

Crystallization, optimization, and crystal structure determination of ligands in complex with galectin-8N



LUNDS
UNIVERSITET

George Larsson

Bachelor Project (KEMK10): 15 hp

Lund University

Department of Biochemistry and Structural Biology

Project length: 2022-03-22 to 2022-06-09

Supervisor: Derek Logan

Version: 3

Abstract:

This thesis focuses on the protein “galectin-8N” which is a lectin protein that takes part in several physiological functions such as inflammation, autophagy, cell migration, and more. The project has three objectives: to observe and increase the reproducibility of crystallization of two variants (wildtype and mutant) of the human protein galectin-8N and to analyze the binding site when three different novel ligands are each inserted. Lastly, attempts of crystallizing mouse galectin-8N was also investigated which there is no prior crystallization data of.

To observe required conditions for crystallizing and to increase the reproducibility of the two variants of human galectin-8N, two commercial plates of “Morpheus™” were used and four manual plates imitating the conditions of Morpheus. After crystals were observed in the droplets of the manual plates, one of the plates was used to soak the ligand MH-3-80 and the other two to cocrystallize with ligands MH-3-80, MH-1FH and NV9.1. All the produced crystals were brought to the MAX IV where data was collected from and through iterative refinement processes with phenix, five models of human galectin-8N were created. From the models, all interactions between inserted ligands and the binding site were observed via PyMOL. Lastly, crystallization attempts of mouse galectin-8N were experimented on in total of eight commercial screens, two buffer screens, and one Morpheus screen combined with “Seed Beads”. The results were that both variants of human galectin-8N crystallizes in the Morpheus commercial plates and well in manual screens imitating Morpheus’ conditions. Crystals were found to prefer conditions of: buffer 1, in precipitant mixture 4, and with additives such as ethylene glycols, monosaccharides and alcohols. Ligands MH-3-80 and MH-1FH were inserted successfully into both variants of the human proteins, while NV9.1 did not. The interactions that occurred between inserted ligands and the residues at the binding site were: Arg52, His72, Arg76, Asn86, and Trp93. Other interactions were also observed however those existed exclusively depending on the inserted ligands’ structure and their moieties. Lastly, the crystallization of mouse galectin-8N was unfortunately difficult and rather unsuccessful, however the commercial screen: JCSG+™ gave the most promising results and is highly recommended for future observations.

Table of Contents:

0.0 Acknowledgement:	5
1.0 Introduction:	6
2.0 Background:	6
2.1 The Galectin Family:	6
2.2 Galectin-8N:	7
2.3 Binding & Ligands:	8
2.4 Protein preparation, extraction, and purification:	11
2.5 X-ray Crystallization Method:	12
2.5.1 Crystallization of protein:	12
2.5.2 Commercial Screening:	13
2.6 Crystallization Requirements:	15
2.7 Data collection:	16
2.7.1 Molecular Replacement:	16
2.7.2 Refinement:	16
2.8 NanoDSF:	18
2.9 NanoDrop:	18
2.10 Seed Bead crystallization:	18
2.11 Serendipity and accidents:	19
3.0 Methods:	20
3.1 Commercial Screening (JCSG+, PACT & Morpheus):	20
3.2 NanoDSF melting point analysis and Buffer screens (RUBIC & Durham):	20
3.2.1 RUBIC Screen:	20
3.2.2 Durham Screen & mouse galectin-8N melting point:	20
3.3 Screen Optimization (Hanging Drop):	21
3.3.1 Ligand Soaking:	22
3.3.2 Ligand Cocrystallization:	22
3.4 Crystal extraction and data refinement:	22
3.5 Mouse protein buffer exchange (Centrifuging to citrate):	23
3.6 Mouse Galectin-8N Screening (5 types):	23
3.7 Seed Beads (human galectin-8N variant M):	23
3.8 Seeding with commercial screen:	24
4.0 Results:	24
4.1 Human Galectin-8N variant M and V Results:	24
4.1.1 Commercial Screens:	24
4.1.2 Manual/Optimization Screens:	26
4.1.3 Ligand Cocrystallization Manual Screens:	26
4.1.4 Data collection and refinement:	27

4.1.5 Data refinement and protein modeling:	29
4.1.6 Ligand observed in the binding site:	29
4.1.7 Refinements and Models:	30
4.1.8 Ligand Interaction in Binding Site:	35
4.1.9 Manual Seed Beads Results:	41
4.2 Mouse Galectin-8N Results:	42
4.2.1 nanoDSF Melting Point Analysis and Buffer Screens:	42
4.2.1.2 RUBIC Screen:	42
4.2.1.3 Durham Screen:	43
4.2.2 Mouse Galectin-8N Commercial Screen (Stock in PBS buffer):	44
4.2.3 Mouse galectin-8N Commercial Screen (Stock in Citrate buffer):	46
4.2.4 Seed Bead with Commercial Screen Results:	46
5.0 Discussion:	46
6.0 References:	48
7.0 Appendix:	49
7.1 Method:	49

0.0 Acknowledgement:

Before the report, I personally would like to express my sincere appreciation to everyone who supported and contributed to this project. This project was only possible thanks to; first, my supervisor during this project Derek Logan for offering help, advice, and suggestions for the project. Also, I would like to thank Ipsita Banerjee for offering advice during lab-work. A thanks to Wolfgang Knecht at LP3 (Lund Protein Production Platform) for donating the protein samples. Furthermore, a big thanks to Maria Gourdon and the staff of LP3 helping me in several experiments. An additional thanks to Ulf Nilsson and Mujtaba Hassan for contributing the ligand samples and their knowledge of them. Lastly, a thanks to all staff of BioMAX beamline at MAX IV for help with data collection.

1.0 Introduction:

The purpose and aim of this project is to find optimal conditions and to increase the reproducibility of crystallizing two variants of the human protein named galectin-8N, utilizing prior data from a bachelor thesis [1] and expanding it further using crystallization methods such as vapor diffusion. In addition, the project also observes how several ligands inserted into the binding site such as; MH-3-80, MH-1FH, and NV9.1 interact with both variants of the human proteins.

Additionally as a side project, crystallization of the protein mouse galectin-8N was attempted and the conditions of the experiments were all noted down. Mouse galectin-8N does not have a model yet available due to no data of successful crystallization has been reported yet. Crystallizations are attempted with the aid of several commercial screens in addition observation of suitable buffers for crystallization are perceived.

2.0 Background:

2.1 The Galectin Family:

Galectins are carbohydrate binding lectin proteins that mainly possess the purpose of binding beta-galactosides where in recent years it has been discovered that these proteins also take part in several physiological functions. These proteins tend to be involved in functions such as; inflammation, cell migration, autophagy and even in cancer [2]. In total 15 types of galectins have been discovered in mammals, of which in humans 12 types can be seen. Galectin-1, 2, 3, 7, 8, 9 and 12 have been observed to induce apoptosis in certain types of cells and also having the purpose to be involved in post-developmental processes and immune regulations [3].

What makes galectins unique is the fact that they are identifiable by inspecting their domain containing up to one or two CRD's (Carbohydrate Recognition Domains), which are domains with high affinity for beta-galactosides. CRDs in galectins are divided into three groups: prototypical, tandem-repeat, and chimera. Prototypical (or dimer) CRDs are the most common and contain two identical CRDs and can be observed in galectin-1, 2, 5, 7, 10, 11, 13, 14, and 15. Tandem-repeat is when two distinct CRDs form and can be inspected to exist in galectin-4, 8, 9, and 12. Lastly, chimera is the most unusual one which either contains one or multiple CRDs and can be observed only in galectin-3 [3].

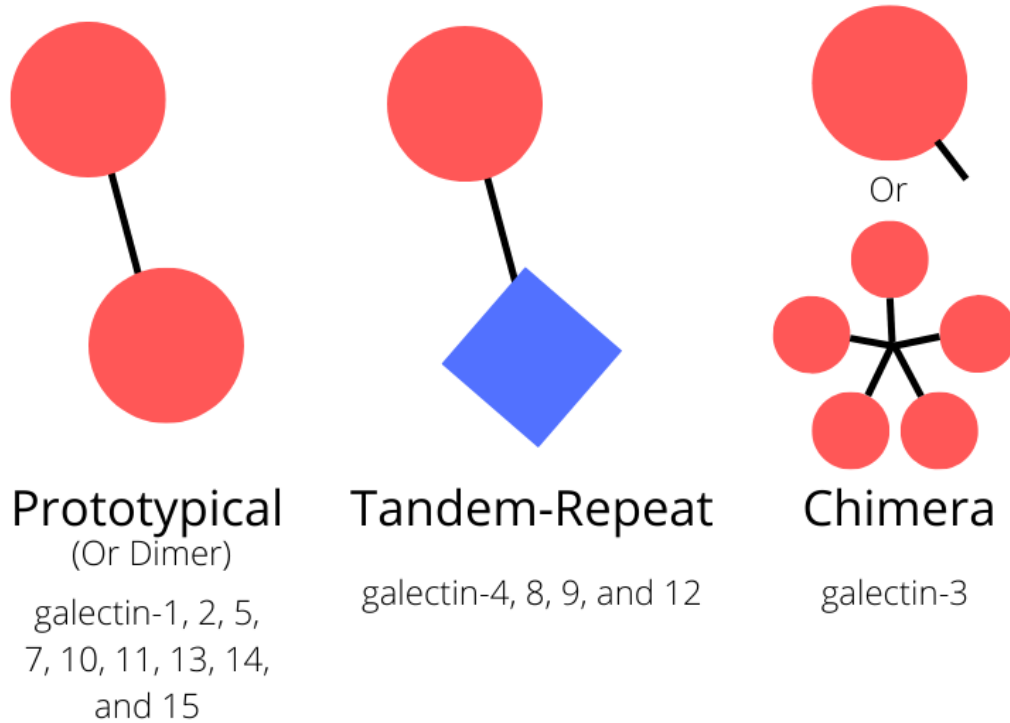


Figure 1: The three types of CRDs that can be recognized in the galectin family protein and which galectins they can be observed in.

