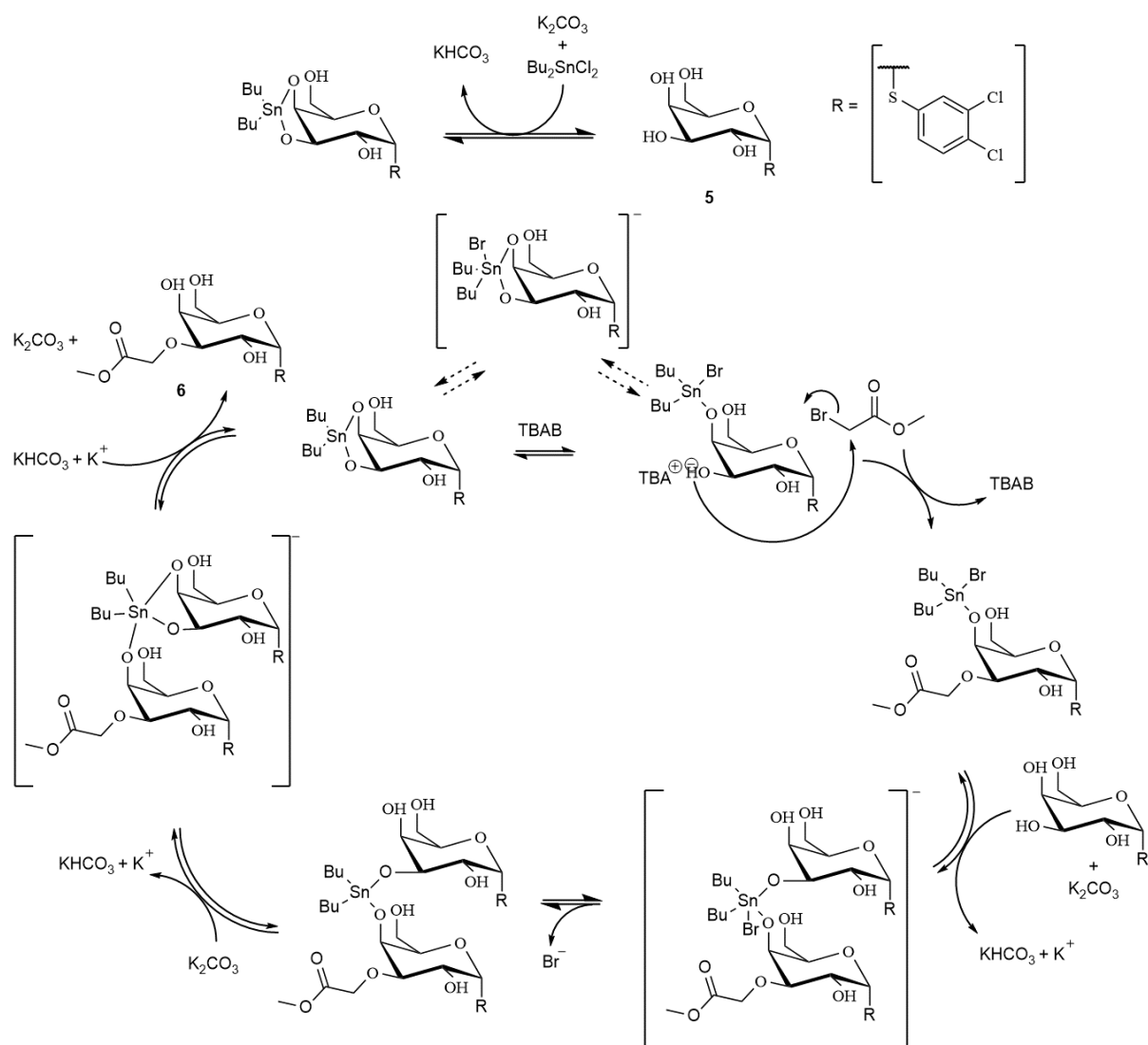

Scheme 5 Proposed [55] *reaction mechanism for regioselective mono-alkylation.*

The experiment was performed twice with yields of 46% and 66% respectively after column purification. The conditions were equal for both experiments except that for the second experiment unreacted starting material (compound **5**) from the first experiment was used. Since the starting material for the second step was purified through the first experiment, it is possible that the higher attained yield was due to a higher purity of compound **5**.

Step d

Compound **6** was mixed with 3-bromo-1,2-diaminobenzene using water as a solvent and heated to 100°C under reflux for 24 hours yielding compound **7**. The reaction mechanism is shown in **Scheme 6** [56]. First, a nucleophilic substitution takes place with one of the amines as the

nucleophile and a methoxide leaving group, following deprotonation by the solvent. Another nucleophilic attack then occurs from the other amine, once again, on the carbonyl followed by protonation of the hydroxide yielding compound **7** with water as a leaving group.
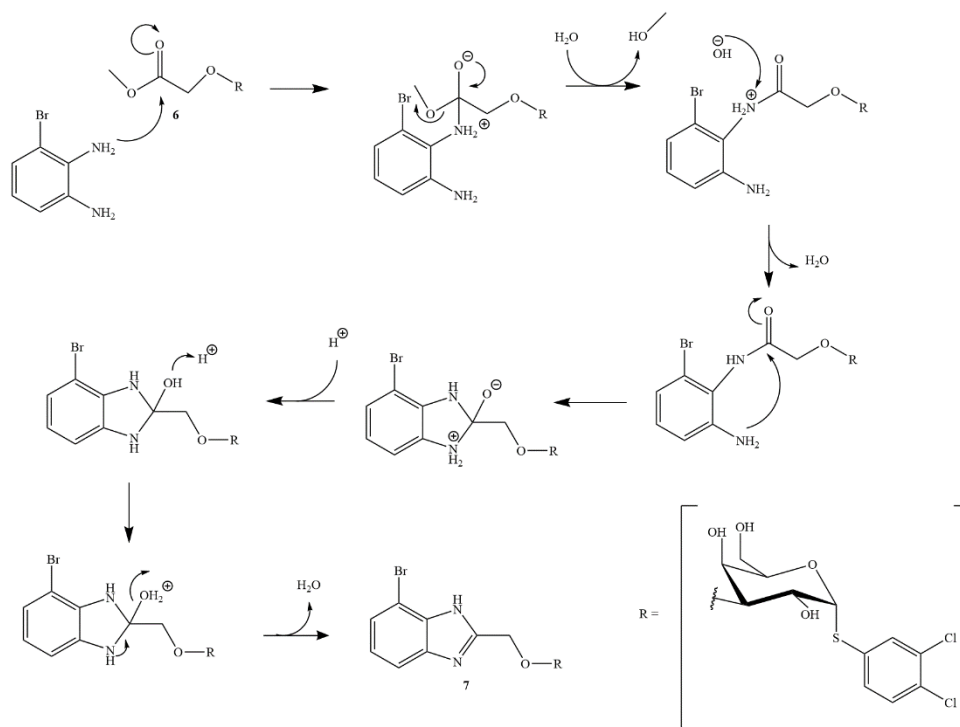

Scheme 6 *Reaction mechanism for benzimidazole cyclization.*

Following the completion of the reaction by monitoring with TLC, different solvent systems were evaluated for the column chromatography using TLC. It was eventually decided to do a gradient flash starting with only DCM and gradually ramping up to 4 parts methanol with a yield of 48.9% of **7**.

Final step

Based on the MD simulation results and availability of boronic acids for Suzuki-coupling, the compounds shown in **Scheme 7** were selected and synthesized as the final compounds to be assessed.
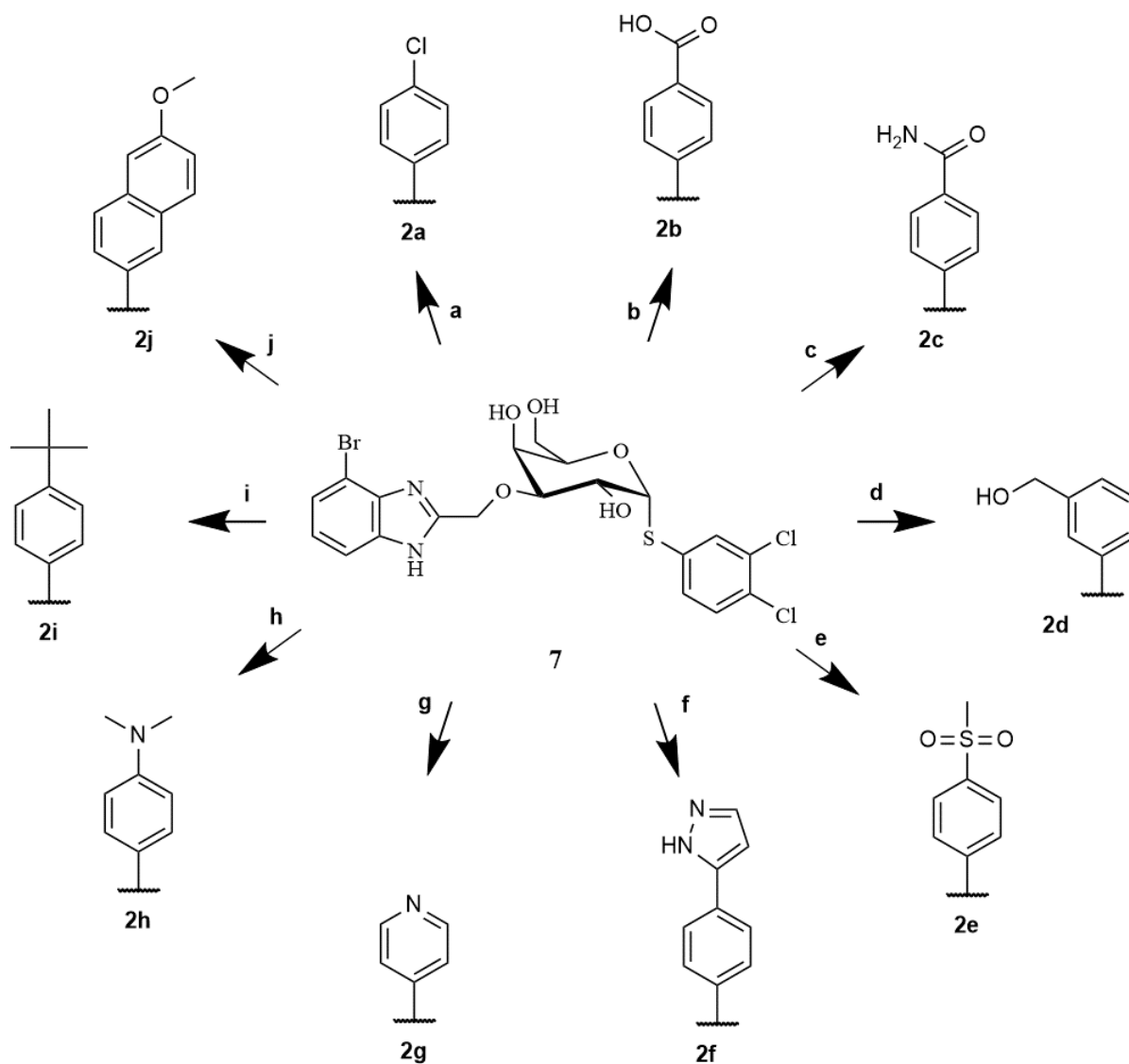
Scheme 7 *Final step achieving D-galactoside-based benzimidazole derivatives through Suzuki-coupling at the C-4 position of the benzimidazole. All reactions (a-j), with the exception of the boronic acid used, were performed under the same conditions: potassium carbonate, Pd(PPH3)4, 4:1 1,4-dioxane/H2O, 110 °C microwave, 1.5h. (a) 4-chlorophenylboronic acid. (48%). (b) 4-carboxyphenylboronic acid. (48%) (c) 4-Carbamoylbenzeneboronic acid. (30%). (d) 3-(Hydroxymethyl)phenylboronic acid. (29%). (e) 4-Methanesulfonylphenylboronic acid. (32%). (f) [4-(1H-Pyrazol-5-yl)phenyl]boronic acid. (28%). (g) 4-Pyridinylboronic acid. (26%). (h) 4-(Dimethylamino)phenylboronic acid. (14%). (i) 4-tert-Butylphenylboronic acid. (25%). (g) 6-Methoxy-2-naphthaleneboronic acid. (27%).*

For the first experiment, three equivalences of sodium carbonate were used as the base with a solvent system of 1:2 water/dioxane along with compound **7** and Pd(PPh$_3$)$_4$ as catalyst. After one and a half hours of microwaving at 120°C, LCMS showed a big peak of compound **7** remaining with a small trace of the product showing up. We hypothesized that the base and the solvent system might have been the culprit since it plays a big part in Suzuki-coupling, so it was decided to change the base to potassium carbonate and a 1:4 water/dioxane mixture instead since it had been successful in previous experiments at the laboratory. This, however, achieved the same results as the previous attempt, yielding a small amount of product with a large amount of compound **7** remaining.

We gathered that the catalyst used had been exposed to air since it presented a dark brown, almost black color, hence a less decomposed catalyst was used for the next experiment. That experiment, sadly, revealed the same results, thus [(dppf)PdCl$_2$] as the catalyst was tried instead, but it turned out much the same.

As a last hurrah, compound **7** was purified by preparative HPLC before attempting the Suzuki-coupling and finally LCMS revealed full consumption of compound **7** in the Suzuki-coupling using Pd(PPh$_3$)$_4$ as catalyst and three equivalences of potassium carbonate as base. With this knowledge, the most likely explanation is that the palladium catalyst was poisoned from byproducts or starting material from previous steps.

Amines are strong EDGs and can easily coordinate to palladium and other transition metals which could lead to poisoning of the catalyst [57], [58]. For the cyclization reaction 3-bromo-1,2-diaminobenzene was used, hence it is possible that this impurity was still prevalent after purification by column chromatography leading to disruption of the Suzuki-cycle.

Another potential catalyst poison may stem from the thiol in the initial step since sulfur [59] and its related compounds strongly bind to the surface of metals [60], deactivating the catalyst.
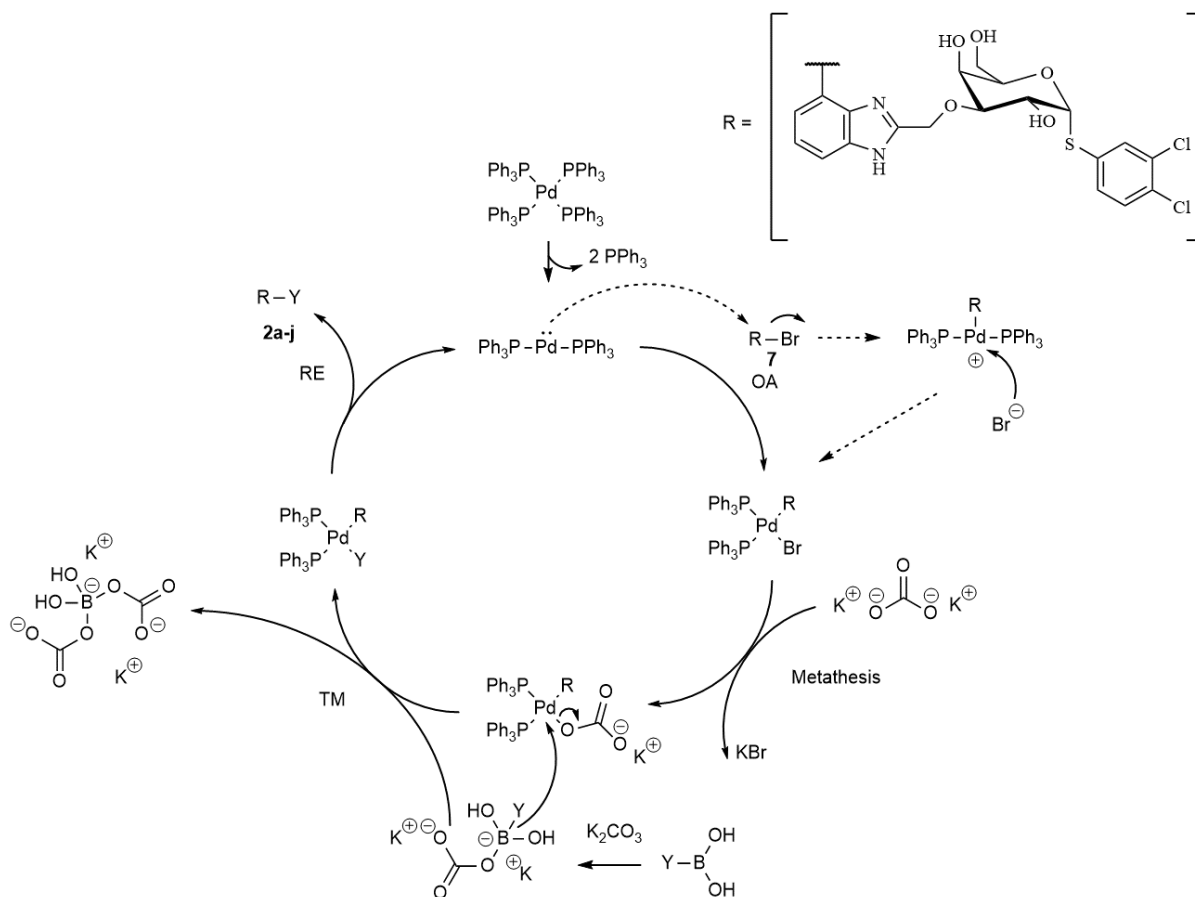
Scheme 8 *Reaction mechanism for Suzuki coupling using Pd(PPH₃)₄*

The mechanism for the reaction using the catalyst Pd(PPh₃)₄ is shown in **Scheme 8** and begins with the disassociation of two PPh₃ ligands to the solvent followed by an oxidative addition that has been proposed to begin with an S_N2-type mechanism [61], with the unsaturated palladium acting as the nucleophile, attacking the bromide-containing carbon of compound **7**. The oxidative addition is completed with the coordination of the bromide to the palladium. An anionic bond exchange then proceeds between the potassium carbonate and the palladium complex with the bromide coupled to the catalyst displaced by the carbonate of the base and potassium bromide being formed as a salt. The base is also important to activate the boronic acid, by forming a borate complex, which then undergoes transmetallation followed by reductive elimination where compound **7** is yielded and palladium (0) is regenerated [62].