2.2 Galectin-8N:

This project focuses on galectin-8N, also known as the N-terminal domain of galectin-8. This galectin has been found to be involved in inducing apoptosis of carcinoma cells by binding to integrins, which are a type of transmembrane receptors on cells [4]. This inhibition mechanism is specific to galectin-8 and is not found in other galectins such as -1 and -3 which are mainly involved in cancer processes. Apoptosis is the base mechanism to prevent cancer a living organism undertakes, which means that if a malfunction occurs it may result in cancer unless T-cells do their functions properly. Thus, obtaining knowledge and research about galectin proteins might be one step closer in the prevention of cancer.

In this project two types of human galectin-8N are researched: galectin-8N aa 4 - 158 M- and V-version. The two are homologues, with molecular weight of about 17380 Da and differ in only three residues, which are located at positions 16, 33, and 54 (see figure 2). In humans the M-version of the protein is found while version V is a mutation version of M and is less common.

```

98.1% identity in 155 residues overlap; Score: 804.0; Gap frequency: 0.0%

M      1  SLNNLQNIYNPVVPFVGTIPDQLDPGTLIVIRGHVPSDADRFQVDLQNGSSMKPRADVA
V      1  SLNNLQNIYNPVPIPVVGTIPDQLDPGTLIVICGHVPSDADRFQVDLQNGSSVKPRADVA
      *****

M      61 FHFNPRFKRAGCIVCNTLINEKWGREEITYDTPFKREKSFEIVIMVLKDKFQVAVNGKHT
V      61 FHFNPRFKRAGCIVCNTLINEKWGREEITYDTPFKREKSFEIVIMVLKDKFQVAVNGKHT
      *****

M      121 LLYGHRIGPEKIDTLGIYGKVNHSIGFSFSSDLQ
V      121 LLYGHRIGPEKIDTLGIYGKVNHSIGFSFSSDLQ
      *****

```

Figure 2: A sequence alignment of galectin-8N variants M and V. They differ in only three amino acids being Phenylalanine-Tyrosine (16), Arg-Cys (33), and Methionine-Valine (54) having a sequence identity of 98.1%.

For the mouse galectin-8N aa 3 - 157 not much is known about it except for having similar functions to human galectin. The sequence which in comparison to its homologue human galectin-8N variant M differs with 30 residues having a sequence identity of 80.6% (see Figure 3). The molecular weight is about 17270 Da.

```

80.6% identity in 155 residues overlap; Score: 681.0; Gap frequency: 0.0%

Human  1  SLNNLQNIYNPVVPFVGTIPDQLDPGTLIVIRGHVPSDADRFQVDLQNGSSMKPRADVA
Mouse  1  SLNNLQNIYNPPIIPVVGITITEQLKPGSLIVIRGHVPKDSERFQVDFQLGNSLKPRAVA
      ***** ** **** ** * ***** * ***** * * *****

Human  61 FHFNPRFKRAGCIVCNTLINEKWGREEITYDTPFKREKSFEIVIMVLKDKFQVAVNGKHT
Mouse  61 FHFNPRFKRSSCIVCNTLTQEKWGWEIITYDMPFRKEKSFEIVFMVLKNKFQVAVNGRHV
      ***** ***** **** ***** ** ***** **** ***** *

Human  121 LLYGHRIGPEKIDTLGIYGKVNHSIGFSFSSDLQ
Mouse  121 LLYAHRISPEQIDTVGIYGKVNHSIGFRFSSDLQ
      *** ** * ** * ***** *****

```

Figure 3: A sequence alignment of human galectin-8N variant M and mouse galectin-8N. They differ in 30 amino acids and have an identity of 80.6%.

2.3 Binding & Ligands:

Galectins are lectins that bind galactosides, hence why the protein is named galectin. It has a high preference for beta-galactosides as mentioned earlier for instance beta-lactose, however depending on the ligand structure it will bind to other compounds that are not saccharides. This project observes how sample ligands interact with the binding site and what kind of ligands attach to them.

There are several structures in the PDB data bank where the binding of natural and synthetic ligands to the binding site of galectin-8's N-terminal domain has been observed. In particular, in Noemi Ströhagen's thesis the ligand interactions are observed by inserting a sample ligand called "MH1" (MH-3-80 in this project) and observing its surrounding interactions with residues at the binding site [1].

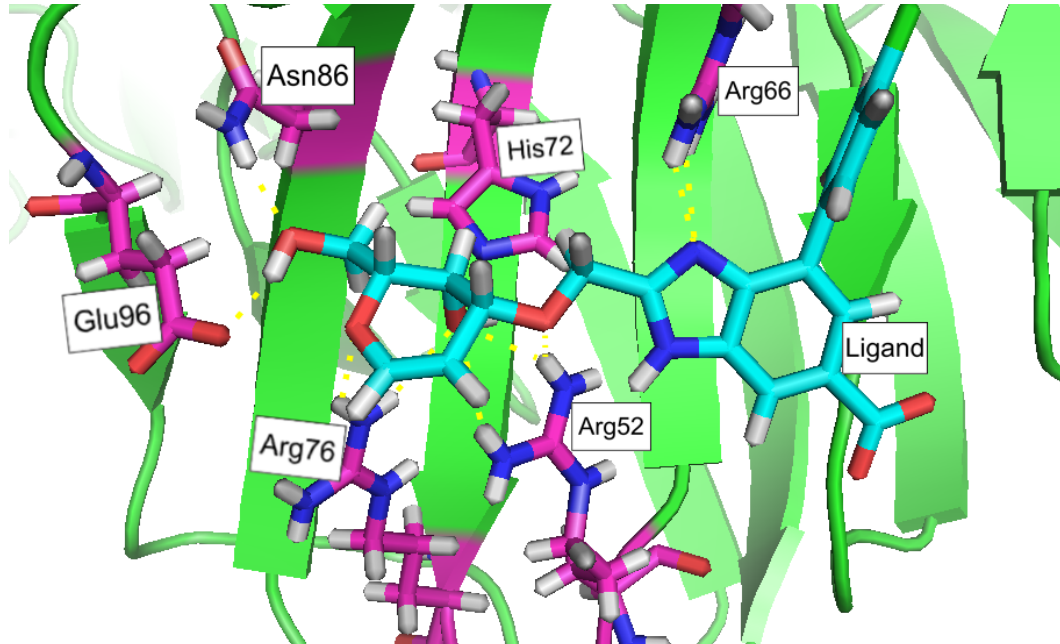


Figure 4: Model of galectin-8N variant M [1]. Possible to observe that six residues (colored in purple) interact with the inserted ligand (colored in cyan).

The following table below shows all interactions between inserted ligands and the binding site from several models available on PDB and UniProt (see table 1):

Table 1: The table below inspects the interaction between an inserted ligand and residues in selected models in PyMOL and on PDB:s website . All models are obtained from [1], PDB or UniProt.

Galectin-8N Models	Inserted Ligand	Residue interactions with ligand in binding site
Ströhagen's Model (Chain A)	MH1 (aka MH-3-80)	Arg52, Arg66, His72, Arg76 Asn86, Glu96
Ströhagen's Model (Chain B)	MH1 (aka MH-3-80)	Arg52, Arg66, His 72, Arg76, Asn86
PDB: 7ALS (Chain A)	Lactose	Arg42, Arg66, His62, Asn76, Glu86
PDB: 7ALS (Chain B)	Lactose	Arg42, Arg66, Glu86
PDB: 7P1M	2-[[[(2R,3R,4R)-2-(hydroxymethyl)-3-oxidanyl-3,4-dihydro-2H-pyran-4-yl]oxymethyl]-3-methyl-benzimidazole-5-carboxylic acid	Arg45, His60, Arg64, Asn72, Glu84
PDB: 5TIU	Glycerol	Arg45, His65, Arg69, Asn79, Glu89
UniProt: O00214	3'-O-sialylated and 3'-O-sulfated glycans	Arg59, Arg69, Asn79, Glu89

From the results of the interactions above (see table 1), between human galectin-8N variant M and chosen ligands, it is possible to see that two arginines always interact in the binding site while histidine, asparagine, and glutamate interact in the majority of the selected models. Not to mention, there are interactions that are not visible in PyMOL such as hydrophobic interaction which exists between phenyl-groups in the ligand and e.g tryptophan at position 93. Due to the similarity in sequence between galectin-8N variant M and V we may expect that residues interacting would also be the same. The only uncertainty is that arginine at position 43 (see figure 2) mutates into a cysteine residue which may or may not affect the overall bonding with a ligand.

On the other hand, mouse galectin-8N has too many mutations compared to the human protein which makes it difficult to predict how a ligand inserted would interact with its surrounding residues at the binding site. However, when observing the sequence alignment (see figure 3) between human variant M and mouse it is possible to notice that the binding site is conserved having arginine, histidine and asparagine. Thus, if the structure tends to be similar enough its binding capability can be predicted to be similar to the human protein.

Between a ligand and residues at the binding site several weak non-covalent interactions are occurring, where the most common types are; hydrogen bond, hydrophobic interactions, or ionic bonds. Observing how ligand MH1 (MH-3-80) interacts with its surrounding residues present (see table 1), it is possible to notice that interactions are mainly hydrogen bonds and hydrophobic interactions.

Three ligands were tested to observe how they interact with human galectin-8N of both variants. The ligands were received by Mujtaba Hassan from the Centre for Analysis and Synthesis and were named as following: MH-3-80, MH-1FH, and NV9.1. Structure-wise the three ligands differ a little, but the most significant difference is where MH-1FH contains a galactose moiety in the middle of the ligand, while MH-3-80 and NV9.1 both have a galactal moiety at the end of the molecule (see figure 5).

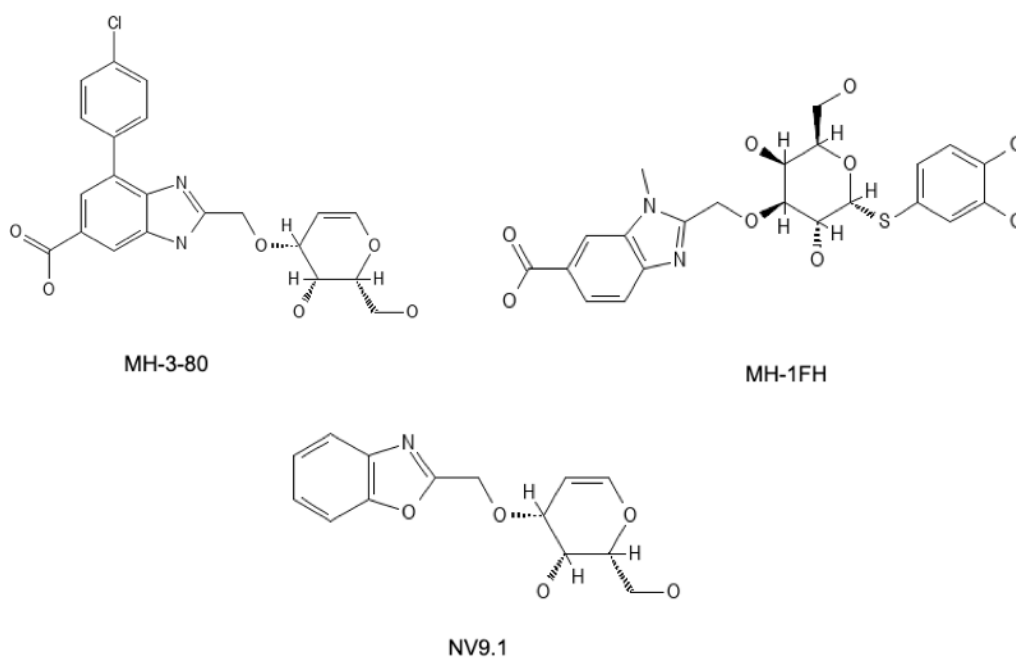


Figure 5: Structures of the three ligands; MH-3-80, MH-1FH, and NV9.1. The molecules were sketched in “PubChem sketcher” by inserting the SMILE strings.

Smile strings for the three ligands:

MH-3-80:

O[C@H]1[C@@H](CO)OC=C[C@H]1OCC2=NC3=C(C4=CC=C(Cl)C=C4)C=C(C(O)=O)C=C3N2

MH-1FH:

OC[C@H]1O[C@H](SC2=CC=C(Cl)C(Cl)=C2)[C@H](O)[C@@H](OCC3=NC(C=CC(C(O)=O)=C4)=C4N3C)[C@H]1O

NV9.1:

O[C@H]1[C@@H](CO)OC=C[C@H]1OCC2=NC3=C(C=CC=C3)O2

The ligands are inserted into the protein either via ligand soaking or cocrystallization, where soaking is a method done by pipetting the ligand after the proteins has crystallized, while cocrystallization is done by adding the ligand together with the protein crystallizing them both forming a ligand-complex directly. In general, ligand cocrystallization has a higher success rating than soaking due to having less risk of destroying the crystals after reaching equilibrium.

When observing how well a ligand binds to a protein, it is important to inspect the binding affinity of that ligand. The affinity is measured by observing the equilibrium between the protein, ligand, and ligand-complex formed where the dissociation constant is written as K_d , where the lower the K_d the greater the binding affinity of that ligand [6].

$$K_d = \frac{[P]+[L]}{[P+L]} \tag{1}$$

Where K_d is the dissociation constant, $[P]$ is the protein concentration, $[L]$ is the ligand concentration, and $[P+L]$ is the concentration of the ligand-complex.

The properties of the ligands used in this project can be found in the table below (see table 2):

Table 2: The table below shows all ligands utilized, the amount received, its molecular weight, and the dissociation constant (K_d).

Ligand	Amount (mg)	Molecular weight (g/mol)	K_d (μ M)
MH-3-80	1.332	430.84	2.9
MH-1FH	1.119	529.39	1.8
NV9.1	1.445	277.28	290

2.4 Protein preparation, extraction, and purification:

The three galectin-8N proteins utilized during the project were received from LP3 (Lund Protein Production Platform) together with three protocols with facts and information about the proteins.

According to the protocols [5] both variants of the human galectin-8N were prepared in similar conditions: Produced in *E. coli* and purified using affinity chromatography in a 19 ml lactosyl-Sepharose column and eluted using MEPBS (mixture of PBS buffer 2 mM EDTA and 4 mM methanol) +150 mM lactose. The buffer was exchanged with PBS (Phosphate-Buffered Saline) buffer at pH 7.4 via dialysis and presumably lactose attached to the binding site was removed. The concentration of variant M was measured to be 13.1 mg/ml and for V to be 4.4 mg/ml using NanoDrop (extinction coefficient of 7.4) and both proteins were estimated to have a final purity of over 95%. The volumes received of both proteins were 1 ml each in two Eppendorf tubes.

The mouse galectin-8N was also extracted in a similar way to the human proteins, however it is mentioned in the protocol [5] that during elution using a 22 ml lactosyl-Sepharose column a lot of precipitate was produced in elution fractions. The final concentration was measured to be 2.6 mg/ml using NanoDrop (extinction coefficient of 10.2) with an estimated purity over 95% and the volume received were 1 ml in an Eppendorf tube.

2.5 X-ray Crystallization Method:

2.5.1 Crystallization of protein:

In order to inspect the structure of a protein using X-ray crystallography, a requirement is to crystallize the target protein first. Crystallization of protein occurs when intermolecular interactions with surrounding proteins take place when various conditions are met, which mainly depends on factors such as: protein purity, pH, protein concentration, precipitant concentration, additives and temperature.

There are various ways to produce protein crystals where during this experiment the method “Vapor diffusion” is utilized. Vapor diffusion is a popular method to crystallize samples where diffusion between a droplet of sample and a reservoir solution well containing reagents reaches equilibrium. Reagents refers to buffer, precipitant, and additives which are different for every protein in order to crystallize. In order to equilibrate and produce fine crystals qualified for further research an optimal mixture must be discovered, which is usually done through trial-and-error.

Most common ways to execute vapor diffusion are through two similar methods: “Hanging Drop” and “Sitting Drop” where both are popular, effective and easy methods to perform. The main difference between the two are where the protein drop is located where in “hanging drop” it is located on the lid directly above the reservoir well, while for “sitting drop” the drop is on a platform above the reservoir well (see figure 5). A hanging drop experiment is illustrious for manual experiments requiring robustness and flexibility of the drops with crystals in and in addition the method is easily observable with a microscope without vision blur caused by plastic. On the other hand, a sitting drop experiment is excellent for crystallization of samples of low volumes with the aid of robots, which saves time significantly [5].