Yield and purity

Following preparative HPLC, compounds **2a-2j** were sent for purity assessment through HPLC-analysis with the yields presented in percentage after the final purification (**Table 1**). As can be seen, the purity was suboptimal with the only compound passing >95% purity being compound **2i**, most likely owing to its higher hydrophobicity and lower polarity giving a better separation in preparative HPLC. A possible explanation for the difficulty of the purification is

the potential prevalence of byproducts and starting material from previous steps of the synthesis process with similar chemical properties resulting in overlapping elution times. This could potentially have been rectified by running column chromatography following the completion of the Suzuki-coupling or by exploring different solvent systems and programs for purification by preparative HPLC.

Interesting to note is the alleged lower purity achieved after running a second preparative HPLC purification. The purity was confirmed by HPLC-analysis, which turned out to be inaccurate and faulty, as reported by fellow laboratory workers and the department, during the period following the second preparative HPLC, hence the reported purity in the last column in **Table 1** should not be considered accurate. However, the compounds were also checked with proton NMR, which confirmed that the purities were less than 95%.
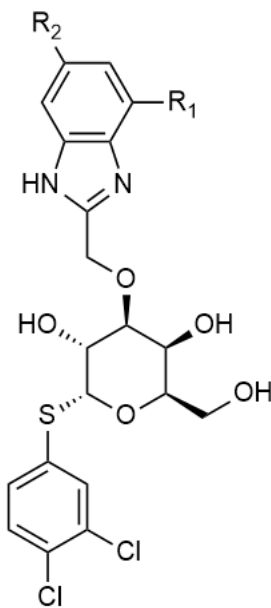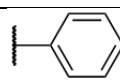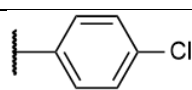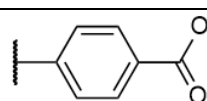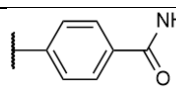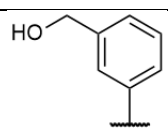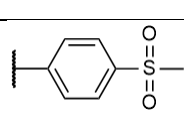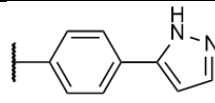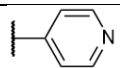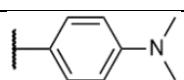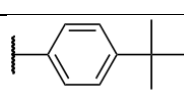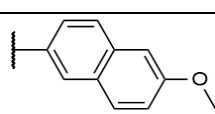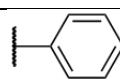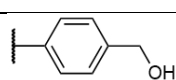
Table 1. *Yield and purity for the final compounds. Yield reported after final purification step. Compounds 2a, 2b, 2g, 2i, 2j were purified by preparative HPLC once. The other compounds twice. Purity results from HPLC.*

Compound	Yield (%)	Purity post first prep. HPLC (%)	Purity post second prep. HPLC (%)
2a	48	94	-
2b	48	87	-
2c	30	85	75
2d	29	90	82
2e	32	77	92
2f	28	76	68.6
2g	26	86	-
2h	14	82	39
2i	25	98	-
2j	27	94	-

Binding affinity data

A competitive fluorescence polarization assay [41] was used to assess the affinity of the final compounds to Gal-8N. During initial inhibition testing, synthesized compounds **2a-2g** except for compound **2b** precipitated from the solution, hence the K_D values under footnote a, in **Table 2**, are not to be considered accurate. The compounds were retested, with no revealed precipitation, and all the available data is presented in **Table 2**.

Table 2. *Binding affinity to Gal-8N reported in μM K_D (dissociation constant; average ± SEM).*

Compound		K_D (μM)	R_1	R_2
	1[c]	5 ± 0.5		COOH
	2a	170[a] 89 ± 23		H
	2b	12 ± 0.8 11 ± 1		H
	2c	20 ± 5[a]		H
	2d	72 ± 13[a] 19 ± 2.6[a] 20 ± 5		H
	2e	107 ± 0.2[a] 35 ± 8[a]		H
	2f	170[a] 210 ± 11		H
	2g	70 ± 12[a]		H
	2h	NT[b]		H
	2i	720 ± 22		H
	2j	430 ± 9.6		H
	8a[d]	39 ± 9.4		H
	8b[d]	16 ± 3.9		H

[a]*Precipitation of compound during assay binding.*
[b]*Not tested.*
[c][45]
[d]Previously synthesized in the Nilsson group by a student, not published.

Compounds **2i** and **2j** showed poor affinity at 720 ± 22 µM respectively 430 ± 9.6 µM. This is not an unexpected result with regard to their substituents, since compound **2i** carries a bulky, nonpolar tert-butyl group and compound **2j** with its naphthalene may give rise to steric hindrance and conformation issues.

MD simulations of the pyrazole **2f** suggested promising interactions with Gal-8N, however as FP-testing revealed it had quite poor affinity at 210 ± 11 µM. This may be due to the substituent being too bulky to conform into the pocket, thus hindering the Ser55 interaction with the pyrazole.

An unexpected result came from the chlorophenyl **2a** having a K_D value of 89 ± 23 µM which is two-fold worse than the inhibitor with an unsubstituted phenyl substituent (**8b**) showing inhibition at 39 ± 9.4 µM. This could be due to a low spread of data points (eight data points spread between $15 - 45$ % inhibition) for K_D evaluation during FP-testing leading to an inaccurate K_D value for compound **2a**. Chloride is also a weak EWG which might factor into a slightly weakened π-cation interaction between the phenyl and the protein, but most likely not by a factor of two. Hence, the chlorophenyl **2a** may warrant FP re-testing given the potential of an inaccurate measurement of the K_D, and the chlorophenyl functional group having shown improved affinities for previously synthesized inhibitors in the lab.

Compound **2d** with a meta-hydroxy methylphenyl group showed the second-best inhibition at 20 ± 5 µM. As a reference, the regioisomeric **8b** with a para-hydroxymethyl showed a slightly better K_D of 16 ± 3.9 µM. This suggests a slight favorability to the para-position over the meta-position as postulated earlier, although this result is within the margins of error.

Compound **2b** showed the best K_D value of 11 ± 1 µM. This is higher than the unsubstituted reference phenyl derivative **1**, which supports that ligand-C-6 carboxylate in **1** interacting with Gly142 yields better affinity to the protein. From the initial MD simulation, we gathered that the carboxylate of benzoic acid in **2b** can interact with Gal-8N through a water-mediated hydrogen bond to Ser141, thus the good affinity of **2b** supports this hypothesis and thus suggests that interactions with Ser141 should be further investigated.

Conclusions

Ten D-galactoside-based benzimidazole inhibitors were successfully synthesized. For each of the Suzuki-couplings, only 20 mg of compound **7** were used, with several 100 mg remaining after the ten final compounds were synthesized. Given the leftover material, it would have been preferable to use a larger amount of compound **7** per coupling, both to have more of the final compounds available for testing, and to ease the purification process.

The carboxyphenyl **2b** showed the best inhibition at 11 ± 1 µM which is worse, but only by ~6 µM, than the previously synthesized regioisomeric carboxylate **1**. This suggests that the Gly142-carboxylate interaction omitted from compound **2b** (and all the synthesized compounds), increases affinity to Gal-8N, and that the Ser55-ligand interaction, absent in compound **1**, increases affinity as well. Thus, including these two interactions should lead to more potent Gal-8N inhibitors. Arg59 was also found to have π-cation interactions with the aromatic ring according to the MD simulations. Hence, it could be interesting to study the π-cation effects of the functional groups on the arenes, *i.e.*, compound **2b** has a carboxylic acid group on the arene which, being an EWG, may weaken the π-cation bond. A good functional group on the arene should, consequently, strengthen the mentioned π-cation bond while keeping the Ser55 interaction, potentially a hydroxy, an amino, or an acylamido group could be interesting.

References

[1]   S. H. Barondes *et al.*, "Galectins: A family of animal β-galactoside-binding lectins," *Cell*, vol. 76, no. 4, pp. 597–598, Feb. 1994, doi: 10.1016/0092-8674(94)90498-7.

[2]   R.-Y. Yang, G. A. Rabinovich, and F.-T. Liu, "Galectins: structure, function and therapeutic potential.," *Expert Rev Mol Med*, vol. 10, p. e17, Jun. 2008, doi: 10.1017/S1462399408000719.

[3]   J. Hirabayashi and K. Kasai, "The family of metazoan metal-independent beta-galactoside-binding lectins: structure, function and molecular evolution.," *Glycobiology*, vol. 3, no. 4, pp. 297–304, Aug. 1993, doi: 10.1093/glycob/3.4.297.

[4]   D. Houzelstein *et al.*, "Phylogenetic analysis of the vertebrate galectin family.," *Mol Biol Evol*, vol. 21, no. 7, pp. 1177–1187, Jul. 2004, doi: 10.1093/molbev/msh082.

[5]   J. Seetharaman, A. Kanigsberg, R. Slaaby, H. Leffler, S. H. Barondes, and J. M. Rini, "X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-A resolution.," *J Biol Chem*, vol. 273, no. 21, pp. 13047–13052, May 1998, doi: 10.1074/jbc.273.21.13047.

[6]   A. R. Kolatkar and W. I. Weis, "Structural basis of galactose recognition by C-type animal lectins.," *J Biol Chem*, vol. 271, no. 12, pp. 6679–6685, Mar. 1996.

[7]   H. Leffler, S. Carlsson, M. Hedlund, Y. Qian, and F. Poirier, "Introduction to galectins.," *Glycoconj J*, vol. 19, no. 7–9, pp. 433–440, 2002, doi: 10.1023/B:GLYC.0000014072.34840.04.

[8]   R. C. Hughes, "Galectins as modulators of cell adhesion.," *Biochimie*, vol. 83, no. 7, pp. 667–676, Jul. 2001, doi: 10.1016/s0300-9084(01)01289-5.

[9]   I. R. Nabi, J. Shankar, and J. W. Dennis, "The galectin lattice at a glance," *J Cell Sci*, vol. 128, no. 13, pp. 2213–2219, 2015, doi: 10.1242/jcs.151159.

[10]   H. Leffler, "Galectins structure and function--a synopsis.," *Results Probl Cell Differ*, vol. 33, pp. 57–83, 2001, doi: 10.1007/978-3-540-46410-5_4.

[11]   I. Kuwabara, H. Sano, and F.-T. Liu, "Functions of galectins in cell adhesion and chemotaxis.," *Methods Enzymol*, vol. 363, pp. 532–552, 2003, doi: 10.1016/S0076-6879(03)01078-4.

[12]   S. Akahani, P. Nangia-Makker, H. Inohara, H. R. Kim, and A. Raz, "Galectin-3: a novel antiapoptotic molecule with a functional BH1 (NWGR) domain of Bcl-2 family.," *Cancer Res*, vol. 57, no. 23, pp. 5272–5276, Dec. 1997.

[13]   I. Kuwabara *et al.*, "Galectin-7 (PIG1) exhibits pro-apoptotic function through JNK activation and mitochondrial cytochrome c release.," *J Biol Chem*, vol. 277, no. 5, pp. 3487–3497, Feb. 2002, doi: 10.1074/jbc.M109360200.

[14]   G. M. Ashraf *et al.*, "Galectins-A Potential Target for Cardiovascular Therapy.," *Curr Vasc Pharmacol*, vol. 15, no. 4, pp. 296–312, 2017, doi: 10.2174/1570161115666170201113046.

[15]   R. Aalinkeel *et al.*, "Galectin-1 Reduces Neuroinflammation via Modulation of Nitric Oxide-Arginase Signaling in HIV-1 Transfected Microglia: a Gold Nanoparticle-Galectin-1 'Nanoplex' a Possible Neurotherapeutic?," *J Neuroimmune Pharmacol*, vol. 12, no. 1, pp. 133–151, Mar. 2017, doi: 10.1007/s11481-016-9723-4.

[16]   N. E. Reticker-Flynn *et al.*, "A combinatorial extracellular matrix platform identifies cell-extracellular matrix interactions that correlate with metastasis.," *Nat Commun*, vol. 3, p. 1122, 2012, doi: 10.1038/ncomms2128.

[17]   C.-H. Li, Y.-C. Chang, M.-H. Chan, Y.-F. Yang, S.-M. Liang, and M. Hsiao, "Galectins in Cancer and the Microenvironment: Functional Roles, Therapeutic Developments, and Perspectives.," *Biomedicines*, vol. 9, no. 9, Sep. 2021, doi: 10.3390/biomedicines9091159.

[18]   S. Sato, M. Ouellet, C. St-Pierre, and M. J. Tremblay, "Glycans, galectins, and HIV-1 infection.," *Ann N Y Acad Sci*, vol. 1253, pp. 133–148, Apr. 2012, doi: 10.1111/j.1749-6632.2012.06475.x.

[19]   S. Sciacchitano *et al.*, "Galectin-3: One Molecule for an Alphabet of Diseases, from A to Z.," *Int J Mol Sci*, vol. 19, no. 2, Jan. 2018, doi: 10.3390/ijms19020379.

[20]   K. V. Mariño, A. J. Cagnoni, D. O. Croci, and G. A. Rabinovich, "Targeting galectin-driven regulatory circuits in cancer and fibrosis," *Nat Rev Drug Discov*, vol. 22, no. April, 2023, doi: 10.1038/s41573-023-00636-2.

[21]   Y. Vinik, H. Shatz-Azoulay, and Y. Zick, "Molecular Mechanisms Underlying the Role of Galectin-8 as a Regulator of Cancer Growth and Metastasis," *Trends in Glycoscience and Glycotechnology*, vol. 30, 2018.

[22]   W.-S. Chen *et al.*, "Pathological lymphangiogenesis is modulated by galectin-8-dependent crosstalk between podoplanin and integrin-associated VEGFR-3.," *Nat Commun*, vol. 7, p. 11302, Apr. 2016, doi: 10.1038/ncomms11302.

[23]   S. Beyer *et al.*, "Galectin-8 and -9 as prognostic factors for cervical cancer," *Arch Gynecol Obstet*, vol. 306, no. 4, pp. 1211–1220, 2022, doi: 10.1007/s00404-022-06449-9.

[24]   H. Zhu *et al.*, "Prognostic Value and Biological Function of Galectins in Malignant Glioma," *Front Oncol*, vol. 12, no. June, pp. 1–16, 2022, doi: 10.3389/fonc.2022.834307.

[25]   Z. Z. Su, J. Lin, R. Shen, P. E. Fisher, N. I. Goldstein, and P. B. Fisher, "Surface-epitope masking and expression cloning identifies the human prostate carcinoma tumor antigen gene PCTA-1 a member of the galectin gene family.," *Proc Natl Acad Sci U S A*, vol. 93, no. 14, pp. 7252–7257, Jul. 1996, doi: 10.1073/pnas.93.14.7252.

[26]   L. Johannes, R. Jacob, and H. Leffler, "Galectins at a glance.," *J Cell Sci*, vol. 131, no. 9, May 2018, doi: 10.1242/jcs.208884.

[27]   Y. Levy *et al.*, "Galectin-8 functions as a matricellular modulator of cell adhesion.," *J Biol Chem*, vol. 276, no. 33, pp. 31285–31295, Aug. 2001, doi: 10.1074/jbc.M100340200.

[28]   F. Ferragut *et al.*, "Dual knockdown of Galectin-8 and its glycosylated ligand, the activated leukocyte cell adhesion molecule (ALCAM/CD166), synergistically delays in vivo breast cancer growth," *Biochim Biophys Acta Mol Cell Res*, vol. 1866, no. 8, pp. 1338–1352, 2019, doi: 10.1016/j.bbamcr.2019.03.010.

[29]   J. F. Sampson, A. Suryawanshi, W. Chen, G. A. Rabinovich, and N. Panjwani, "Galectin-8 promotes regulatory T cell differentiation by modulating IL-2 and TGFβ signaling," vol. 94, no. 2, pp. 213–219, 2016, doi: 10.1038/icb.2015.72.Galectin-8.

[30]   T. L. M. Thurston, M. P. Wandel, N. von Muhlinen, A. Foeglein, and F. Randow, "Galectin 8 targets damaged vesicles for autophagy to defend cells against bacterial invasion.," *Nature*, vol. 482, no. 7385, pp. 414–418, Jan. 2012, doi: 10.1038/nature10744.

[31]   N. Pied *et al.*, "TBK1 is part of a galectin 8 dependent membrane damage recognition complex and drives autophagy upon Adenovirus endosomal escape.," *PLoS Pathog*, vol. 18, no. 7, p. e1010736, Jul. 2022, doi: 10.1371/journal.ppat.1010736.

[32]   D. Liu, H. Zhu, and C. Li, "Galectins and galectin-mediated autophagy regulation: new insights into targeted cancer therapy," *Biomark Res*, vol. 11, no. 1, pp. 1–17, 2023, doi: 10.1186/s40364-023-00466-9.

[33]   A. Wittrup *et al.*, "Visualizing lipid-formulated siRNA release from endosomes and target gene knockdown," *Nat Biotechnol*, vol. 33, no. 8, pp. 870–876, 2015, doi: 10.1038/nbt.3298.

[34]   Y. R. Hadari, M. Eisenstein, R. Zakut, and Y. Zick, "Galectin-8: on the Road from Structure to Function.," *Trends in Glycoscience and Glycotechnology*, vol. 9, no. 45, pp. 103–112, 1997, doi: 10.4052/tigg.9.103.

[35]   Y. R. Hadari, K. Paz, R. Dekel, T. Mestrovic, D. Accili, and Y. Zick, "Galectin-8," *Journal of Biological Chemistry*, vol. 270, no. 7, pp. 3447–3453, Feb. 1995, doi: 10.1074/jbc.270.7.3447.

[36]   H. Ideo, "The N-terminal carbohydrate recognition domain of galectin-8 recognizes specific glycosphingolipids with high affinity," *Glycobiology*, vol. 13, no. 10, pp. 713–723, Jun. 2003, doi: 10.1093/glycob/cwg094.

[37]   S. R. Stowell, C. M. Arthur, K. A. Slanina, J. R. Horton, D. F. Smith, and R. D. Cummings, "Dimeric Galectin-8 Induces Phosphatidylserine Exposure in Leukocytes through Polylactosamine Recognition by the C-terminal Domain," *Journal of Biological Chemistry*, vol. 283, no. 29, pp. 20547–20559, Jul. 2008, doi: 10.1074/jbc.M802495200.

[38]   Y. Levy *et al.*, "It depends on the hinge: a structure-functional analysis of galectin-8, a tandem-repeat type lectin," *Glycobiology*, vol. 16, no. 6, pp. 463–476, Jun. 2006, doi: 10.1093/glycob/cwj097.

[39]   S. A. Hollingsworth and R. O. Dror, "Molecular Dynamics Simulation for All," *Neuron*, vol. 99, no. 6, pp. 1129–1143, 2018, doi: https://doi.org/10.1016/j.neuron.2018.08.011.

[40]   W. A. Lea and A. Simeonov, "Fluorescence polarization assays in small molecule screening.," *Expert Opin Drug Discov*, vol. 6, no. 1, pp. 17–32, Jan. 2011, doi: 10.1517/17460441.2011.537322.

[41]   P. Sörme, B. Kahl-Knutsson, M. Huflejt, U. J. Nilsson, and H. Leffler, "Fluorescence polarization as an analytical tool to evaluate galectin–ligand interactions," *Anal Biochem*, vol. 334, no. 1, pp. 36–47, 2004, doi: https://doi.org/10.1016/j.ab.2004.06.042.

[42]   X. Du *et al.*, "Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods.," *Int J Mol Sci*, vol. 17, no. 2, Jan. 2016, doi: 10.3390/ijms17020144.

[43]   M. Hassan *et al.*, "Structure-Guided Design of d-Galactal Derivatives with High Affinity and Selectivity for the Galectin-8 N-Terminal Domain.," *ACS Med Chem Lett*, vol. 12, no. 11, pp. 1745–1752, Nov. 2021, doi: 10.1021/acsmedchemlett.1c00371.

[44] M. Hassan *et al.*, "Benzimidazole-galactosides bind selectively to the Galectin-8 N-Terminal domain: Structure-based design and optimisation.," *Eur J Med Chem*, vol. 223, p. 113664, Nov. 2021, doi: 10.1016/j.ejmech.2021.113664.

[45] R. Juškaitė, "Design and Synthesis of D -Galactal-Based Benzimidazole and Benzoxazole Derivatives as Inhibitors of Galectin-8 N-Terminal Domain," 2022.

[46] K. B. Pal *et al.*, "Quinoline-galactose hybrids bind selectively with high affinity to a galectin-8 N-terminal domain," *Org Biomol Chem*, vol. 16, no. 34, pp. 6295–6305, 2018, doi: 10.1039/c8ob01354c.

[47] C. Ballatore, D. M. Huryn, and A. B. 3rd Smith, "Carboxylic acid (bio)isosteres in drug design.," *ChemMedChem*, vol. 8, no. 3, pp. 385–395, Mar. 2013, doi: 10.1002/cmdc.201200585.

[48] J. G. N. W. S. Clayden, *Organic Chemistry*, 2nd ed., Oxford University Press, 2012, pp. 800-805.

[49] J. G. N. W. S. Clayden, *Organic Chemistry*, 2nd ed., Oxford University Press, 2012, pp. 130.

[50] J. G. N. W. S. Clayden, *Organic Chemistry*, 2nd ed., Oxford University Press, 2012, pp. 340-342.

[51] C. Pezzetta, S. Tshepelevitsh, I. Leito, and G. Poli, "Comment on 'Zemplén transesterification: a name reaction that has misled us for 90 years' by B. Ren{,} M. Wang{,} J. Liu{,} J. Ge{,} X. Zhang and H. Dong{,} Green Chemistry{,} 2015{,} 17{,} 1390–1394," *Green Chem.*, vol. 20, no. 10, pp. 2392–2394, 2018, doi: 10.1039/C7GC03795C.

[52] Y. Zhou, J. Li, Y. Zhan, Z. Pei, and H. Dong, "Halide promoted organotin-mediated carbohydrate benzylation: mechanism and application," *Tetrahedron*, vol. 69, no. 13, pp. 2693–2700, 2013, doi: https://doi.org/10.1016/j.tet.2013.02.024.

[53] M. M. Whalen, B. G. Loganathan, and K. Kannan, "Immunotoxicity of Environmentally Relevant Concentrations of Butyltins on Human Natural Killer Cells in Vitro," *Environ Res*, vol. 81, no. 2, pp. 108–116, 1999, doi: https://doi.org/10.1006/enrs.1999.3968.

[54] H. Xu *et al.*, "Regioselective mono and multiple alkylation of diols and polyols catalyzed by organotin and its applications on the synthesis of value-added carbohydrate intermediates," *Tetrahedron*, vol. 72, no. 24, pp. 3490–3499, 2016, doi: 10.1016/j.tet.2016.04.076.

[55] H. Xu *et al.*, "Regioselective Benzylation of Diols and Polyols by Catalytic Amounts of an Organotin Reagent," *Adv Synth Catal*, vol. 356, no. 8, pp. 1735–1740, May 2014, doi: 10.1002/adsc.201301152.

[56] J. G. N. W. S. Clayden, *Organic Chemistry*, 2nd ed., Oxford University Press, 2012, pp. 126-127.

[57] P. Marcazzan, B. O. Patrick, and B. R. James, "Amine Products and Catalyst Poisoning in the Homogeneous H2 Hydrogenation of Imines Catalyzed by the [Rh(COD)(PPh3)2]PF6 Precursor," *Organometallics*, vol. 22, no. 6, pp. 1177–1179, Mar. 2003, doi: 10.1021/om020992c.

[58] J. Streuff, C. H. Hövelmann, M. Nieger, and K. Muñiz, "Palladium(II)-catalyzed intramolecular diamination of unfunctionalized alkenes," *J Am Chem Soc*, vol. 127, no. 42, pp. 14586–14587, 2005, doi: 10.1021/ja055190y.

[59] J. Liu, K. Kimmel, K. Dao, Y. Liu, and M. Qi, "Identification and Elimination of an Unexpected Catalyst Poison in Suzuki Coupling," *Org Process Res Dev*, vol. 22, no. 1, pp. 111–116, Jan. 2018, doi: 10.1021/acs.oprd.7b00342.

[60]   C. H. Bartholomew, P. K. Agrawal, and J. R. Katzer, "Sulfur Poisoning of Metals," D. D. Eley, H.
         Pines, and P. B. Weisz, Eds., in Advances in Catalysis, vol. 31. Academic Press, 1982, pp. 135–
         242. doi: https://doi.org/10.1016/S0360-0564(08)60454-X.

[61]   K. S. Y. Lau, R. W. Fries, and J. K. Stille, "Stereochemistry of oxidative addition of alkyl halides
         to palladium(0) complexes," *J Am Chem Soc*, vol. 96, no. 15, pp. 4983–4986, Jul. 1974, doi:
         10.1021/ja00822a044.

[62]   S. Fortun, P. Beauclair, and A. R. Schmitzer, "Metformin as a versatile ligand for recyclable
         palladium-catalyzed cross-coupling reactions in neat water," *RSC Adv*, vol. 7, no. 34, pp.
         21036–21044, 2017, doi: 10.1039/C7RA01197K.

[63]   M. Hassan, "MH-3-104," *Lund University*, 2021.

[64]   K. F. Ansari, C. Lal, and D. L. Parmar, "Some Novel Oxadiazolyl/Azetidinyl Benzimidazole
         Derivatives: Synthesis and in Vitro Biological Evaluation," *Synth Commun*, vol. 42, no. 24, pp.
         3553–3568, Dec. 2012, doi: 10.1080/00397911.2011.580881.

[65]   M. Hassan, "MH-4-45," *Lund University*, 2022.

Appendix I: Experimental

All reagents and solvents were well-dried prior to use according to standard methods and commercial reagents were used without further purification. Silica-gel 60 F254 aluminum sheets (Merck) were used for TLC analysis with detection by UV and charring by a solution of 10% H_2SO_4 in EtOH. Purification was done through column chromatography on SiO_2 (technical grade, 60 Å pore size, 230-400 mesh, 40-63 µm) or by preparative HPLC on an Agilent 1260 infinity system, column SymmetryPrep C18 with a mobile phase gradient of X% ACN in H_2O to Y% with 0.1% FA. Microwave reactions were performed in a Biotage Initiator$^+$. ^{1}H NMR, ^{13}C NMR, 2D COSY, HMQC, and HMBC spectra were obtained via a Bruker Avance II 400 MHz spectrometer (400 Hz for 1H, 100 Hz for 13C). Chemical shifts are reported in d parts per million (ppm), with multiplicity (b = broad, s = singlet, d = doublet, t = triplet, q = quartet, quin = quintet, hept = heptet, m = multiplet, app = apparent), coupling constants (in Hz) and integration.

3,4-Dichlorophenyl 1-thio-α-D-galactopyranoside tetraacetate (**4**) [63].
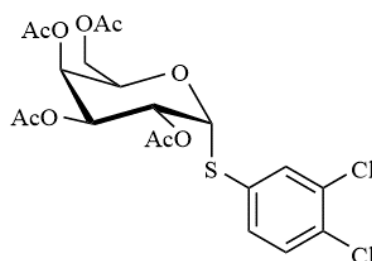
B-D-galactose pentaacetate (5.5239 g, 14.15 mmol) were added to an oven-dried round bottom flask and dissolved in dry DCM (30 ml) under N_2 with molecular sieves. Phosphorous pentachloride (3.25 g, 15.6 mmol) was added followed by a slow addition of $BF_3O(Et)_2$ (0.09 ml, 0.73 mmol) TLC (1:1 heptane/EtOAc) checked after ten minutes. Quenched reaction with 200ml DCM. The organic phase was washed with 100 ml ice cold water, followed by 2x100 ml ice cold saturated $NaHCO_3$ and a final wash with 2x80 ml ice cold water. Dried with Na_2SO_4 then put in vacuo. Purged an oven-dried round bottom flask with N_2. Added NaH (60%, 1.022 g, 25.47 mmol) and dissolved in 20 ml dry DMF under N_2. 3.4-dichlorobenzenethiol (1.24 ml, 2.2494 g [crystallized], 22.3 mmol) was then added. Dissolved the beta-glycoside in 24 ml DMF and added it to the thiol mixture after half an hour. Let the mixture react at 50 °C overnight. Diluted with 200 ml DCM followed by a wash with 100 ml saturated NH_4Cl and 2x100 ml H_2O. Dried over sodium sulfate. Co-evaporated the solvents with toluene and purification with flash chromatography (SiO_2, gradient mobile phase: 3:1 heptane/EtOAc down to 1:1 heptane/EtOAc). 1.981 g (27.5%) of **4** was obtained as yellow powder. **^{1}H NMR** (400

MHz, CDCl$_3$) δ 7.57 (dd, J = 6.1, 2.2 Hz, 1H), 7.40 – 7.31 (m, 2H), 7.29 (ddd, J = 8.4, 3.4, 2.2 Hz, 1H), 5.49 (dd, J = 3.3, 1.3 Hz, 1H), 5.34 (dd, J = 11.0, 5.5 Hz, 1H), 5.25 (dd, J = 11.0, 3.3 Hz, 1H), 4.66 (t, J = 6.4 Hz, 1H), 4.18 – 4.03 (m, 2H), 2.19 – 2.06 (m, 5H), 2.06 – 1.99 (m, 3H), 2.00 (s, 2H), 1.27 (d, J = 9.7 Hz, 2H), 0.90 – 0.81 (m, 1H).

3,4-Dichlorophenyl-1-thio-a-D-galactoside (**5**) [46].
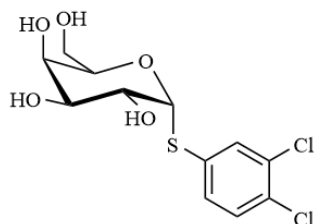
Compound **4** (1981 mg, 3.889 mmol) was dissolved in dry methanol (26 ml) followed by addition of sodium methoxide (9.2 ml, 0.5 M in MeOH) and left to stir overnight. TLC (1:1 heptane/EtOAc) was used to check for completion. Filtered out the white precipitate and dried with clean DOWEX. Purified using flash chromatography (SiO$_2$, gradient mobile phase: 2:98 MeOH/DCM followed by 5:95 MeOH/DCM and finally 10:90 MeOH/DCM). Solvents were then evaporated. 873 mg (66%) of **5** was obtained as a yellow powder. **^{1}H NMR** (400 MHz, MeOD) δ 7.74 (d, J = 2.0 Hz, 1H), 7.53 – 7.42 (m, 2H), 5.67 (d, J = 5.5 Hz, 1H), 4.30 (ddd, J = 6.8, 5.2, 1.3 Hz, 1H), 4.20 (dd, J = 10.2, 5.5 Hz, 1H), 3.99 (dd, J = 3.3, 1.2 Hz, 1H), 3.82 – 3.58 (m, 3H).

3,4-Dichlorophenyl-3-O-[O-methyl(carboxymethyl)]-1-thio-a-D-galactoside (**6**) [54].
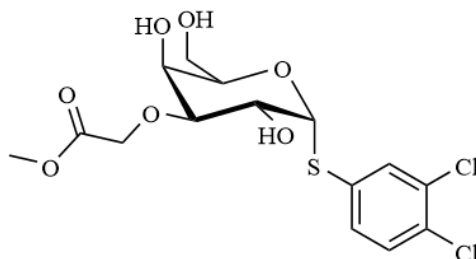
Compound **5** (861.7 mg, 2.525 mmol) was dissolved in 11 ml mixture of 10:1 ACN/DMF, followed by methyl bromoacetate (0.6 ml, 6.3 mmol), dibutyltin dichloride (80 mg, 0.26 mmol), TBAB (92.6 mg, 28.7 mmol), and potassium carbonate (504.6 mg, 3.94 mmol). Reaction was run at 80°C overnight then checked with TLC (95:5 DCM/MeOH). Solvents were then co-evaporated with toluene and purified using flash chromatography (SiO$_2$, mobile phase: 2:98 MeOH/DCM followed by 5:95 MeOH/DCM). 485 mg (46%) of **6** was obtained as yellow

powder.

Experiment was repeated under the same conditions using unreacted compound **5** (347 mg, 1.02 mmol) retrieved from the column with 279.9 mg (66%) of **6** was obtained as yellow powder. **¹H NMR** (400 MHz, CDCl₃) δ 8.04 (s, 1H), 3.87 – 3.73 (m, 1H), 2.98 (d, J = 0.5 Hz, 2H), 2.91 (d, J = 0.6 Hz, 2H), 1.28 (s, 1H), 0.86 (s, 3H), 0.09 (s, 5H).

3,4-Dichlorophenyl-3-O-[(4-bromo-1H-benzoimidazole-2-yl)-methyl]-1-thio-a-D-galactoside (**7**) [64]
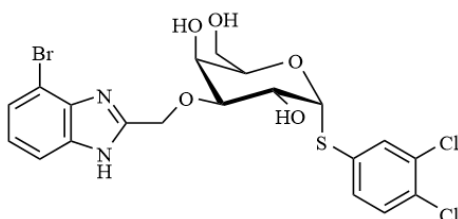
Added 3-bromo-1,2-diaminobenzene (655.2 mg, 3.503 mmol) to the round-bottom flask containing compound **6** (717.4 mg, 1.736 mmol) followed by 33 ml of water. Let the reaction run for 24h at 100°C under reflux. TLC (5:95 MeOH/DCM) was checked for completion of reaction. Solvents were evaporated and purification was done using flash chromatography (SiO₂, gradient mobile phase: 1:99 MeOH/DCM, finishing at 4:96 MeOH/DCM). 467.5 mg (48.9%) of **7** was obtained as yellow powder. **¹H NMR** (400 MHz, MeOD) δ 7.80 – 7.69 (m, 1H), 7.58 – 7.50 (m, 1H), 7.53 – 7.38 (m, 3H), 7.17 (td, J = 7.9, 4.7 Hz, 1H), 5.73 (d, J = 5.6 Hz, 1H), 5.14 – 4.94 (m, 2H), 4.46 (dd, J = 10.1, 5.6 Hz, 1H), 4.41 – 4.19 (m, 2H), 3.92 – 3.66 (m, 3H).