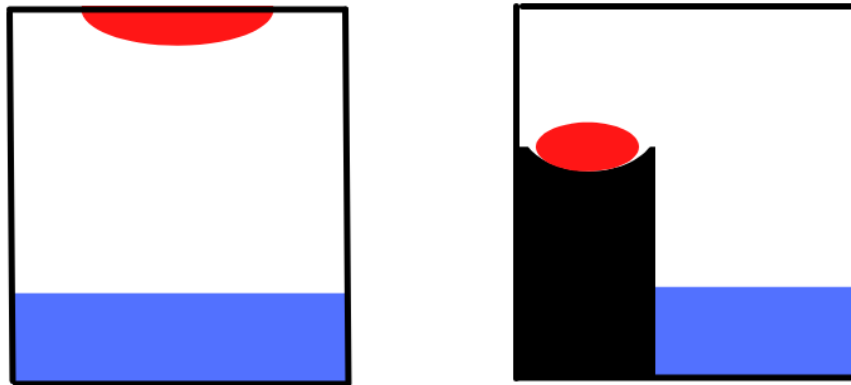


Figure 5: The left figure is “Hanging Drop” and the right is “Sitting Drop”. The red color indicates the protein sample drop while the blue is the reservoir solution.

2.5.2 Commercial Screening:

When crystallization conditions of a protein are unknown commercial crystallization screens are a good starting point, which are plates with multiple wells containing a variety of components covering a broad range such as: pH, precipitant and additives to find a match in crystallization, which can then be optimized for further investigation for optimal conditions. Commonly, in a screen plate there are 96 different conditions to experiment in whether it meets the requirement to crystallize. On the website “Molecular Dimensions™” [9] information about all commercial screen-products utilized during this project can be found, which were in total five: JCSG+™, PACT premier™, Morpheus™, Basic Chemical Space (BCS)™, and ProPlex™.

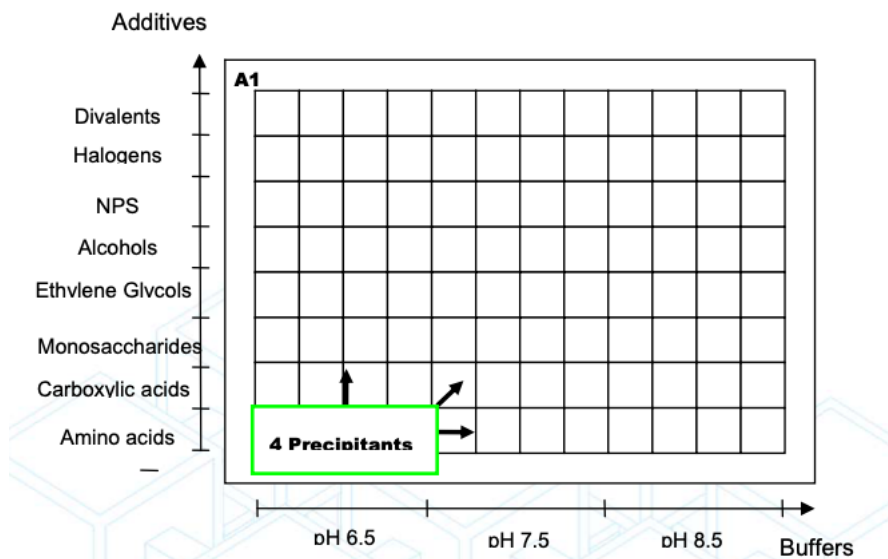


Figure 6: Example of a commercial screen (Morpheus) showing how components vary on the X and Y axis. The picture is from Molecular Dimensions’ website [9]

The main difference between all the screens above is that JCSG+ covers a wide range of pH (4.0~10.5) with different precipitant mixtures while PACT premier observes factors such as pH, anion and cation

interactions in presence of PEG precipitant [9]. These two screens are good for observing crystal conditions of new proteins with no prior crystallization information. Morpheus has additional factors: “additives” and “precipitant mixtures” where additives are components which either bind or help the surrounding of the protein to increase its stability to crystallize, and precipitant mixtures are as the name suggests mixtures of PEG (polyethylene glycol). These additives vary and contain everything from alcohols, sugars, amino acids etc. BCS focuses on a broad range of molecular weight and changes PEG precipitant smears (mixture) that reduces the amount of PEG variables in all 96 wells. Lastly, ProPlex’s main objective is on maintaining protein-protein interactions in complexes under crystallization conditions.

Commercial screen plates are stored in JANSi UVEXps 256/600 plate hotel at 20 °C at LP3, equipped with UV-light and a camera scanning the protein samples frequently. Examples of pictures of proteins samples are inspected through a computer and appear as following (see figure 7 and 8):

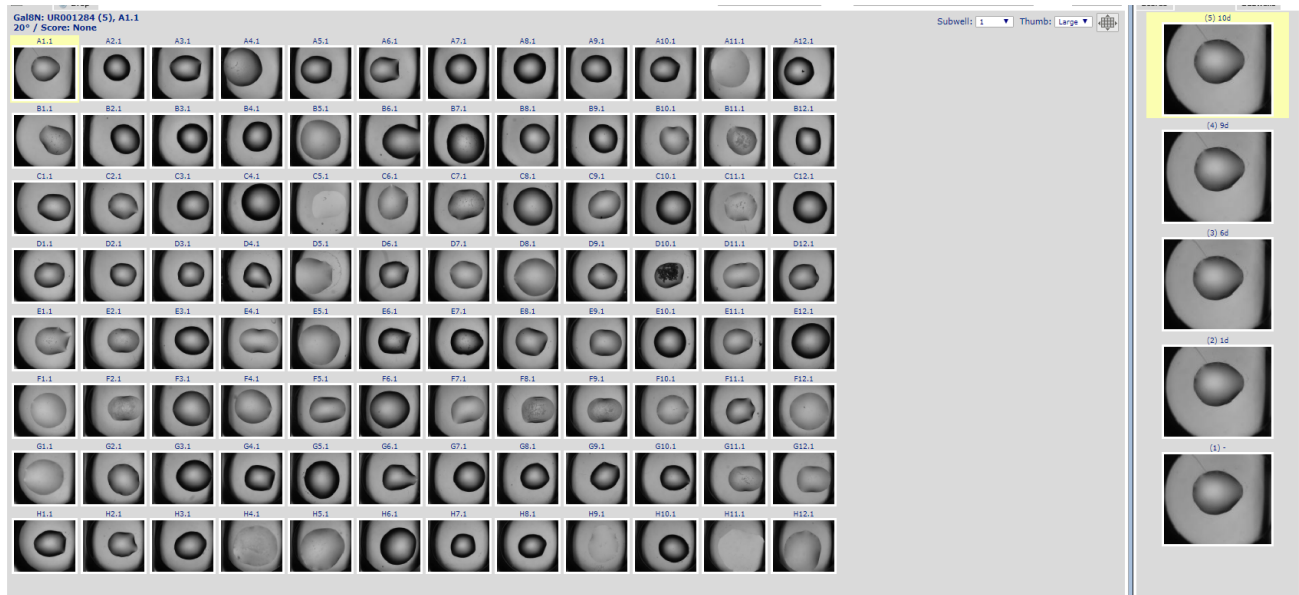


Figure 7: Example of how protein drops are observed through computer software. The pictures were taken at LP3.

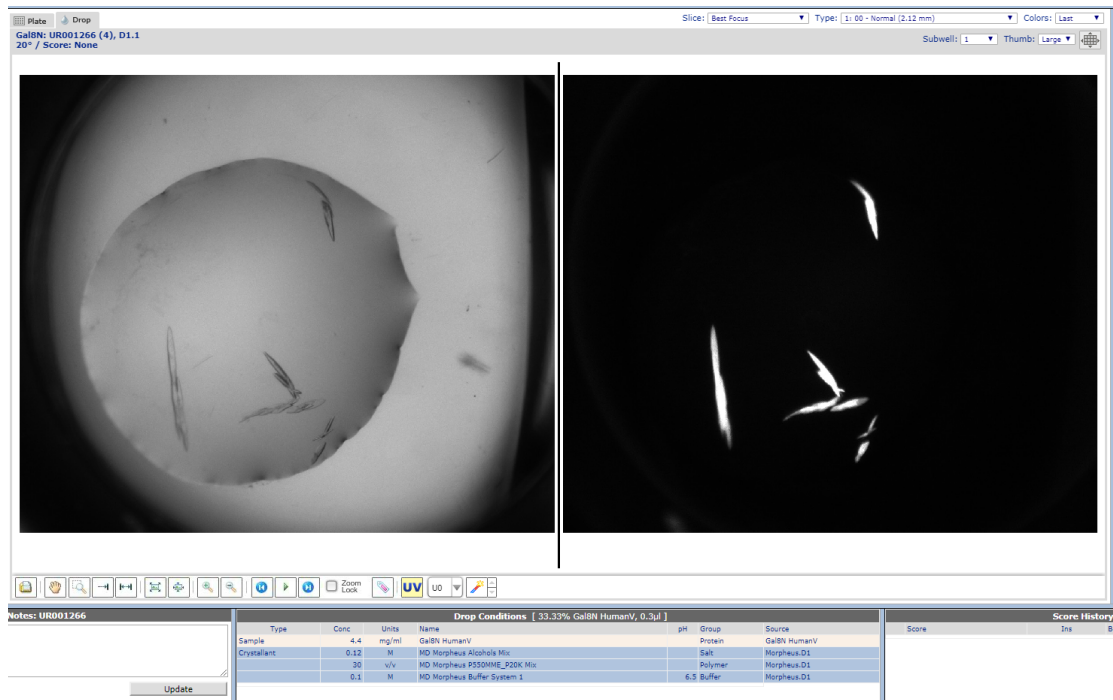


Figure 8: Example of several crystals produced in the protein drop, visible with UV-light. The conditions are also written below the two pictures of the protein crystals.

In addition, during this experiment commercial buffer screens were utilized in order to observe the optimal buffer for mouse galectin-8N, which has the purpose of inspecting how buffers with different conditions and pH affects the thermal stability of a protein. The higher the melting temperature of the protein increases the better the buffer, due to stabilizing the protein. The optimal buffer is found when the melting point increases with 1°C or more. Two buffer screens are used to find the optimal buffer which was also produced by Molecular Dimensions [9] and were: RUBIC™ and Durham™. Both screens have the same goal to observe how thermal stability is affected but the main difference between them is that RUBIC contains small molecule additives that helps the folding/unfolding process, while on other hand Durham has 30 salts in the buffer screen that affects the stability of the protein. By combining these two buffer screens it increases the accuracy of finding the optimal buffer required for the mouse protein.

2.6 Crystallization Requirements:

From observing Ströhagen's data of crystallizing human galectin-8N version M the best crystals were obtained via Morpheus™ commercial screen and then by imitating the conditions manually similar crystals were reproduced [1]. According to the data the optimal crystals were formed in the condition: pH 6.5 buffer mixture 1 (61% MES and 39% imidazole), precipitant mix either 1 & 4, and alcohol additives. To confirm this data and increase the reproducibility, Morpheus and manual screening are tested during this project.

Additionally, a buffer to dilute the protein samples for commercial screening named "buffer 1" in Ströhagen's thesis [1] contained: 10 mM Tris/HCl (pH 8.0), 150 mM NaCl, 1 mM TCEP, 10 mM lactose, and rest water. However, to form apoproteins without neither ligand nor additives a new buffer was prepared where the lactose was removed and instead the amount of water was increased. For future naming, the buffer without lactose is named "Ströhagen's buffer-alpha".

2.7 Data collection:

Data collection and analysis were performed after data was obtained by x-rays hitting the crystal and creating a diffraction pattern at BioMAX beamline of MAX IV lab in Sweden. The diffraction pattern containing structure factor, amplitudes and other viable information is saved through a MTZ-file and can be downloaded from MAX IV's website. About five pipelines simultaneously process the data and convert it into an MTZ-file where all five pipelines have different qualities in several variables such as resolution, Rmerge, Completeness, and etc. The MTZ-file from the pipeline with the best quality is chosen for further modeling. During this project all protein models were created using the MTZ-files gathered from the STARANISO™ pipeline due to producing the best results for anisotropic data compared to the other pipelines.

2.7.1 Molecular Replacement:

A problem to mention that occurs during X-ray diffraction is “phase problem”, which means that the phase is difficult to find due to not being able to measure the phase of each individual diffracted X-ray, therefore must be measured. There are several methods such as; Molecular Replacement (MR), Multiple Isomorphous Replacement (MIR), or MAD/SAD to solve the phase problem and in this experiment molecular replacement is utilized. Molecular replacement is a method where a homologous protein with a high sequence similarity is used to obtain the phase. During this experiment, the phase of variant V is obtained with molecular replacement from Ströhagen's model of galectin-8N variant M [1].

Molecular Replacement is done via the program Phaser-MR in the software Phenix (Python-based Hierarchical ENvironment for Integrated Xtallography). Phaser-MR requires input of an MTZ file of the analyzed crystal, a model of the (PDB file) of the homologous protein, sequence identity (%), space group, and lastly molecular weight. This produces a new MTZ-file with phases correct or close to the real value which then can be further refined for better models.

2.7.2 Refinement:

The MTZ file obtained from the X-ray diffraction experiment contains important information such as miller indices (h,k,l), intensity, amplitude, structure factor of each reflections, completeness, and so on. All these combined are able to produce an electron density map which takes the form of a mesh and can be viewed in the modeling software COOT. A protein model skeleton and its residues are built based on the shape of the mesh where green blobs indicate the lack of something in the data unexplained by the model while red indicates excess of it. An example of COOT modeling can be seen below (see figure 9):

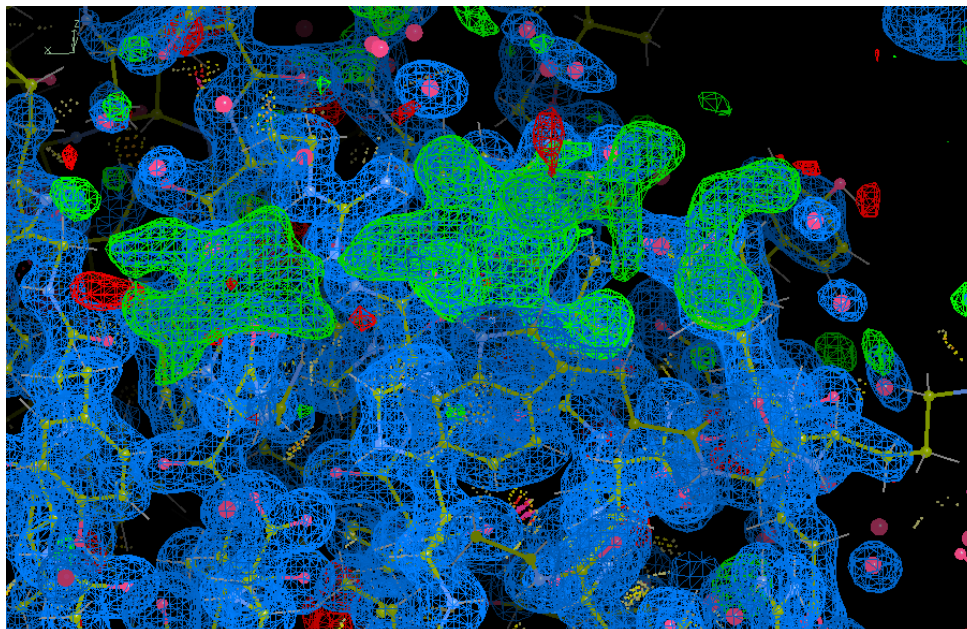


Figure 9: Example of an electron density map where the blue mesh is a “2Fo-Fc” map, green blobs are lacking something in the data unexplained by the model and red blobs have excess of it. The big green blob in the middle is likely a ligand bonded to the binding site of the protein.

After a preliminary model is created the refinement takes place, which is done using phenix.refine. A refinement begins with input of necessary files; PDB format, MTZ, and CIF files. Once required files are input the optimal refinement strategy is selected and the computer will through an iterative process refine the model. After going through selected numbers of cycles the outputs are: refined PDB format, MTZ and CIF files available for further refinement via COOT and phenix until the best structure is produced [7].

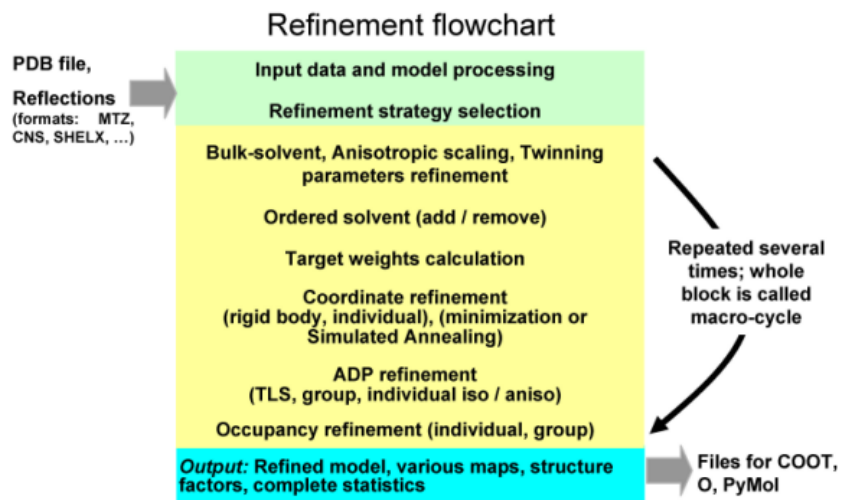


Figure 10: A flowchart of how the refinement process works, taken from Phenix’s website [7]

The goal of the refinement is to find right rotamers, angles, and electron density which through refinement will decrease R-values such as R-work and R-free which is a measurement of how good a model is; the lower the R-value the better the model. R-work tends to have a lower value than R-free but R-free is the

important R-value that determines the quality of the structure. The refinement is repeated until the R-values stop decreasing.