3,4-Dichlorophenyl-3-O-[(4-(4-chlorophenyl)-1H-benzoimidazole-2-yl)-methyl]-1-thio-a-D-galactoside (**2a**) [65]
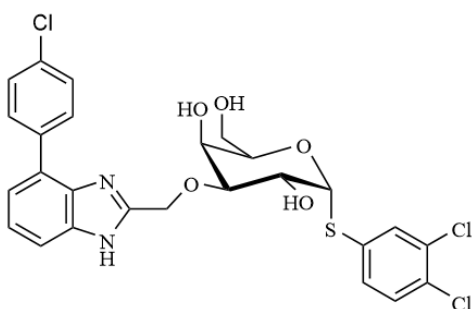
Potassium carbonate (18.6 mg, 0.134 mmol), prep HPLC purified compound 7 (19.9 mg, 0.0362 mmol), 4-chlorophenylboronic acid (8.0 mg, 0.051 mmol), and Pd(PPh3)4 (6.7 mg, 0.0058 mmol) were added in order to a microwave vial. The vial was sealed and put in vacuo, followed by a 10-minute nitrogen flush. A round-bottom flask was sealed and degassed, flushed with nitrogen, followed by the addition of a 1:4 H2O/1,4-dioxane solvent mixture. 1 ml of the solvent mixture was then added to the microwave vial and the reaction ran in a microwave at 110°C, high, 1.5h. The mixture was filtered through celite, and solvents were evaporated. Purification was performed by prep HPLC. 10.1 mg (48.0%) of **2a** was obtained as white powder. **^{1}H NMR** (400 MHz, MeOD) δ 7.81 – 7.69 (m, 3H), 7.58 (dd, J = 6.9, 2.2 Hz, 1H), 7.58 – 7.42 (m, 5H), 7.39 – 7.27 (m, 3H), 5.73 (d, J = 5.6 Hz, 1H), 5.06 (d, J = 14.0 Hz, 1H), 4.99 (d, J = 14.1 Hz, 1H), 4.42 (dd, J = 10.2, 5.6 Hz, 1H), 4.34 – 4.19 (m, 2H), 3.87 – 3.74 (m, 1H), 3.84 – 3.65 (m, 2H). **^{13}C NMR** (101 MHz, MeOD) δ 152.95, 133.04, 131.31, 130.27, 129.95, 128.46, 122.72, 121.98, 89.57, 79.93, 72.05, 67.45, 65.83, 64.22, 60.98, 47.38, 47.17, 46.96. **HRMS** calculated for $C_{26}H_{23}N_2O_5Cl_3S$ – H-: 579.0315, found: 579.0300

3,4-Dichlorophenyl-3-O-[(4-(4-Carboxyphenyl)-1H-benzoimidazole-2-yl)-methyl]-1-thio-a-D-galactoside (**2b**) [65].
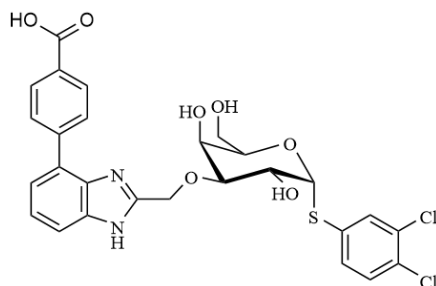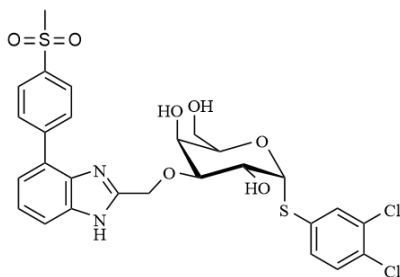

Potassium carbonate (17.4 mg, 0.126 mmol), prep HPLC purified compound **7** (20.6 mg, 0.0374 mmol), 4-carboxyphenylboronic acid (10.2 mg, 0.0607 mmol), and Pd(PPh3)4 (4.2 mg, 0.0036 mmol) were added in order to a microwave vial. The vial was sealed and put in vacuo, followed by a 10-minute nitrogen flush. A round-bottom flask was sealed and degassed, flushed with nitrogen, followed by the addition of a 1:4 H2O/1,4-dioxane solvent mixture. 1 ml of the solvent mixture was then added to the microwave vial and the reaction ran in a microwave at 110°C, high, 1.5h. The mixture was filtered through celite, and solvents were evaporated. Purification was performed by prep HPLC. 9.4 mg (48.0%) of **2b** was obtained as white powder. **^{1}H NMR** (400 MHz, MeOD) δ 8.22 – 8.11 (m, 2H), 7.93 – 7.83 (m, 2H), 7.81 – 7.73 (m, 1H), 7.61 (td, J = 7.4, 1.9 Hz, 1H), 7.53 – 7.29 (m, 4H), 5.72 (d, J = 5.6 Hz, 1H), 5.03 (q,

J = 14.1 Hz, 2H), 4.42 (dd, J = 10.2, 5.6 Hz, 1H), 4.38 – 4.28 (m, 1H), 4.29 (s, 1H), 3.87 – 3.68 (m, 3H). **^{13}C NMR** (101 MHz, MeOD) δ 133.08, 131.37, 130.25, 129.80, 128.40, 122.72, 122.26, 89.62, 79.92, 72.06, 67.45, 65.83, 64.21, 61.00, 47.38, 47.17, 46.96. **HRMS** calculated for $C_{27}H_{24}N_2O_7Cl_2S - H^-$: 589.0603, found: 589.0608

3,4-Dichlorophenyl-3-O-[(4-(4-aminocarbonyl)phenyl)-2-1H-benzimidazolylmethyl]-1-thio-a-D-galactoside (**2c**) [65].
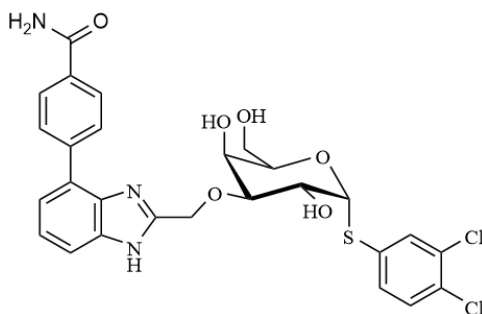
Potassium carbonate (15.6 mg, 0.113 mmol), prep HPLC purified compound **7** (17.5 mg, 0.0318 mmol), 4-carbamoylbenzeneboronic acid (7.9 mg, 0.048 mmol), and Pd(PPh$_3$)$_4$ (4.9 mg, 0.0042 mmol) were added in order to a microwave vial. The vial was sealed and put in vacuo, followed by a 10-minute nitrogen flush. A round-bottom flask was sealed and degassed, flushed with nitrogen, followed by the addition of a 1:4 H_2O/1,4-dioxane solvent mixture. 1 ml of the solvent mixture was then added to the microwave vial and the reaction ran in a microwave at 110°C, high, 1.5h. The mixture was f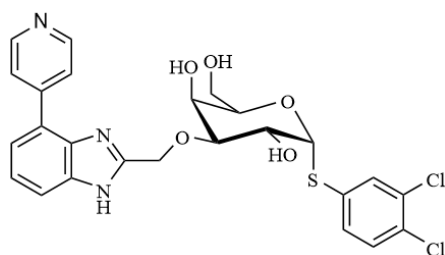iltered through celite, and solvents were evaporated. Purification was performed twice by prep HPLC. 5.6 mg (29.8%) of **2c** was obtained as white powder. **^{1}H NMR** (400 MHz, MeOD) δ 8.08 – 7.99 (m, 2H), 7.90 (d, J = 8.1 Hz, 2H), 7.75 (d, J = 2.0 Hz, 1H), 7.65 – 7.56 (m, 1H), 7.53 – 7.29 (m, 4H), 5.73 (d, J = 5.5 Hz, 1H), 5.07 (d, J = 14.0 Hz, 1H), 5.00 (d, J = 14.1 Hz, 1H), 4.43 (dd, J = 10.1, 5.6 Hz, 1H), 4.34 – 4.22 (m, 2H), 3.78 (dd, J = 11.5, 5.3 Hz, 1H), 3.77 – 3.67 (m, 2H). **^{13}C NMR** (101 MHz, MeOD) δ 135.34, 133.07, 131.35, 130.26, 128.50, 127.75, 122.72, 89.58, 79.95, 72.05, 67.47, 65.85, 60.99, 47.88. **HRMS** calculated for $C_{27}H_{25}N_3O_6Cl_2S + H^+$: 590.0919, found: 590.0910

3,4-Dichlorophenyl-3-O-[(4-(3-(hydroxymethyl)phenyl)-1H-benzoimidazole-2-yl)-methyl]-1-thio-a-D-galactoside (**2d**) [65].
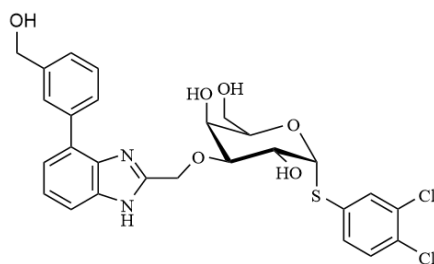
Potassium carbonate (20.9 mg, 0.151 mmol), prep HPLC purified compound **7** (20.0 mg, 0.0363 mmol), 3-(hydroxymethyl)phenylboronic acid (8.6 mg, 0.057 mmol), and Pd(PPh$_3$)$_4$ (5.8 mg, 0.0050 mmol) were added in order to a microwave vial. The vial was sealed and put in vacuo, followed by a 10-minute nitrogen flush. A round-bottom flask was sealed and degassed, flushed with nitrogen, followed by the addition of a 1:4 H$_2$O/1,4-dioxane solvent mixture. 1 ml of the solvent mixture was then added to the microwave vial and the reaction ran in a microwave at 110°C, high, 1.5h. The mixture was filtered through celite, and solvents were evaporated. Purification was performed twice by prep HPLC. 6.0 mg (28.6%) of **2d** was obtained as white powder. **^{1}H NMR** (400 MHz, MeOD) δ 7.76 (dd, J = 8.8, 1.9 Hz, 2H), 7.66 (dt, J = 7.6, 1.5 Hz, 1H), 7.63 – 7.55 (m, 1H), 7.55 – 7.30 (m, 6H), 5.72 (d, J = 5.6 Hz, 1H), 5.07 (d, J = 14.0 Hz, 1H), 4.99 (d, J = 14.0 Hz, 1H), 4.75 (s, 2H), 4.42 (dd, J = 10.1, 5.6 Hz, 1H), 4.34 – 4.26 (m, 2H), 3.82 – 3.67 (m, 3H). **^{13}C NMR** (101 MHz, MeOD) δ 133.04, 131.32, 130.26, 63.94. **HRMS** calculated for C$_{27}$H$_{26}$N$_2$O$_6$Cl$_2$S – H$^-$: 575.0810, found: 575.0797

3,4-Dichlorophenyl-3-O-[(4-(4-(methylsulfonyl)phenyl)-1H-benzoimidazole-2-yl)-methyl]-1-thio-a-D-galactoside (**2e**) [65].

Potassium carbonate (16.0 mg, 0.116 mmol), prep HPLC purified compound **7** (19.7 mg, 0.0358 mmol), 4-methanesulfonylphenylboronic acid (9.7 mg, 0.048 mmol), and Pd(PPh$_3$)$_4$ (8.4 mg, 0.0073 mmol) were added in order to a microwave vial. The vial was sealed and put in vacuo, followed by a 10-minute nitrogen flush. A round-bottom flask was sealed and degassed, flushed with nitrogen, followed by the addition of a 1:4 H$_2$O/1,4-dioxane solvent mixture. 1 ml of the solvent mixture was then added to the microwave vial and the reaction ran in a microwave at 110°C, high, 1.5h. The mixture was filtered through celite, and solvents were

evaporated. Purification was performed twice by prep HPLC. 7.1 mg (31.5%) of **2e** was obtained as white powder. **¹H NMR** (400 MHz, MeOD) δ 8.29 (s, 0H), 8.14 – 8.03 (m, 4H), 7.75 (d, J = 2.0 Hz, 1H), 7.68 – 7.56 (m, 1H), 7.52 – 7.36 (m, 3H), 5.73 (d, J = 5.6 Hz, 1H), 5.13 – 4.94 (m, 2H), 4.43 (dd, J = 10.2, 5.6 Hz, 1H), 4.31 (dd, J = 6.7, 4.7 Hz, 2H), 3.82 – 3.68 (m, 3H), 3.20 (s, 3H). **¹³C NMR** (101 MHz, MeOD) δ 133.08, 131.36, 130.28, 129.45, 127.40, 122.81, 79.97, 72.04, 67.49, 65.87, 60.98, 43.10. **HRMS** calculated for $C_{27}H_{26}N_2O_7Cl_2S_2 + H^+$: 625.0637, found: 625.0652

3,4-Dichlorophenyl-3-O-[(4-(4-(1H-pyrazol-5-yl)phenyl)-1H-benzoimidazole-2-yl)-methyl]-1-thio-a-D-galactoside (**2f**) [65].
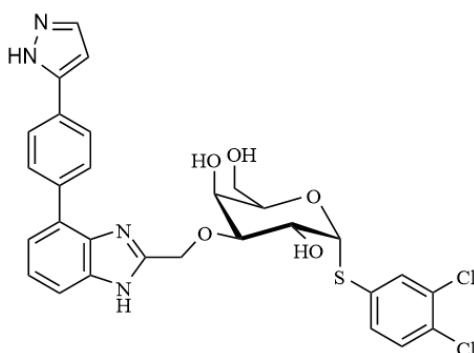
Potassium carbonate (16.8 mg, 0.122 mmol), prep HPLC purified compound **7** (20.0 mg, 0.0363 mmol), [4-(1H-Pyrazol-5-yl)phenyl]boronic acid (9.0 mg, 0.048 mmol), and Pd(PPh₃)₄ (5.5 mg, 0.0048 mmol) were added in or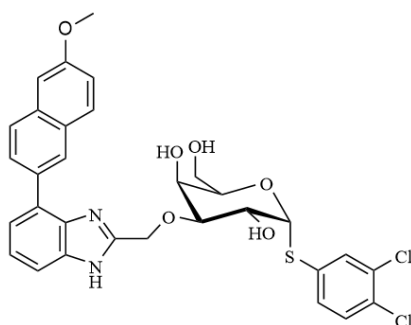der to a microwave vial. The vial was sealed and put in vacuo, followed by a 10-minute nitrogen flush. A round-bottom flask was sealed and degassed, flushed with nitrogen, followed by the addition of a 1:4 H_2O/1,4-dioxane solvent mixture. 1 ml of the solvent mixture was then added to the microwave vial and the reaction ran in a microwave at 110°C, high, 1.5h. The mixture was filtered through celite, and solvents were evaporated. Purification was performed twice by prep HPLC. 6.2 mg (28.1%) of **2f** was obtained as white powder. **¹H NMR** (400 MHz, MeOD) δ 7.97 – 7.89 (m, 2H), 7.85 (d, J = 8.3 Hz, 2H), 7.75 (d, J = 2.0 Hz, 1H), 7.71 (d, J = 2.3 Hz, 1H), 7.59 (dd, J = 7.1, 2.1 Hz, 1H), 7.52 – 7.42 (m, 2H), 7.42 – 7.29 (m, 2H), 6.75 (d, J = 2.2 Hz, 1H), 5.72 (d, J = 5.6 Hz, 1H), 5.07 (d, J = 14.0 Hz, 1H), 5.00 (d, J = 14.0 Hz, 1H), 4.42 (dd, J = 10.1, 5.6 Hz, 1H), 4.30 (dd, J = 6.5, 4.7 Hz, 2H), 3.82 – 3.68 (m, 3H). **¹³C NMR** (101 MHz, MeOD) δ 133.04, 131.34, 130.25, 125.72, 102.01. **HRMS** calculated for: $C_{27}H_{26}N_2O_6Cl_2S - H^-$: 611.0923, found: 611.0928

3,4-Dichlorophenyl-3-O-[(4-(pyridine-4-yl)-1H-benzoimidazol-2-yl)-methyl]-1-thio-a-D-galactoside (**2g**) [65].

Potassium carbonate (25.1 mg, 0.182 mmol), prep HPLC purified compound **7** (20.0 mg, 0.0363 mmol), 4-Pyridinylboronic acid (6.5 mg, 0.053 mmol), and Pd(PPh$_3$)$_4$ (5.8 mg, 0.0050 mmol) were added in order to a microwave vial. The vial was sealed and put in vacuo, followed by a 10-minute nitrogen flush. A round-bottom flask was sealed and degassed, flushed with nitrogen, followed by the addition of a 1:4 H$_2$O/1,4-dioxane solvent mixture. 1 ml of the solvent mixture was then added to the microwave vial and the reaction ran in a microwave at 110°C, high, 1.5h. The mixture was filtered through celite, and solvents were evaporated. Purification was performed twice by prep HPLC. 5.2 mg (26.3%) of **2g** was obtained as white powder. **^{1}H NMR** (400 MHz, DMSO) δ 12.63 (s, 1H), 8.66 (d, J = 5.0 Hz, 2H), 8.15 (d, J = 5.2 Hz, 2H), 7.76 (d, J = 2.1 Hz, 1H), 7.66 – 7.53 (m, 3H), 7.46 (dd, J = 8.4, 2.2 Hz, 1H), 7.35 (td, J = 7.7, 4.2 Hz, 1H), 5.78 (d, J = 5.5 Hz, 1H), 5.70 (s, 1H), 5.05 – 4.87 (m, 3H), 4.67 (s, 1H), 4.31 – 4.22 (m, 1H), 4.18 (s, 1H), 4.13 (s, 0H), 4.04 (t, J = 6.3 Hz, 1H), 3.65 – 3.49 (m, 3H), 2.01 (q, J = 7.6 Hz, 0H), 1.27 – 1.22 (m, 2H). **^{13}C NMR** (101 MHz, DMSO) δ 132.45, 131.33, 131.05, 60.60. **HRMS** calculated for C$_{25}$H$_{23}$N$_3$O$_5$Cl$_2$S + H$^+$: 548.0814, found: 548.0822

3,4-Dichlorophenyl-3-O-[(4-(-4(dimethylamino)phenyl)-1H-benzoimidazol-2-yl)-methyl]-1-thio-D-galactoside (**2h**) [65].
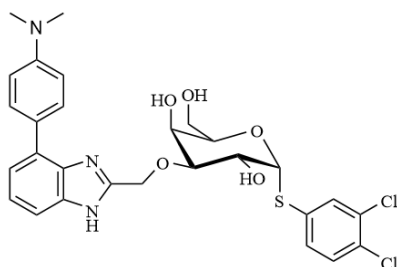
Potassium carbonate (18.3 mg, 0.132 mmol), prep HPLC purified compound **7** (19.2 mg, 0.0349 mmol), 4-(dimethylamino)phenylboronic acid (10.5 mg, 0.0636 mmol), and Pd(PPh$_3$)$_4$ (8.2 mg, 0.0071 mmol) were added in order to a microwave vial. The vial was sealed and put

in vacuo, followed by a 10-minute nitrogen flush. A round-bottom flask was sealed and degassed, flushed with nitrogen, followed by the addition of a 1:4 H_2O/1,4-dioxane solvent mixture. 1 ml of the solvent mixture was then added to the microwave vial and the reaction ran in a microwave at 110°C, high, 1.5h. The mixture was filtered through celite, and solvents were evaporated. Purification was performed twice by prep HPLC. 3.0 mg (14.1%) of **2h** was obtained as white powder. **¹H NMR** (400 MHz, MeOD) δ 7.75 (d, J = 2.0 Hz, 1H), 7.64 (d, J = 8.4 Hz, 2H), 7.53 – 7.42 (m, 3H), 7.34 – 7.24 (m, 2H), 6.98 – 6.88 (m, 2H), 5.75 – 5.65 (m, 1H), 5.09 – 5.01 (m, 1H), 4.98 (d, J = 13.8 Hz, 1H), 4.91 (s, 1H), 4.59 (s, 1H), 4.42 (dd, J = 10.1, 5.6 Hz, 1H), 4.37 – 4.27 (m, 2H), 3.89 – 3.61 (m, 2H), 3.02 (s, 6H). **HRMS** calculated for $C_{28}H_{29}N_3O_5Cl_2S - H^-$: 588.1127, found: 588.1114

3,4-Dichlorophenyl-3-O-[(4-(-4(tert-butyl)phenyl)-1H-benzoimidazole-2-yl)-methyl]-1-thio-a-D-galactoside (**2i**) [65].
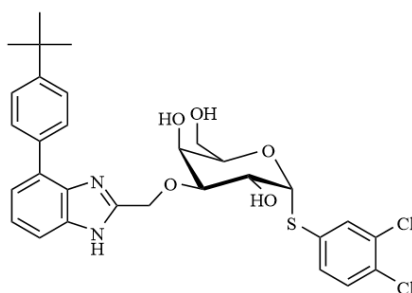
Potassium carbonate (17.5 mg, 0.127 mmol), prep HPLC purified compound **7** (19.6 mg, 0.0356 mmol), 4-tert-butylphenylboronic acid (8.9 mg, 0.050 mmol), and $Pd(PPh_3)_4$ (7.3 mg, 0.0063 mmol) were added in order to a microwave vial. The vial was sealed and put in vacuo, followed by a 10-minute nitrogen flush. A round-bottom flask was sealed and degassed, flushed with nitrogen, followed by the addition of a 1:4 H_2O/1,4-dioxane solvent mixture. 1 ml of the solvent mixture was then added to the microwave vial and the reaction ran in a microwave at 110°C, high, 1.5h. The mixture was filtered through celite, and solvents were evaporated. Purification was performed by prep HPLC. 5.5 mg (25.3%) of **2i** was obtained as white powder. **1H NMR** (400 MHz, MeOD) δ 7.75 (d, J = 1.9 Hz, 1H), 7.69 (d, J = 8.2 Hz, 2H), 7.60 – 7.53 (m, 3H), 7.52 – 7.41 (m, 2H), 7.38 – 7.29 (m, 2H), 5.72 (d, J = 5.6 Hz, 1H), 5.04 (d, J = 13.7 Hz, 1H), 4.97 (d, J = 13.6 Hz, 1H), 4.42 (dd, J = 10.1, 5.6 Hz, 1H), 4.30 (d, J = 6.0 Hz, 1H), 3.84 – 3.66 (m, 2H), 1.40 (s, 7H), 1.41 – 1.35 (m, 1H), 1.35 – 1.27 (m, 4H). **¹³C NMR** (101 MHz, MeOD) δ 125.34, 72.06, 30.40. **HRMS** calculated for $C_{30}H_{32}N_2O_5Cl_2S - H^-$: 601.1331, found: 601.1315

(3,4-dichlorophenyl)-3-O-[(4-(6-methoxynaphthalen-2-yl)-1H-benzoimidazole-2-yl)-methyl]-1-thio-a-D-galactoside (**2j**) [65].