Requirements for good maps are: R-free about 10-15% of the resolution (e.g if the resolution is 1.2 Å the R-free value should be lower than 0.12-0.18), root-mean-square deviation from ideal bond length of less than 0.020 Å, and Ramachandran outliers of 0-2% which are aimed to be achieved during the model building of the project.

2.8 NanoDSF:

Since the mouse galectin-8N does not have any prior data for crystallization, to find factors such as melting point and optimal buffer to stabilize the protein is crucial to proceed with experiments. This can be done via a technique called “Differential Scanning Fluorimetry (DSF)” which is a simple tool similar to melting point analysis of solid compounds. Commonly in DSF to observe the fluorimetry of a compound a hydrophobic fluorescent dye is required however in nanoDSF no such thing is needed, instead it observes the fluorescence of residues such as tyrosine and tryptophan which fluoresce between 330~350 nm [10]. The two residues tend to be located inside of a protein therefore by increasing the temperature the protein unfolds thus the fluorescence increases until all the fluorescing residues are fully exposed.

The process above is done via an instrument called Prometheus NT.plex which can store up to 48 capillary tubes for one experiment and measure the fluorescence then produce graphs measuring the melting point, where the melting point of a protein can be found at the center of the sigmoidal curve in the graph.

2.9 NanoDrop:

During the experiment we want to determine the protein concentration with methods other than reading the received protocol [5] and this is done by a method called NanoDrop. NanoDrop is a method of measuring the protein concentration by observing the absorbance by sending light through a sample. Proteins tend to absorb light with a wavelength of 280 nm. The principle of NanoDrop follows “Beer-Lambert’s law” which is a simple equation of measuring absorbance and appears as following:

$$A = \epsilon * b * c \quad (2)$$

Where A is the absorbance, ϵ is the molar extinction coefficient, b is the length of the droplet, and c is the concentration.

The method is called NanoDrop but the name comes from the instrument released by Thermo Fisher Scientific: Thermo Scientific™ NanoDrop™ 2000/2000c UV-Vis Spectrometer. This machine containing adjustable light wavelengths can observe anything from DNA, RNA, or protein by adding a sample droplet of volume 1-2 μ l [11]. In the protocol received from LP3 [5] it is mentioned that all three protein concentrations were measured using this instrument, so it is important to once again measure the concentration during the middle of the project. The reason why we observe the sample twice is due to wanting to avoid concentration changes as much as possible, mainly caused by impurities or contamination of the stock protein.

2.10 Seed Bead crystallization:

When crystallization experiments have been successfully performed and produce high quality crystals, the next task is to observe and improve the reproducibility of the experiment. A perfect kit for such is called “Seed Beads” released by Hampton Research™ and is available for purchase [12]. Seed bead is

conducted by crushing successful crystals and mixing it with the reservoir solution the crystals were produced in. After extracting the crushed crystal pieces from successful drops to an Eppendorf tube, a tiny ball made out of ceramic material or steel is inserted into the tube and vortexed in order to further crush the crystals. According to the manual the best stock solution is produced when mixing the completely crushed crystals solution with 50 μ l reservoir solution. The stock solution can then be fold-diluted to 0.1, 0.01, 0.001 and so on by taking a portion of the solution and adding additional reservoir solution.

The stock or diluted seed solution is then stroked through a protein drop with the use of a fiber string tool included in the kit if the seeding is done manually. Since the fiber string (soaked in seed solution) has been stroked in a line the crystals will also get produced in a straight line often having round or diamond shapes and similar sizes. It also works to do seeding automatically with the help of robotic tools such as Mosquito Crystal™ by pipetting the seed solution into the protein drop directly. The crystals will form randomly in the drop but will maintain similar size and shape.

Furthermore, the Seed bead kit can be used not only for reproducing crystals but also for producing new crystals for homologous proteins. For instance, seed solution made out of variant M of galectin-8N crystals can be utilized for variant V with a sequence identity of 98.1%. In theory, the method may also be possible for mouse galectin-8N which has a sequence identity of 80.6% to human galectin-8N variant M. The seeding process of mouse galectin-8N is done by combining automatic seeding with commercial screens to observe how well unknown crystals may or may not crystallize in several of the available conditions. During this experiment, seeding screen of mouse galectin-8N is done in Morpheus commercial screen which has one of the highest success rates of crystallizing unknown proteins according to Molecular Dimensions [9].

2.11 Serendipity and accidents:

During the project crystals were found to exist in the Eppendorf tube containing the stock protein solution for human galectin-8N version V. The protein was stored exactly like the other two proteins (human galectin-8N M and mouse galectin-8N) with similar conditions; in PBS buffer and stored at 4 °C, which makes it unexpected to find crystals considering crystals were not observable in the remaining two protein stock solutions. Suspecting salt crystals were formed, the protein concentration was measured with NanoDrop and resulted in a concentration decrease from 4.4 mg/ml to 2.2 mg/ml (extinction coefficient of 1.588), which tends to happen when proteins crystallize. After observing the crystals at MAX IV the crystals were confirmed to be large protein crystals (see figure 11) with atomic resolutions (<1.0 Å).

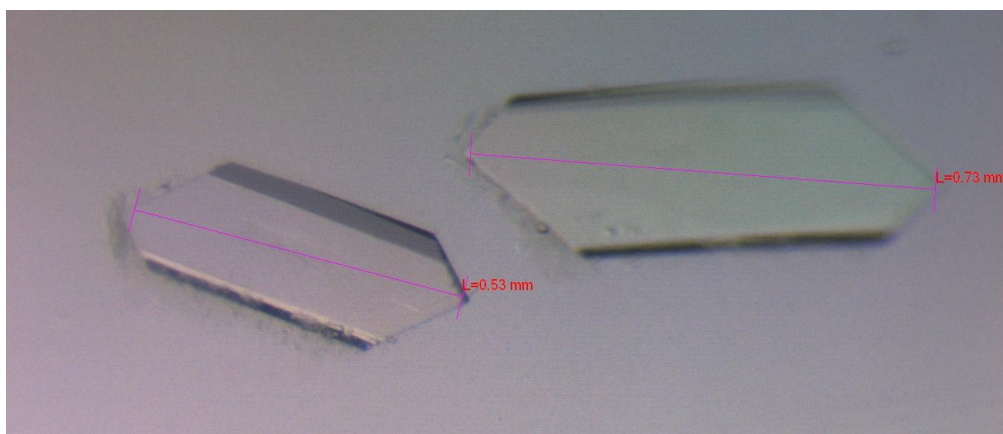


Figure 11: A picture of galectin-8N variant V taken from a microscope. Two well-shaped protein crystals with length of 0.53 and 0.73 mm were photographed, but in the Eppendorf tube there were over hundreds of similar crystals.

3.0 Methods:

3.1 Commercial Screening (JCSG+, PACT & Morpheus):

During the start of this experiment three types of commercial screens all obtained from Molecular Dimensions™ were utilized which were: JCSG+, PACT Premier and Morpheus.

The human galectin-8N variant M with concentration 13.1 mg/ml was diluted to 1.31 mg/ml and variant V with concentration 4.4 mg/ml to 0.44 mg/ml were both prepared in two separate Morpheus commercial screens with three protein wells in each well (total 96). 6 µl x 8 wells in the Mosquito Crystal™ robot were filled with protein and were distributed to all 96x3 wells of the commercial screens. Reservoir solution was then added into all the protein wells in volume ratios of: 1:2, 1:1, and 2:1 (protein to reservoir), in this case the volumes were 100:200 nl, 100:100 nl, and 200:100 nl

The mouse galectin-8N with concentration 2.6 mg/ml diluted to 0.234 mg/ml was experimented on using all three types of commercial screening plates (JCSG+, PACT premier, and Morpheus). Two of the protein wells on the Mosquito Crystal robot were filled with 4.2 µl x 8 and an additional 1.8 µl x 8 of protein+ligand MH-3-80 in the last well, and were distributed to all 96x3 wells. The volume ratios of the protein to reservoir were: 1:2, 1:1, and 1:1, in this case 100:200 nl, 100:100 nl, and 100:100 nl.

All the five commercial screens were brought and stored into a plate hotel (JANSi UVEXps 256/600) at a temperature of 20 °C and were observed with a consistent frequency for several days at LP3.

3.2 NanoDSF melting point analysis and Buffer screens (RUBIC & Durham):

3.2.1 RUBIC Screen:

The mouse galectin-8N was diluted by adding 144 µl to 106 µl buffer (Ströhagen's buffer-alpha) which resulted in a volume of 250 µl and a protein concentration of 1.5 µg/µl. 2µl of the diluted protein solution was pipetted into the first eight wells (column) on the RUBIC screen with aluminum foil covering the wells and was then centrifuged (in 1500 g) for about 5 seconds to achieve a mixture with the buffer solution on the bottom of the well. The process above was repeated until 48 wells (6 rows x 8 columns) were mixed with the diluted protein and buffer. Using capillary tubes all 48 solutions were collected and rowed up inside of the NanoDSF instrument Prometheus NT.plex and the fluorescence was gradually measured from a temperature starting at 20°C to 95°C with a pacing of 1°C/minute and an excitation power of 40%. During the melting process the remaining 48 wells were mixed with diluted protein and buffer then the solution mixture was collected utilizing the exact same method as earlier. After 1 hour and 30 minutes the denatured 48 samples were discarded and newly 48 samples were loaded into the nanoDSF instrument and was measured at the same conditions.

3.2.2 Durham Screen & mouse galectin-8N melting point:

The mouse galectin-8N was diluted by adding 192 µl protein to 58 µl buffer (Ströhagen's buffer-alpha) resulting in a solution of 250 µl diluted protein. The rest of the method proceeded exactly the same as RUBIC-screen (see section above), and the fluorescence was measured for a total of 96 samples. After the denaturation was complete two additional samples of diluted protein and diluted protein mixed with

ligand MH-8-30 were inserted into the instrument with similar temperature increase and interval as the two experiments above but with an excitation power of 7% and the melting point was measured. All the data obtained was later observed using the computer software PR.ThermControl and all appropriate buffers changing the melting point with +1°C were noted down.

3.3 Screen Optimization (Hanging Drop):

By using the data and results obtained from the Morpheus commercial screen the Human galectin-8N variant M and V were further crystallized manually using the method “Hanging drop”. To imitate the conditions of the Morpheus screen five types of solution were prepared: buffer, precipitant (two mixtures), additives and water.

The most successful buffer was a mixture of 61.2% MES and 38.8% Imidazole with pH 6.5. 100 ml of 1M MES solution was prepared by dissolving 21.34 g MES in water, and 100 ml of 1M Imidazole solution was prepared by dissolving 6.82 g Imidazole in water. The two solutions were mixed in ratio of 61.2/38.8 by adding 6.12 ml of MES-solution to 3.88 ml Imidazole resulting in a 10 ml 1M buffer solution with pH 6.48 measured using a calibrated pH-meter. The solution was shaken and stored in a refrigerator at 4°C

Two of the best types of precipitant mixtures obtained from Morpheus were prepared named 1 and 4. Precipitant mixture 1 was a mixture of 40% PEG550MME, 20% PEG20000 and rest water, which was prepared by adding 12.53 g of PEG550MME to 6.27 g of PEG20000 and filled up with water up to the 31.3 ml mark of a 50ml tube. The other mixture 4 was a combination of 25% MPD, 25% PEG1000, 25% PEG3350 and rest water which was prepared by adding 12.58 g MPD with 12.508 g PEG1000, 12.54 g PEG3350 and water was filled up to the 50ml mark of a 50ml tube. The two precipitant mixtures were stored at room temperature (circa 20°C).

Table 3: The table below showing the two precipitant mixtures (1 & 4) utilized during this experiment. In addition, their concentration and mixture contents are also noted down.

Precipitant	Total Concentration	Mixture Contents
Mixture 1	60%	40% PEG550MME 20% PEG20000
Mixture 4	75%	25% MPD 25% PEG1000 25% PEG3350

The best and available additive mixture (alcohol) giving crystallization results in Morpheus was prepared by mixing 0.3 M 1,6-Hexadiol, 0.3 M 1-Butanol, 0.32 M 1,2-Propanediol, and 0.38 M 2-Propanol with water. The mixture was prepared by adding 1.78 g 1,6-Hexadiol, 1.12 g 1-Butanol, 1.20 g 1,2-Propanediol, 1.13 g 2-Propanol, and was filled with water until the 50 ml mark of a 50 ml tube. The mixture was shaken until the mixture was homogeneous and was then stored in a refrigerator at 4°C.

Hanging drops experiments were prepared on two 5x3 EasyXtal plates and 1 ml of reservoir solution containing; buffer, precipitant mixture, additive mixture, water and TCEP was mixed in required ratios (see table 1 and 2 in appendix) for each of the 15 wells.

2 μ l x 2 drops of the reservoir solution was extracted from the well and pipetted on the lid of the EasyXtal plates. 2 μ l of human galectin-8N variant M (13.03 mg/ml) was pipetted into one of the drops and 4 μ l of human galectin-8N variant V (2.15 mg/ml) on the other. This process was repeated for all remaining 14 wells and all the lids were screwed on. Both hanging-drop plates were stored in 20°C letting them equilibrate and crystallize for a few days.

3.3.1 Ligand Soaking:

In preparation, ligand MH-3-80 (1.3 mg) received in a small jar was dissolved in 62 μ l DMSO which resulted in a concentration of about 50 mM.

After observing that crystals formed in the manual screens, 0.2 μ l of dissolved MH-3-80 was pipetted into the protein drops in well A4 (0.08 M alcohol & 32% PM4) and 0.3 μ l in B4 (0.12 M & 32% PM4). After soaking the crystals in ligand for 24 hours, a few crystals in wells A4 and B4 were extracted using loops and an additional 0.2 μ l of MH-3-80 was inserted into well A4 and soaked for an additional hour, then extracted.

3.3.2 Ligand Cocrystallization:

In preparation, the two ligands MH-1FH (1.119 mg) and NV9.1 (1.445 mg) also received in small jars were dissolved in DMSO by adding 42.3 μ l DMSO to MH-1FH and 52 μ l in NV9.1. The final concentration of the ligand was 50 mM for MH-1FH and 100 mM for NV9.1.

Two EasyXtal plates for each variant of the human proteins were prepared and mixed with the same compounds and amounts from the screen optimization section (see table 2 in appendix). Two Eppendorf tubes each containing 40 μ l of the human galectin-8N variant M (13.1 mg/ml) were mixed with 4 μ l of MH-1FH ligand in one tube and 4 μ l of NV9.1 in the other. In addition, three Eppendorf tubes each containing 70 μ l human galectin-8N variant V (2.15 mg/ml) were mixed with 7 μ l of ligands MH-3-80, MH-1FH, and NV9.1 in each tube.

Two 2 μ l drops of reservoir solution were pipetted on the lid and 2 μ l of each protein+ligand solution was pipetted into the two reservoir drops for the M-variant. On the other plate (V-variant), three drops of 2 μ l reservoir solution was pipetted onto the lid and 4 μ l of each protein+ligand solution was pipetted into the three drops. After screwing on the lid the two processes were repeated a total of 14x2 additional times and both plates were stored in 20°C letting them crystallize.

3.4 Crystal extraction and data refinement:

A universal puck (UniPuck) containing room for 16 crystals was opened and sunk into liquid nitrogen. Using loops of polymer sufficient amounts of crystals suitable for analysis were harvested, by fishing up the wanted crystals manually. After the crystal attached onto the loop, immediately the crystals were sunk in liquid nitrogen and placed in the puck. This process was repeated several times until all the wanted samples were placed in the pucks.

All selected crystals suitable for analysis were: 13 crystals from Morpheus (both variant M and V), 3 manually crystallized variant V, 8 manually crystallized variant V soaked in the same ligand, 3 from the spontaneous crystallized variant V, and 16 cocrystallized with three different kind of ligands (both variant M and V). This resulted in 43 samples and required three pucks in total to execute.

The pucks were stored in a dewar containing liquid nitrogen and were delivered to MAX IV lab for analysis at the beamline. All the data obtained were analyzed and data refinement was performed several times using softwares such as COOT, phenix.refine and eLBOW to model the necessary proteins (total five models).

3.5 Mouse protein buffer exchange (Centrifuging to citrate):

The mouse galectin-8N did not result in any crystals therefore the buffer was changed into a different one. From the results of RUBIC and Durham buffer screens the appropriate buffer for the protein resulted in; a 50mM citrate buffer at pH 6.5 (see results). 10 ml of 50 mM citrate buffer was prepared by mixing 375 μ l 4 M NaCl, 20 μ l 0.5 M TCEP, 313 μ l 1.6 M Citrate, and 9.29 ml of water.

A volume of 2.5 ml 50 mM citrate buffer was pipetted into a centrifuge tube (Sartorius Vivaspin 2, 5000 MWCO) in addition the remaining 2.6 mg/ml mouse galectin-8N (about 400 μ l) was added into the tube. The centrifuge tube containing the protein and citrate buffer was placed in a centrifuge together with a counter-weight tube and spun at 4000 rpm, 4°C for 30 minutes. Due to nothing significant happening, the process was repeated again at 4000 rpm, 4000g, 4°C for another 30 minutes. Due to nothing noticeable happening again, the process was repeated four times at 4000 rpm, 4000g, 12°C until the buffer exchanged from PBS to citrate resulting in a total centrifuge time of four hours.