Potassium carbonate (17.5 mg, 0.127 mmol), prep HPLC purified compound **7** (20.1 mg, 0.0365 mmol), 6-methoxy-2-naphthaleneboronic acid (9.8 mg, 0.049 mmol), and Pd(PPh$_3$)$_4$ (10 mg, 0.0087 mmol) was added in order to a microwave vial. The vial was sealed and put in vacuo, followed by a 10-minute nitrogen flush. A round-bottom flask was sealed and degassed, flushed with nitrogen, followed by the addition of a 1:4 H$_2$O/1,4-dioxane solvent mixture. 1 ml of the solvent mixture was then added to the microwave vial and the reaction ran in a microwave at 110°C, high, 1.5h. The mixture was filtered through celite, and solvents were evaporated. Purification was performed twice by prep HPLC. 4.4 mg (26.7%) of **2j** was obtained as white powder. **HRMS** calculated for C$_{31}$H$_{28}$N$_2$O$_6$Cl$_2$S − H$^-$: 625.0967, found: 625.0986

Appendix II: NMR spectra

¹H NMR spectra

Figure 9. *¹H NMR spectrum of **compound 4** at 400 MHz, CDCl₃*
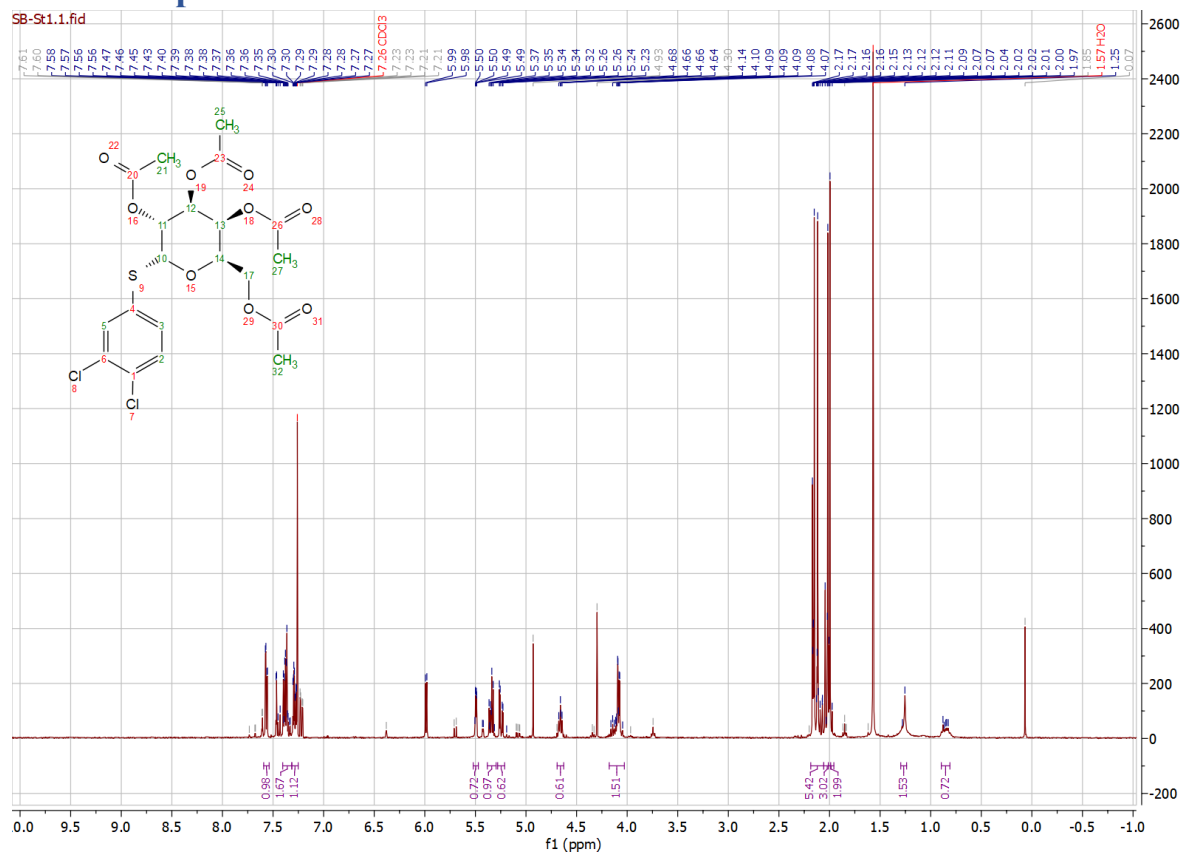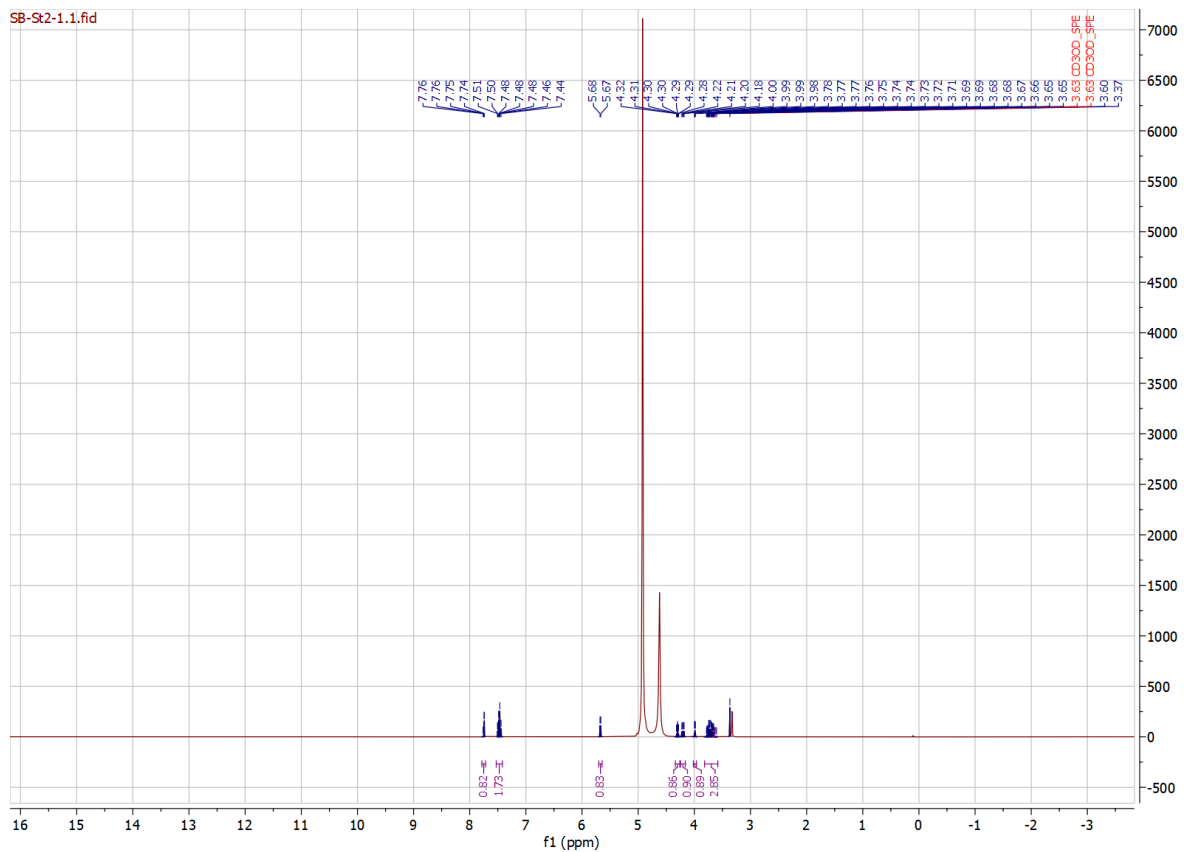
Figure 10. *¹H NMR spectrum of* ***compound 5*** *at 400 MHz, MeOD*
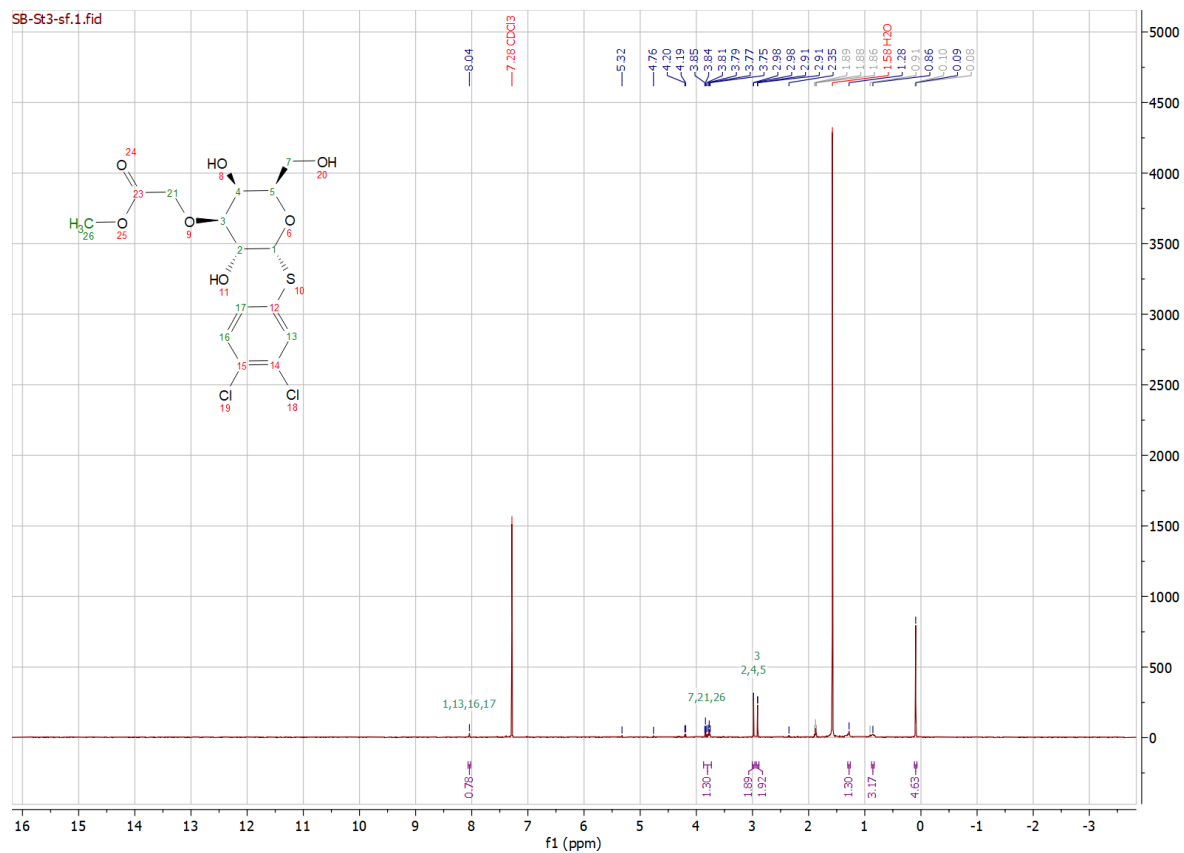
Figure 11. *^{1}H NMR spectrum of **compound 6** at 400 MHz, CDCl₃*
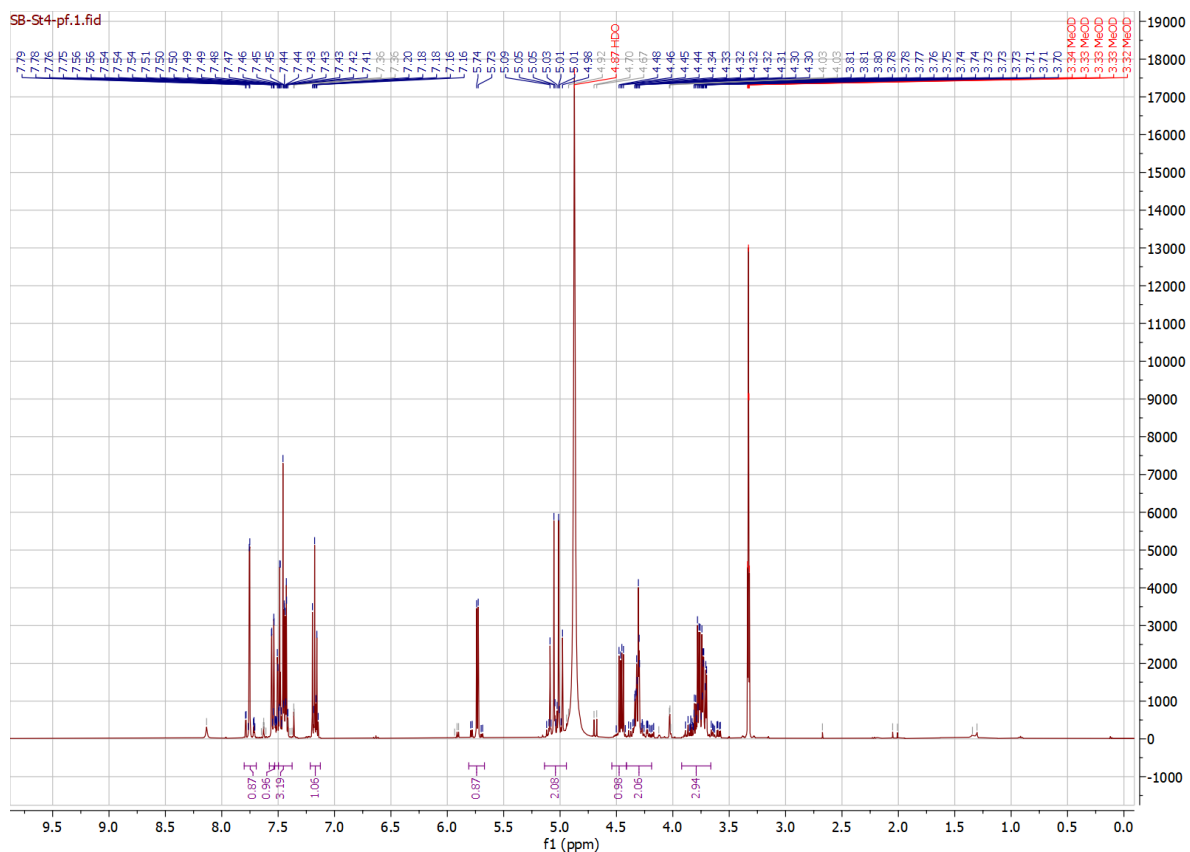
Figure 12. *¹H NMR spectrum of* **compound 7** *at 400 MHz, MeOD*
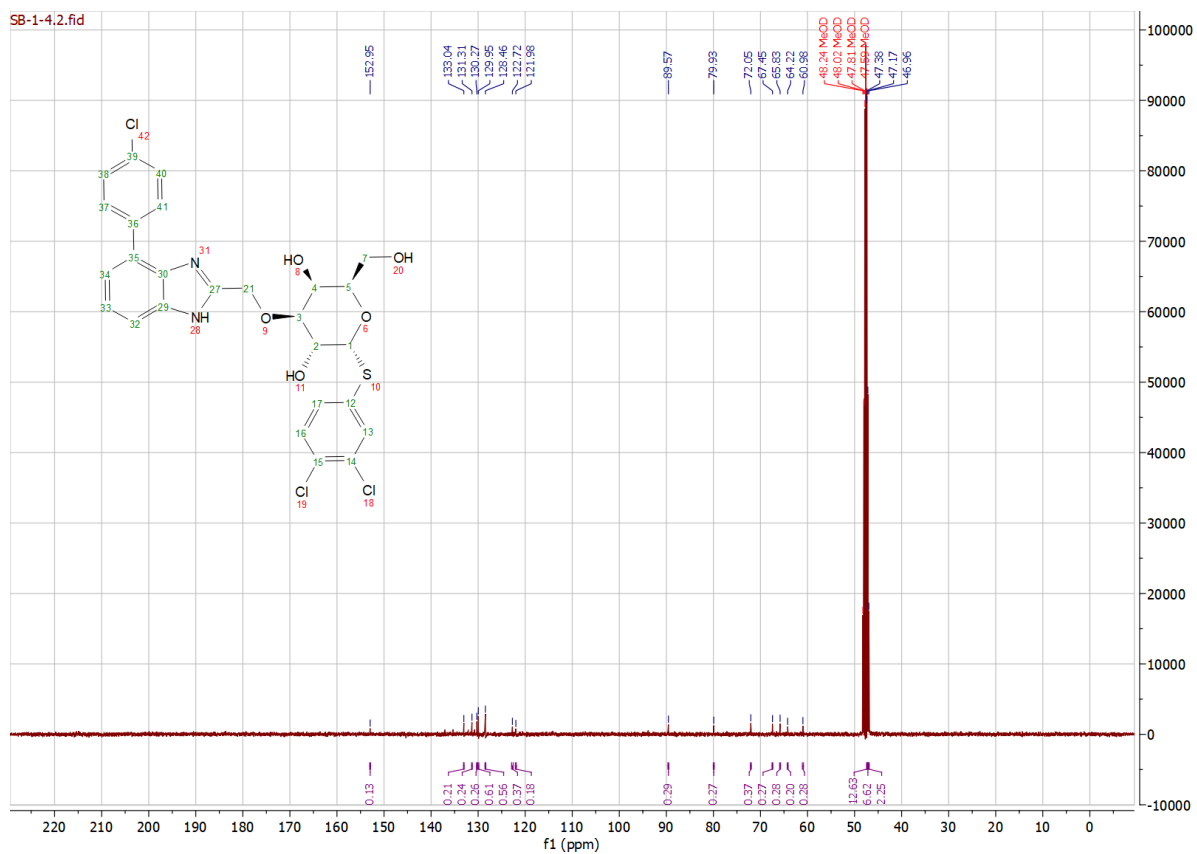
Figure 13. ^{1}H NMR spectrum of **compound 2a** at 400 MHz, MeOD
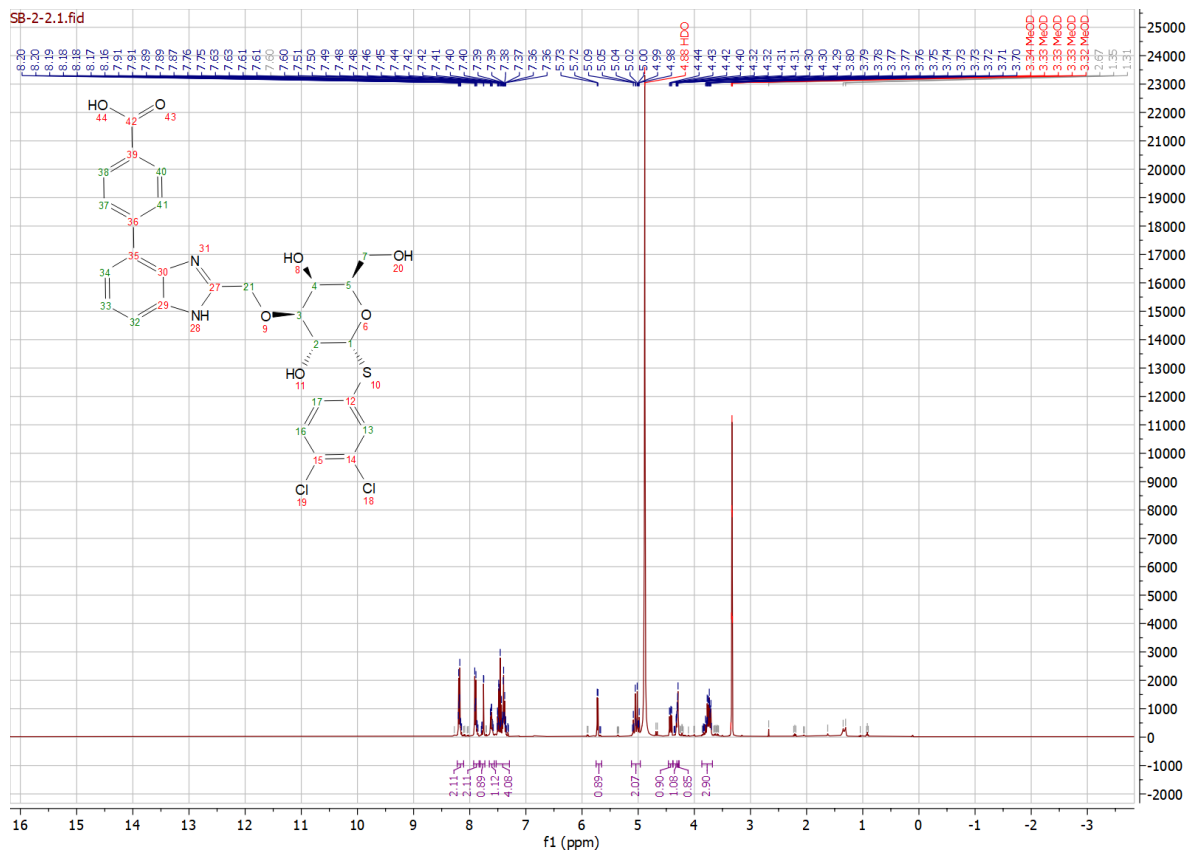
Figure 14. *${}^1H$ NMR spectrum of **compound 2b** at 400 MHz, MeOD*
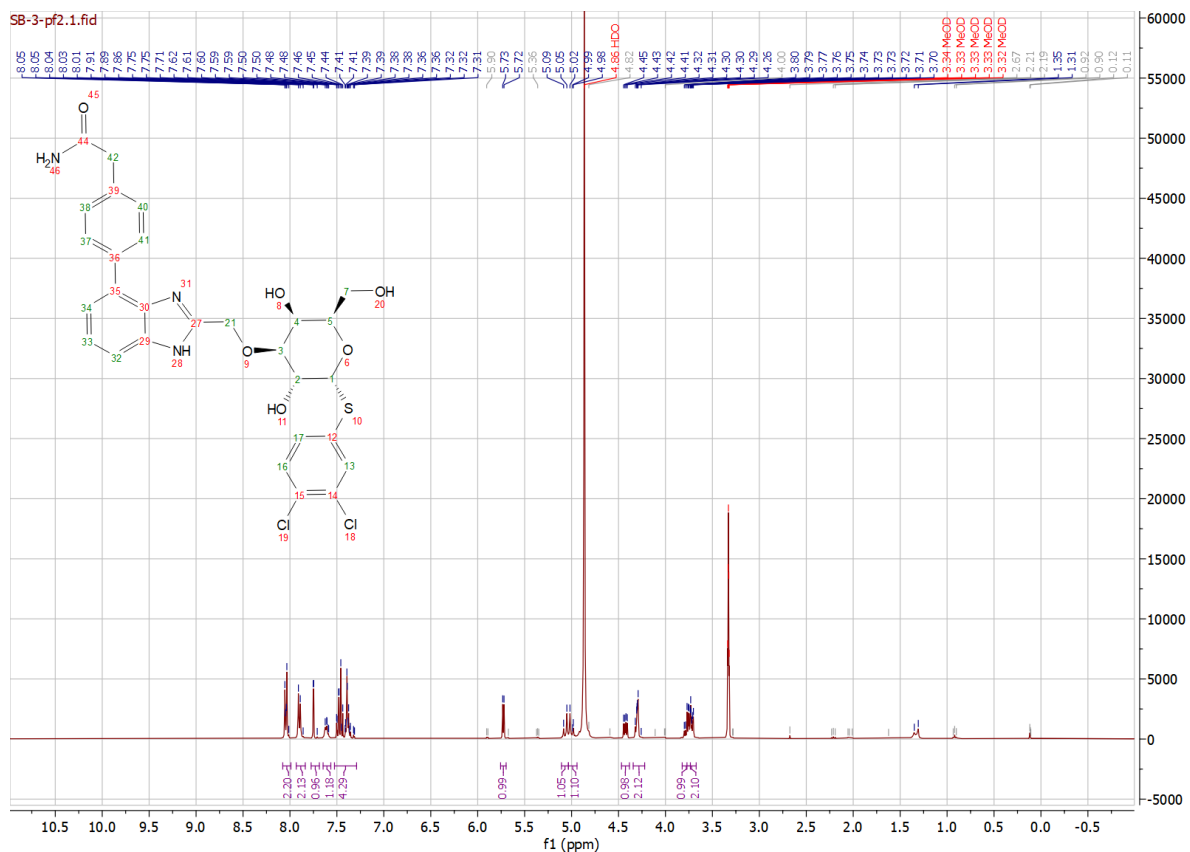
Figure 15. *¹H NMR spectrum of* **compound 2c** *at 400 MHz, MeOD*
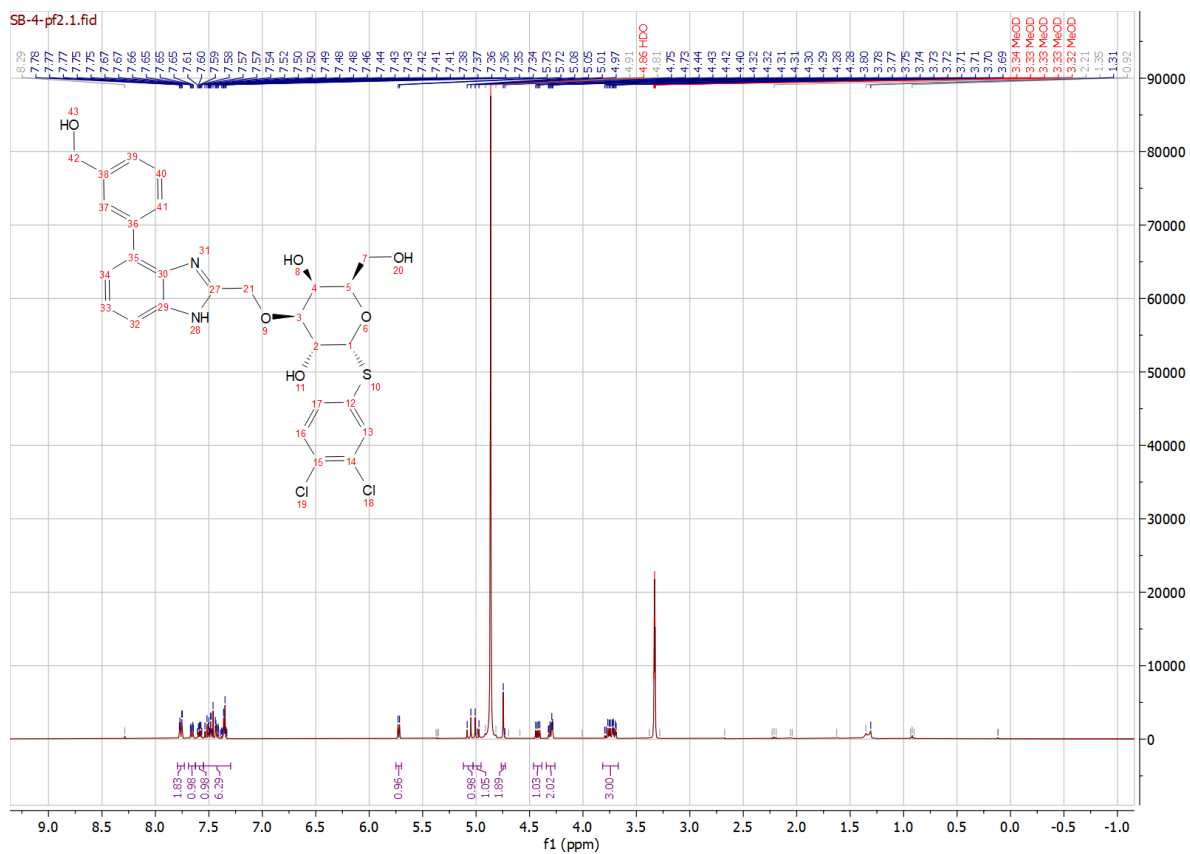
Figure 16. *¹H NMR spectrum of* **compound 2d** *at 400 MHz, MeOD*
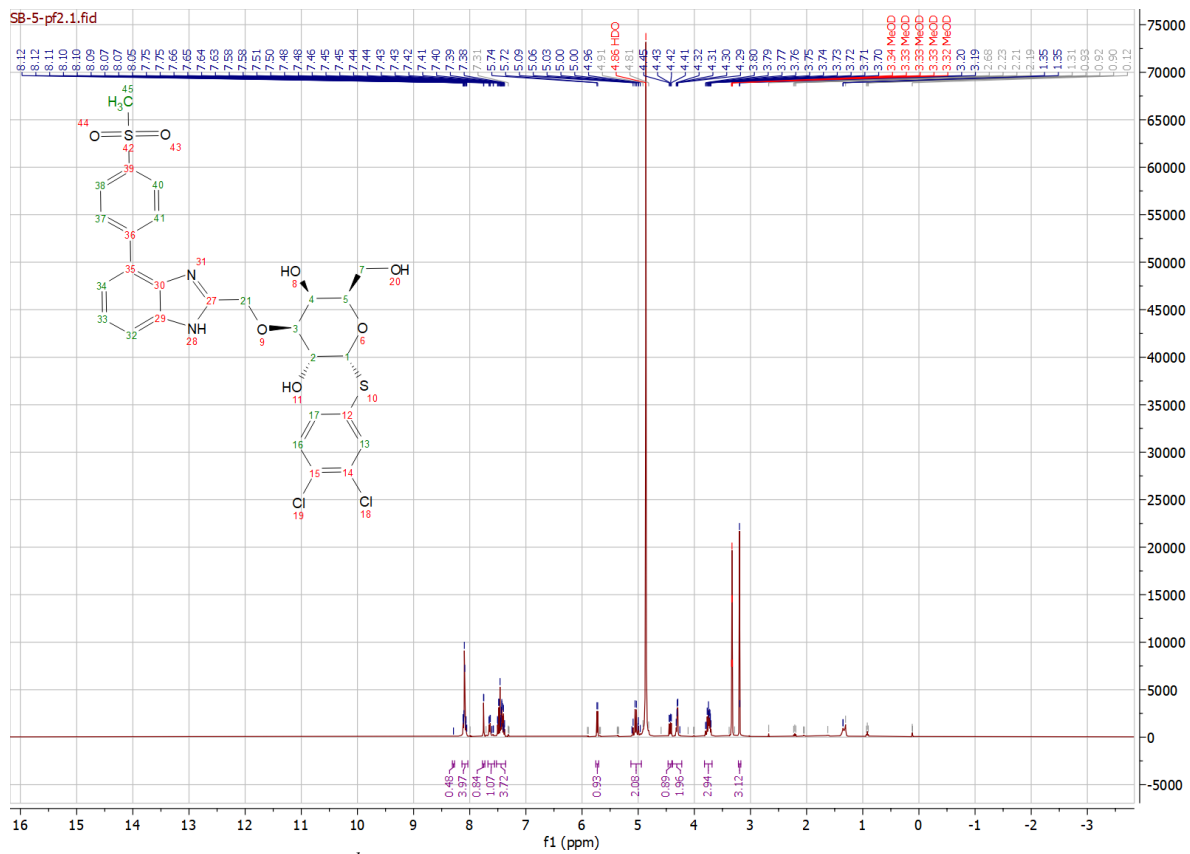
Figure 17. ^{1}H NMR spectrum of **compound 2e** at 400 MHz, MeOD
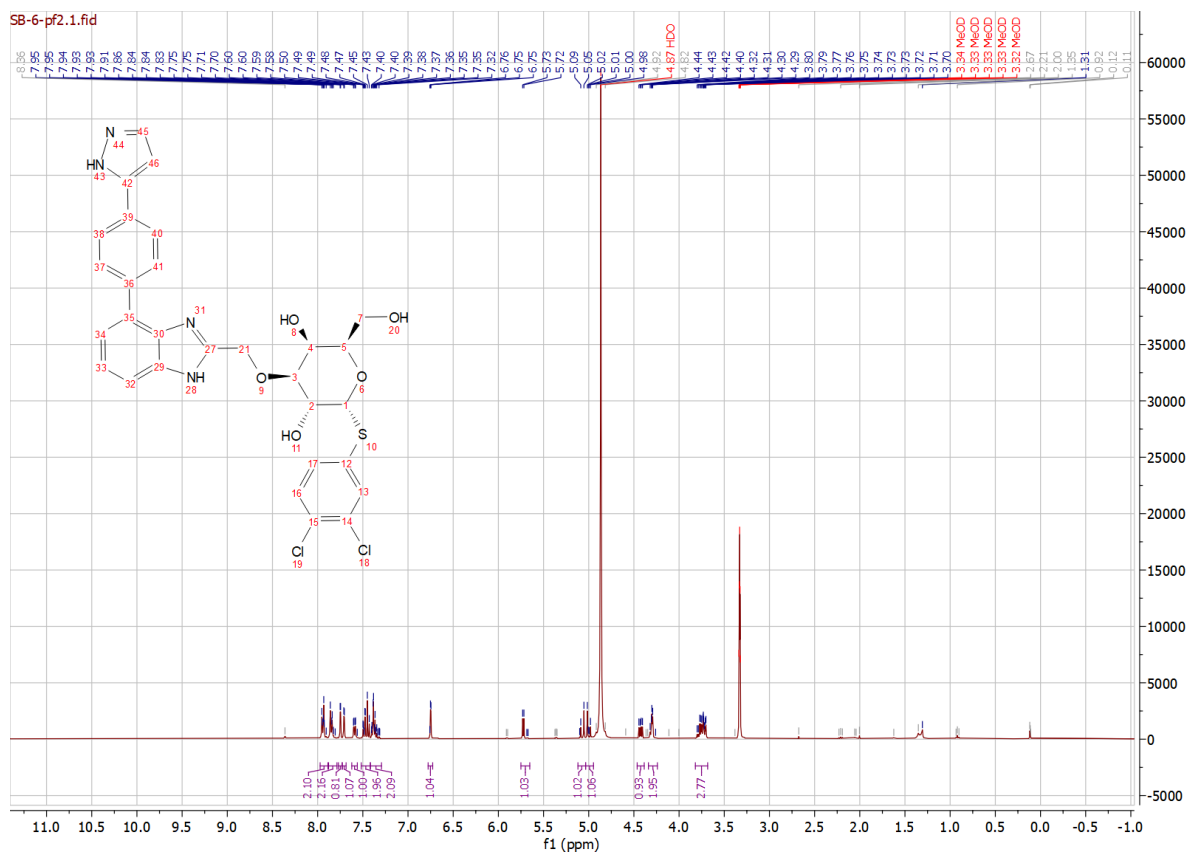
Figure 18. *^{1}H NMR spectrum of* **compound 2f** *at 400 MHz, MeOD*
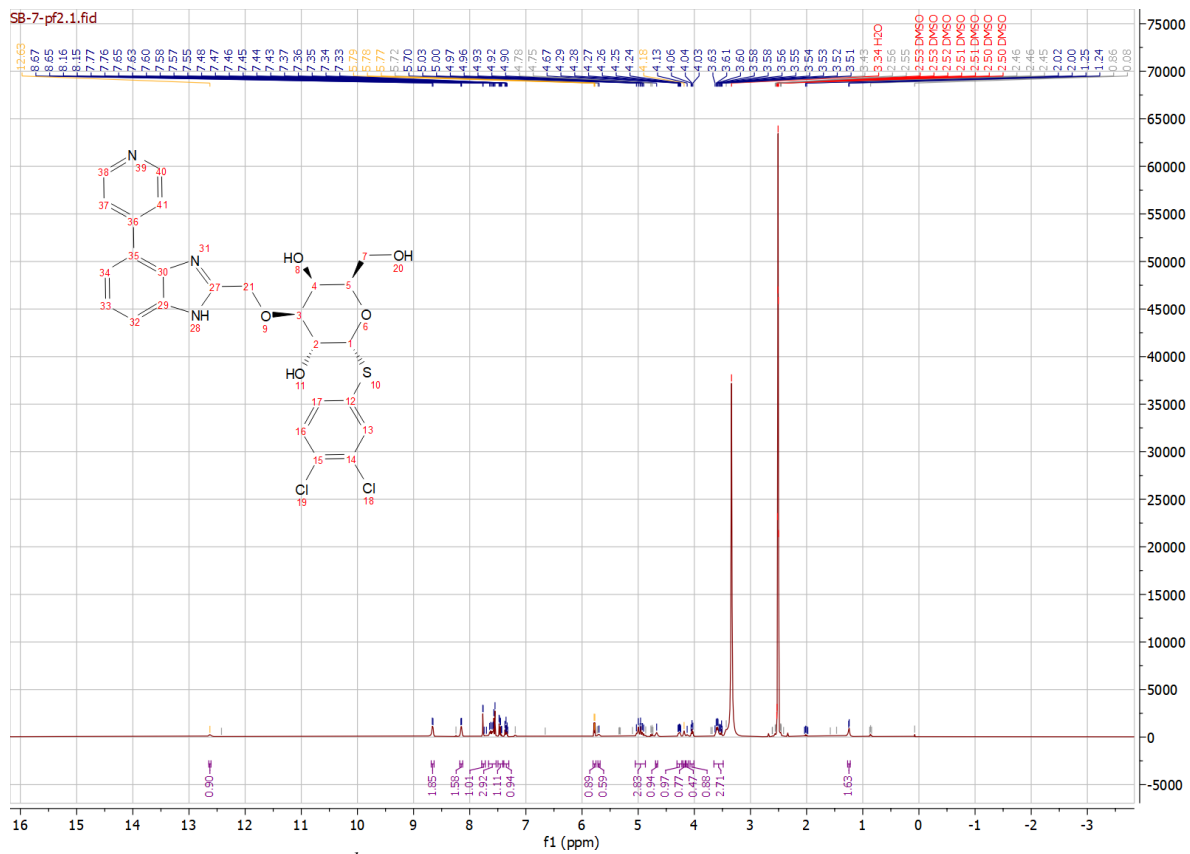
Figure 19. ^{1}H NMR spectrum of **compound 2g** at 400 MHz, DMSO
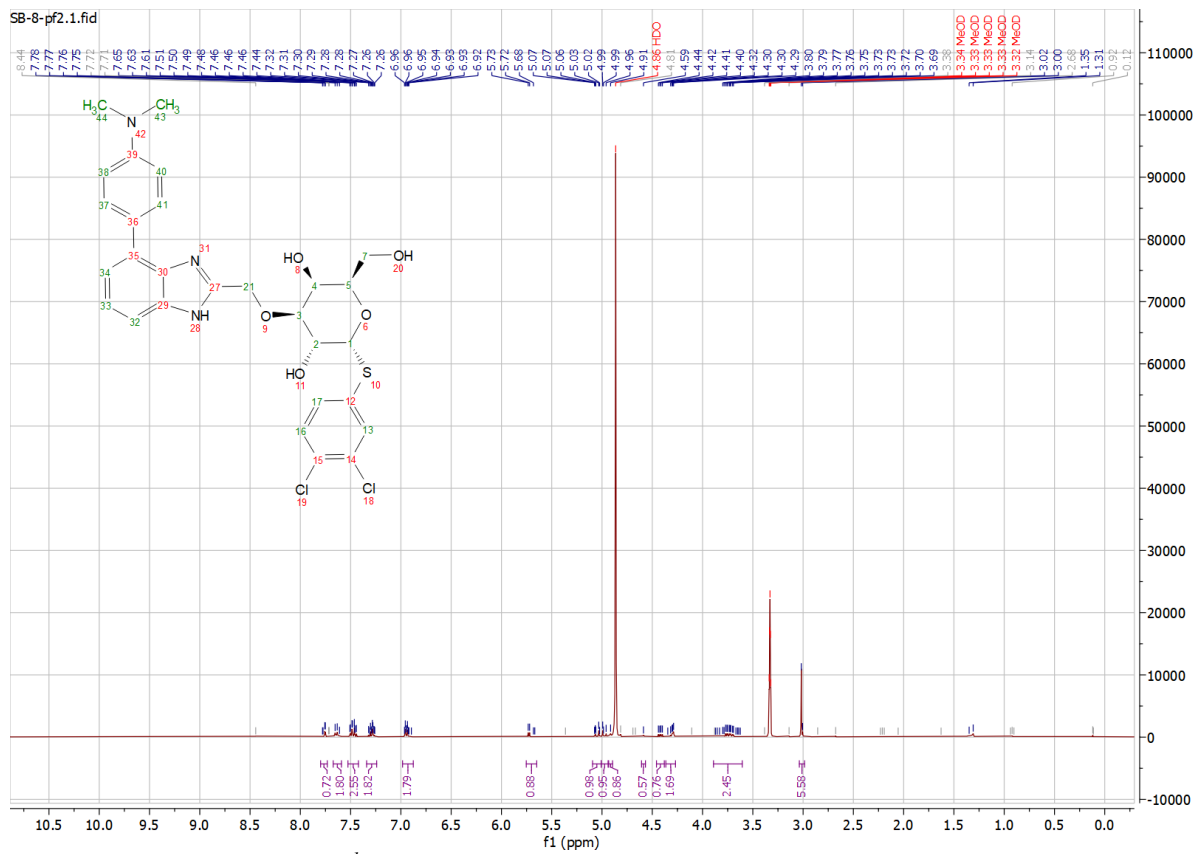
Figure 20. *¹H NMR spectrum of **compound 2h** at 400 MHz, MeOD*
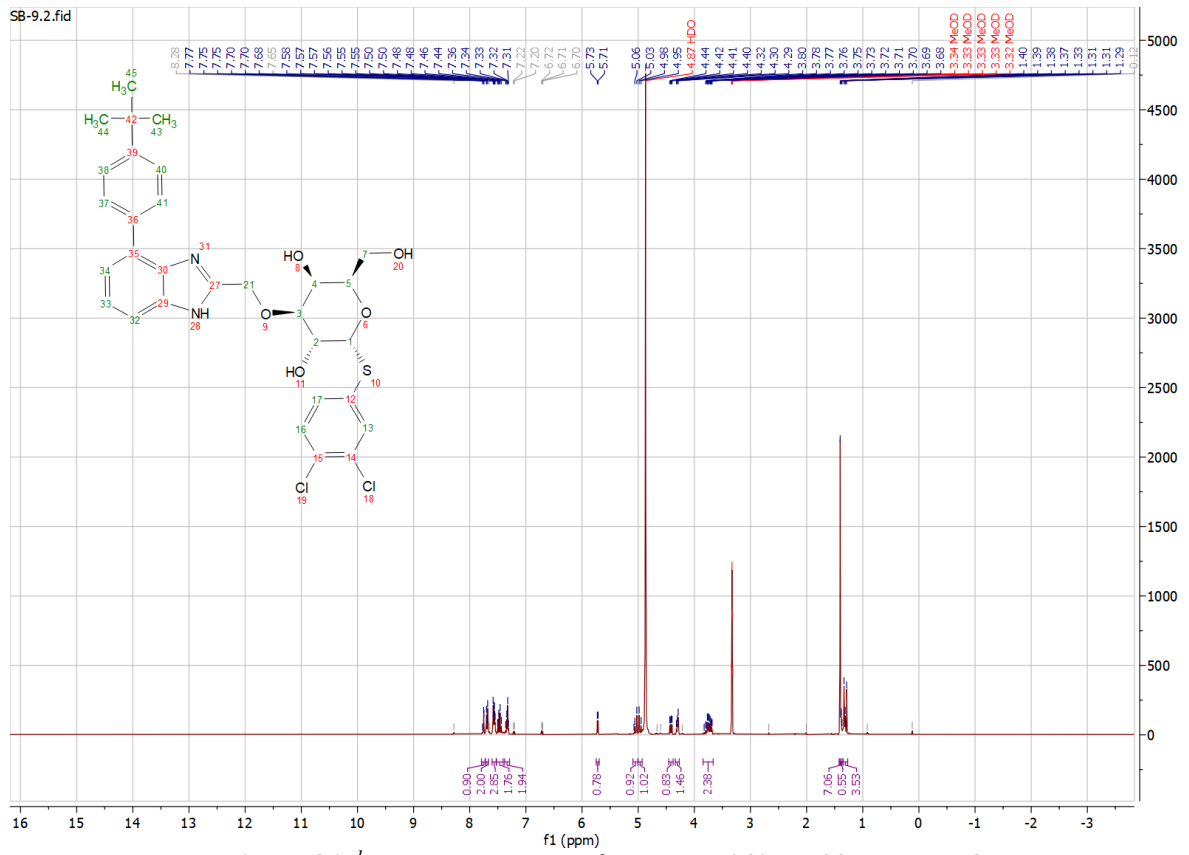
Figure 21. *¹H NMR spectrum of **compound 2i** at 400 MHz, MeOD*

^{13}C NMR spectra
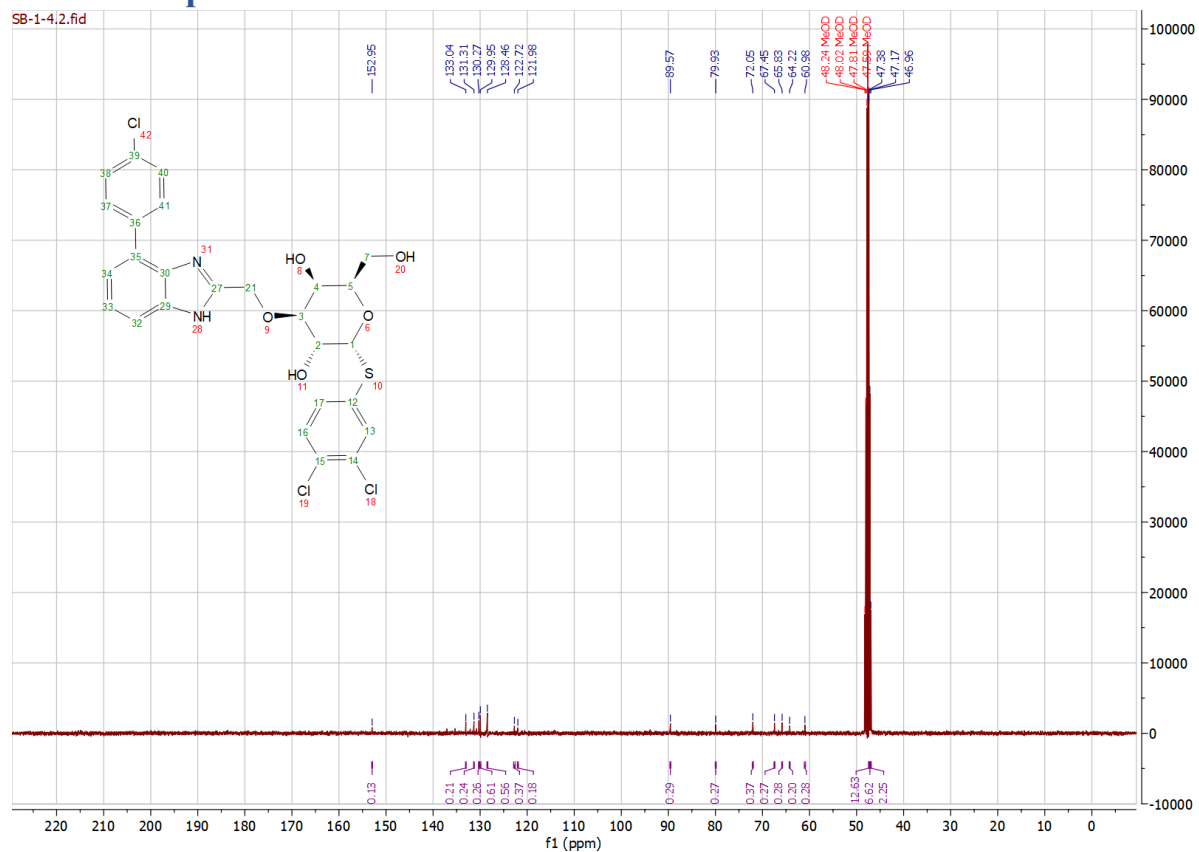
Figure 22. *^{13}C NMR spectrum of **compound 2a** at 100 MHz, MeOD*
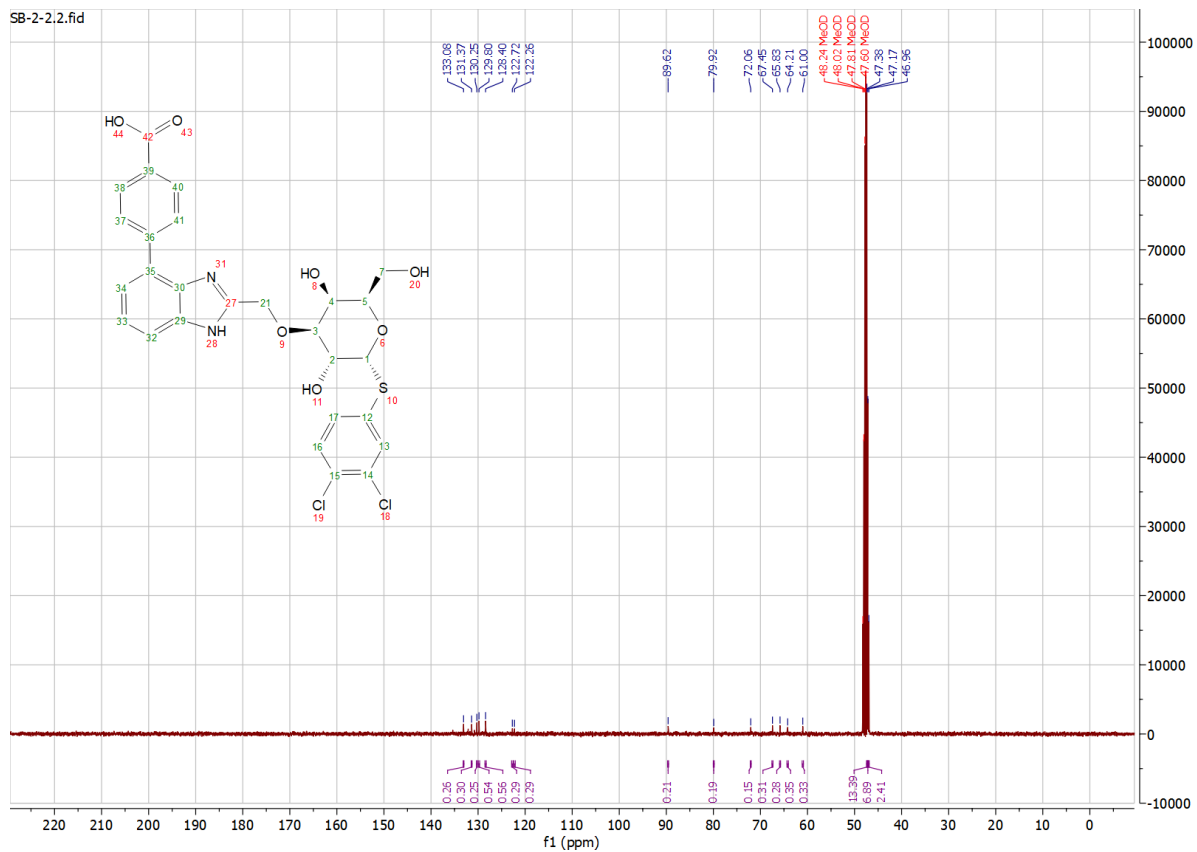
Figure 23. *^{13}C NMR spectrum of **compound 2b** at 100 MHz, MeOD*
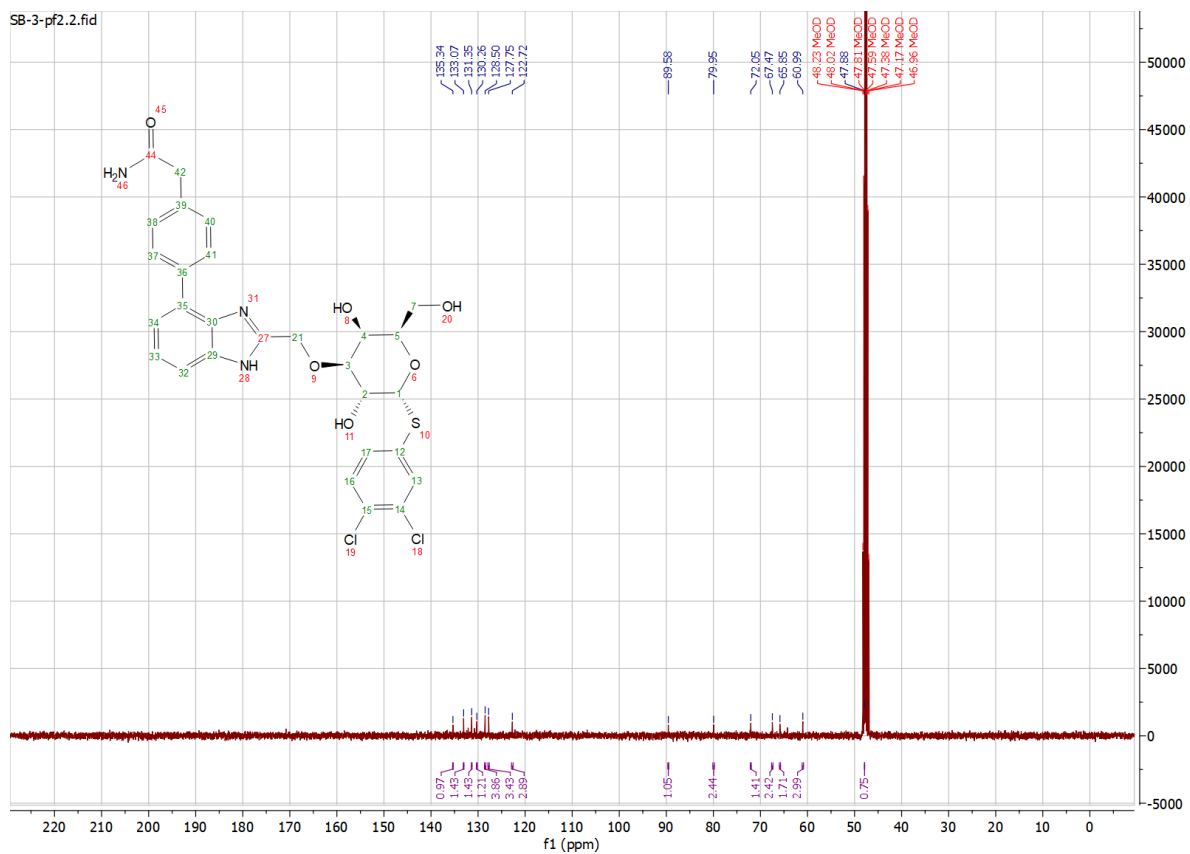
Figure 24. *^{13}C NMR spectrum of **compound 2c** at 100 MHz, MeOD*
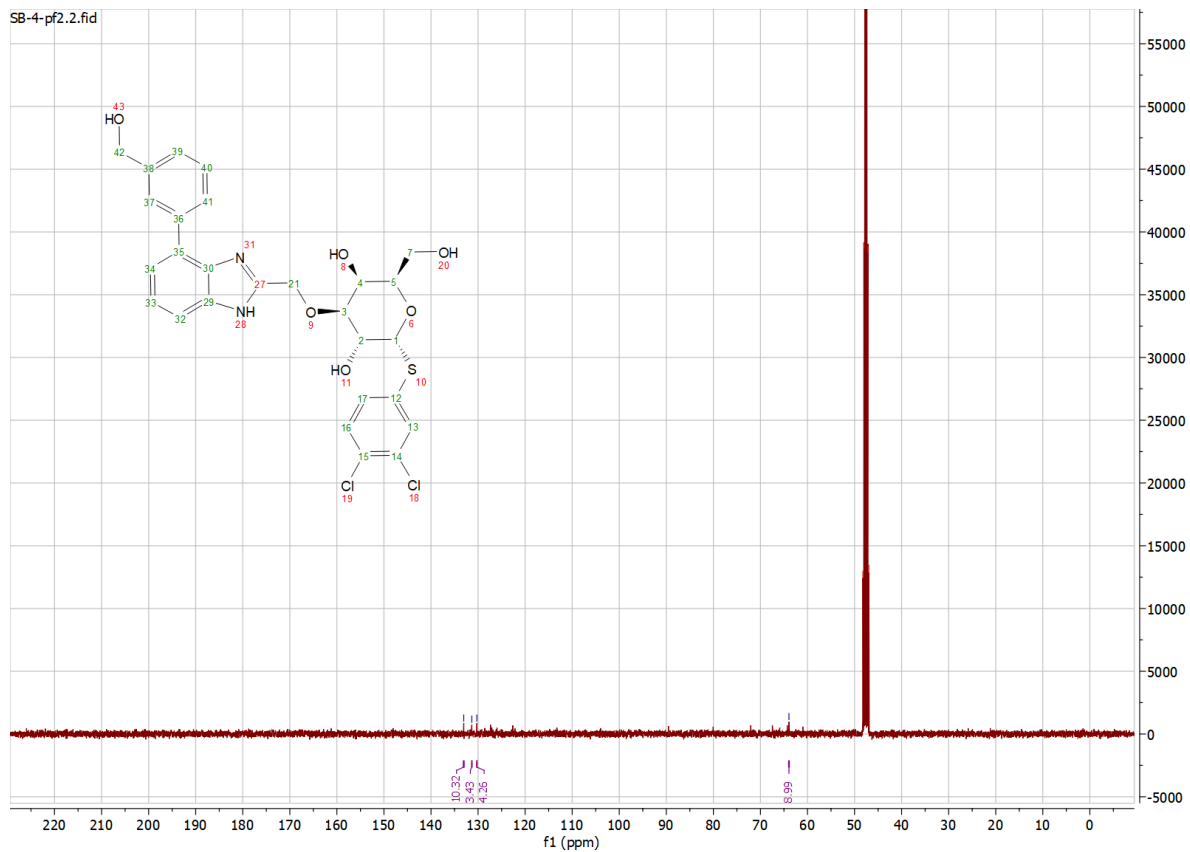
Figure 25. *^{13}C NMR spectrum of **compound 2d** at 100 MHz, MeOD*
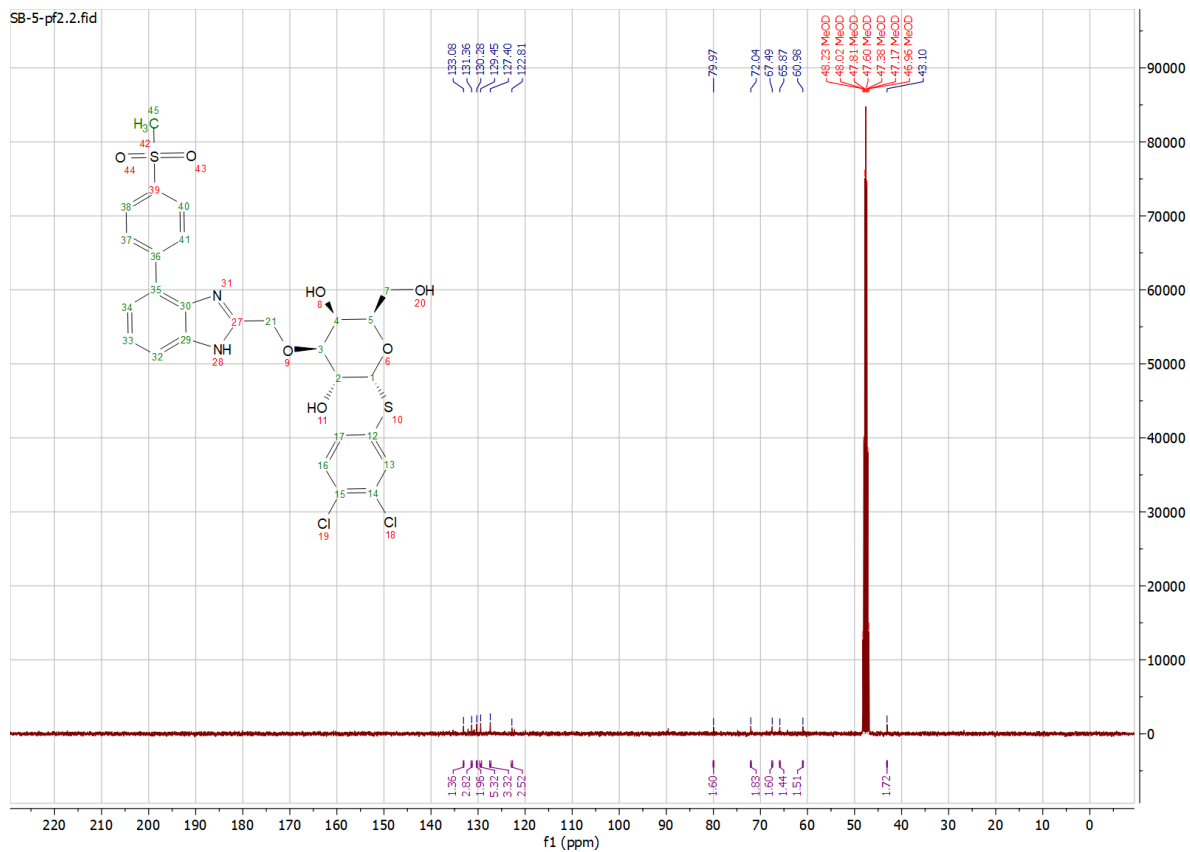
Figure 26. *¹³C NMR spectrum of* ***compound 2e*** *at 100 MHz, MeOD*
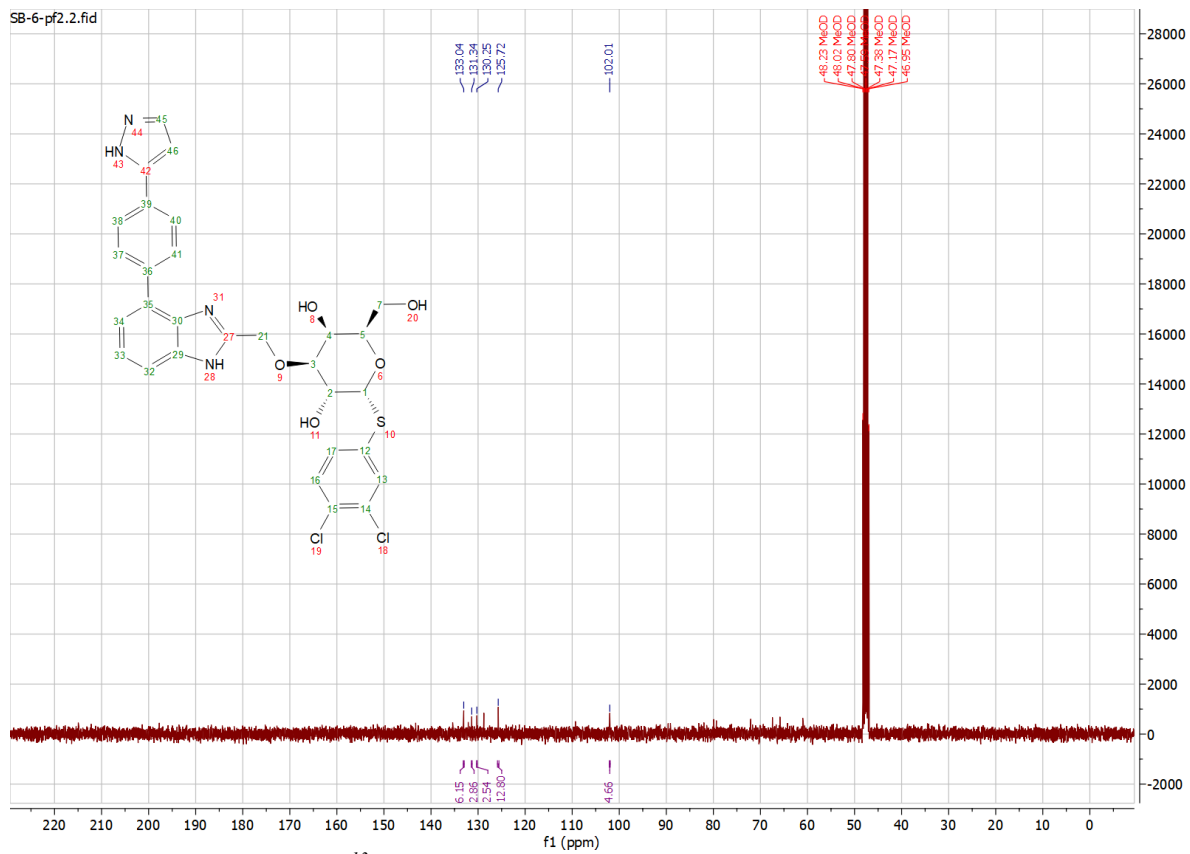
Figure 27. *^{13}C NMR spectrum of* **compound 2f** *at 100 MHz, MeOD*

Figure 28. *^{13}C NMR spectrum of **compound 2f** at 100 MHz, DMSO*
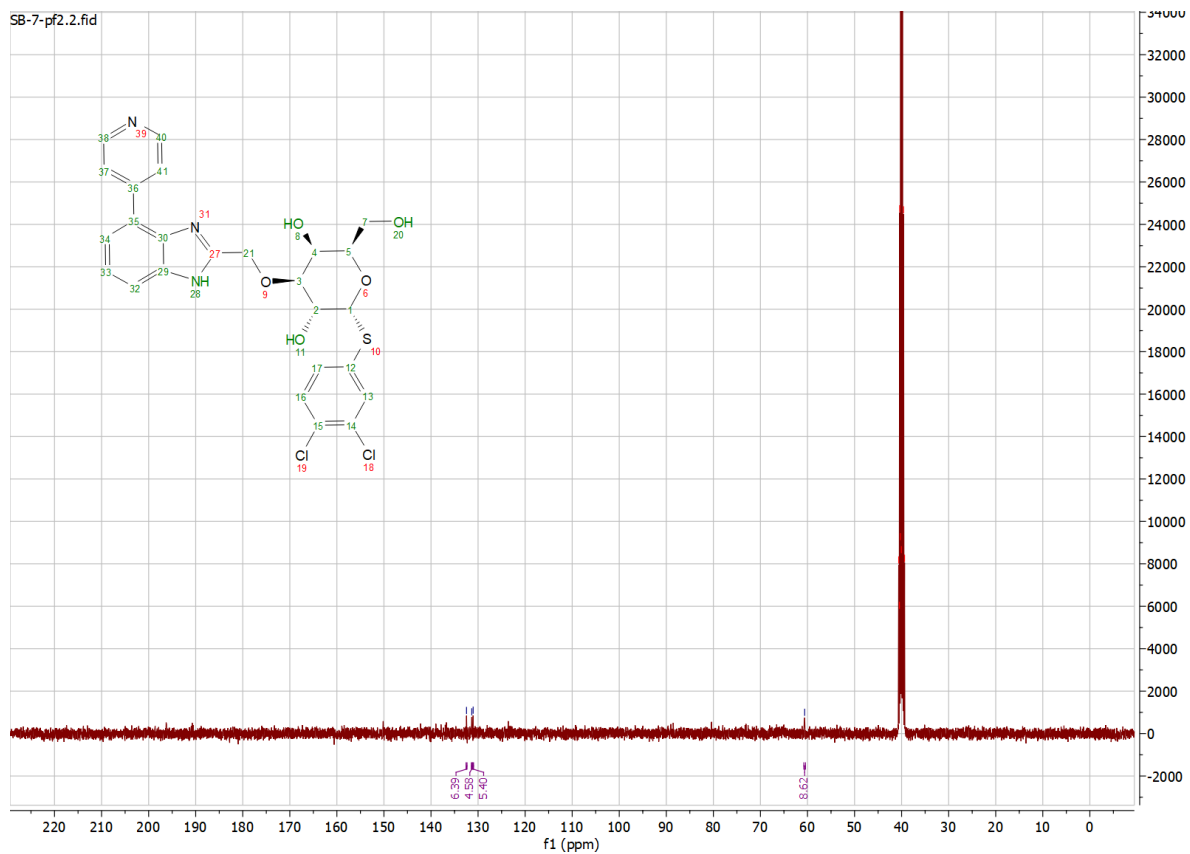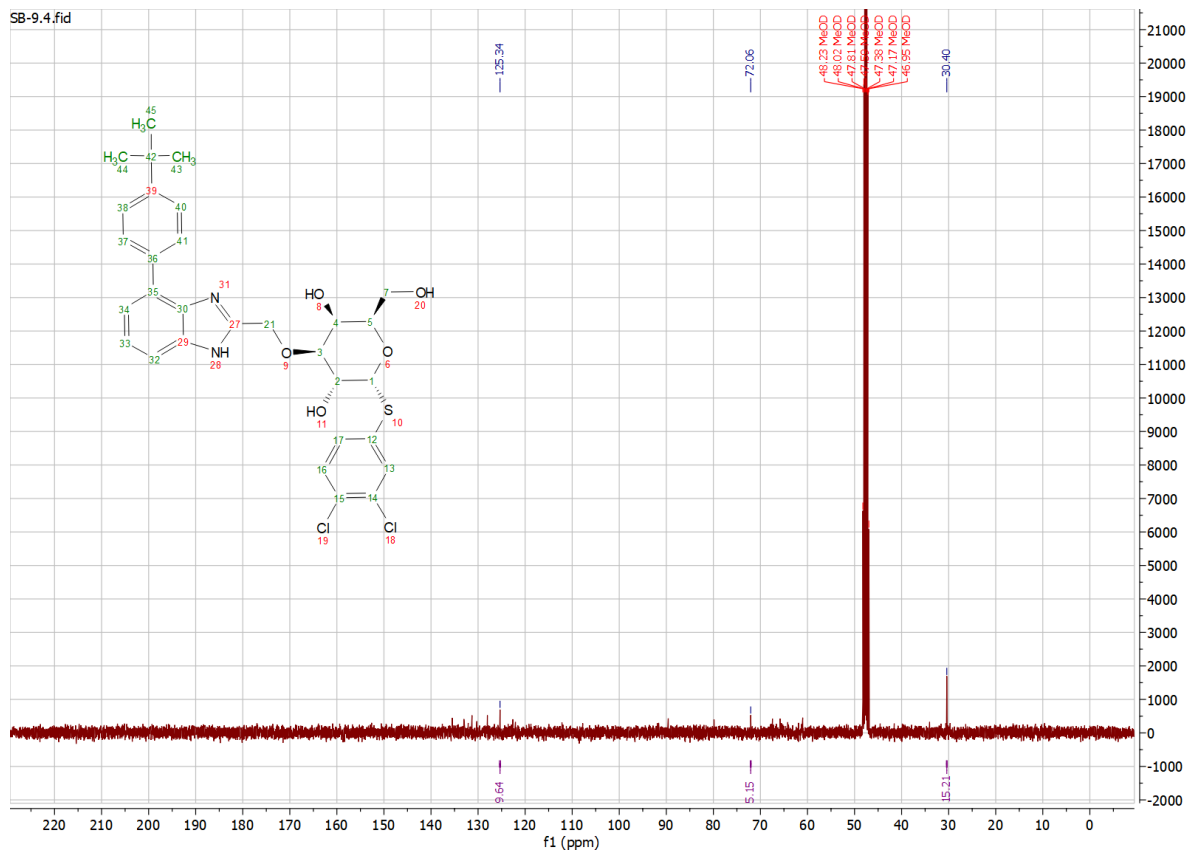
Figure 29. *^{13}C NMR spectrum of* **compound 2i** *at 100 MHz, MeOD*