2 μ l of the mouse protein now in the citrate buffer was pipetted and placed on NanoDrop and the concentration was measured to be 2.2 mg/ml.

3.6 Mouse Galectin-8N Screening (5 types):

Due to JCSG+, PACT Premier, and Morpheus screens not resulting in anything for mouse galectin-8N in the PBS buffer five new screens were prepared, this time in citrate buffer. The five new commercial screens utilized were: JCSG+, PACT Premier, Morpheus, BCS, and ProPlex. The mouse protein with 100 mM lactose binded to it with concentration of 2.2 mg/ml was pipetted into 3 μ l x 8 wells of the Mosquito Crystal™, then was distributed to all 12 protein wells (only one subwell) and mixed with the reservoir solution. This process was repeated for the remaining four screens. The five plates were then stored in the plate hotel (JANSi UVEXps 256/600) at 20°C and were observed frequently using computer softwares at LP3.

3.7 Seed Beads (human galectin-8N variant M):

From the manual screening plate (from section "Screen Optimization" see table 2 in appendix) containing 75% PEG precipitant, all M-variants of the human protein crystals from well A3 were crushed into pieces using a glass rod from the Seed Bead kit. Using an automated pipette the crushed protein crystal solution was extracted and placed in an Eppendorf tube. A tiny ceramic ball was dropped into the Eppendorf tube and was vortexed for 30 seconds and placed in ice, then 50 μ l of the reservoir solution in well A3 was pipetted into the tube (Stock). 5 μ l the stock solution was extracted and placed in a new tube and 45 μ l of the same reservoir solution (A3) was added and then vortexed (Dilution 1). 5 μ l of the first dilution was extracted, placed in a new tube and 45 μ l reservoir solution was added then vortexed (Dilution 2). Lastly, 5 μ l of the second dilution was pipetted to a new tube and 45 μ l of reservoir solution was added, then vortexed (Dilution 3).

An EasyXtal plate (5x3) was prepared where eight wells were utilized for experiment. In all the wells were 2 μ l TCEP, 100 μ l buffer 1, and 100 μ l alcohol mixture added. In addition, from left to right were 25, 27.5,

30.0, and 32.5% of precipitant mixture 4 added, and rest water to reach a volume of 1 ml in each well (see table 3 in appendix). Two drops of 2 μ l of galectin-8N variant M was pipetted onto the lid for all eight wells (in total 16 drops for 8 wells), and using a fiber string tool contained in the Seed Bead kit the string was dipped into each of the stock and three dilution solutions then stroked in a line through the protein drop. All eight lids were screwed on and stored at a temperature of 20°C.

Due to no crystals being formed after three days in any of the wells, 0.5 μ l of stock solution was pipetted into all proteins drop 2 (the drops containing dilution 1) and additionally the stock solution was stroked through the protein drop 1 (stock) and the result was observed. Immediately, crystals formed in drop 2 with high PEG concentration (>30%) and nothing in protein drops 1. To further test this out, in all protein drops 3 and 4 0.5 μ l of respective dilution solution 2 and 3 was pipetted into all drops and incubated at 20°C for a day.

3.8 Seeding with commercial screen:

A Morpheus commercial screen plate with 96 wells was prepared and placed inside of the Mosquito crystal™ robot, loaded with with eight side wells pipetted with 2.4 μ l mouse galectin-8N (2.2 mg/ml) each and additionally eight new side wells with 1.2 μ l seed stock solution each. This resulted in, in each well there were 200 nl protein and 100 nl seed stock in each of the 96 protein-wells, then 100 nl of the reservoir solution was extracted from the reservoir-well and added into the protein well. The plate was sealed with plastic and stored at 20°C and observed for days with regular frequency at LP3.

4.0 Results:

4.1 Human Galectin-8N variant M and V Results:

4.1.1 Commercial Screens:

From prior Morpheus data by Ströhagen's experiments [1] it was expected for galectin-8N variant M to produce crystals and result in the same, while variant V presumably producing similar results due to slight difference in three amino acids not affecting the binding site. The crystallization results were as following:

Just for clarification, the names of the crystals come from the well they were extracted from. The starting letter in the crystal well indicates the column, the number after indicates the row and the last number after _ indicates the subwell. For instance, D4_1 means "column D, row 4 and subwell 1".

Variant M:

Table 4: Results of all the wells containing crystals of human galectin-8N variant M in Morpheus commercial screen, and in what conditions they crystallized in.

Crystal well	Protein:Reservoir Ratio	Additives	Buffer	pH	Precipitant Mixture
D4_1	1:2	Alcohol	1	6.5	4
D4_2	1:1	Alcohol	1	6.5	4
D8_1	1:2	Alcohol	2	7.5	4
D12_1	1:2	Alcohol	3	8.5	4

E4_1	1:2	Ethylene Glycols	1	6.5	4
E4_2	1:1	Ethylene Glycols	1	6.5	4
E8_1	1:2	Ethylene Glycols	2	7.5	4
E12_1	1:2	Ethylene Glycols	3	8.5	4
F4_1	1:2	Monosaccharides	1	6.5	4
F4_2	1:1	Monosaccharides	1	6.5	4
F8_1	1:2	Monosaccharides	2	7.5	4
H4_1	1:2	Amino Acids	1	6.5	4
H8_1	1:2	Amino Acids	2	7.5	4
H12_1	1:2	Amino Acids	3	8.5	4

Variant V:

Table 5: Results of all the wells containing crystals of human galectin-8N variant V in Morpheus commercial screen, and in what conditions they crystallized in.

Crystal well	Protein/Reservoir Ratio	Additives	Buffer	pH	Precipitant Mixture
D1_1	1:2	Alcohol	1	6.5	1
D1_2	1:1	Alcohol	1	6.5	1
D1_3	2:1	Alcohol	1	6.5	1
E1_1	1:2	Ethylene Glycols	1	6.5	1
E1_2	1:1	Ethylene Glycols	1	6.5	1
E4_1	1:2	Ethylene Glycols	1	6.5	4
E4_2	1:1	Ethylene Glycols	1	6.5	4
E8_1	1:2	Ethylene Glycols	2	7.5	4
E9_1	1:2	Ethylene Glycols	3	8.5	1
E12_1	1:2	Ethylene Glycols	3	8.5	4
E12_2	1:1	Ethylene Glycols	3	8.5	4
F1_1	1:2	Monosaccharides	1	6.5	1
F1_2	1:1	Monosaccharides	1	6.5	1
F4_1	1:2	Monosaccharides	1	6.5	4
F4_2	1:1	Monosaccharides	1	6.5	4
F5_1	1:2	Monosaccharides	2	7.5	1

F8_1	1:2	Monosaccharides	2	7.5	4
F9_1	1:2	Monosaccharides	3	8.5	1
F12_1	1:2	Monosaccharides	3	8.5	4
H1_1	1:2	Amino Acids	1	6.5	1
H1_2	1:1	Amino Acids	1	6.5	1

From the two screen results above, it is possible to observe that variant V is more versatile in crystallization in several conditions over M. The majority of both protein variants have preferences of crystallizing in conditions mainly containing: ethylene glycols and buffer of pH 6.5.

4.1.2 Manual/Optimization Screens:

Buffer (pH 6.5), Precipitant Mixture 1 (60% PEG), and Alcohol Additives: No crystals were formed.

Buffer (pH 6.5), Precipitant Mixture 4 (75% PEG), and Alcohol Additives:

Table 6: Results of all wells containing galectin-8N variant M and V crystals of the manual/optimization screen. M indicates variant M and V indicates V.

Mix 4 (75%)	32.5%	35%	37.5%	40%	42.5%
0.08M	M	M	M	M & V	M & V
0.12M	M	M	M & V	M & V	M & V
0.16M	M	M & V	M & V	M & V	M & V

The result above shows that galectin-8N in presence of alcohol additives will have a higher probability to crystallize in precipitant mixture 4 over 1. In the conditions tested containing alcohol additives and precipitant mixture 4 variant M will 100% crystallize, while variant V requires high concentration of either alcohol, precipitant, or both.

4.1.3 Ligand Cocrystallization Manual Screens:

Buffer (pH 6.5), Precipitant Mixture 4, Alcohol Additives, and Galectin-8N variant M (1 week after):

Table 7: Results of all wells containing galectin-8N variant M crystals crystallized with ligand MH-1FH and NV9.1. The empty boxes indicate no crystals being produced.

M variant	32.5%	35%	37.5%	40%	42.5%
0.08M				Both	NV9.1
0.12M			NV9.1	Both	Both
0.16M	NV9.1	NV9.1	NV9.1	NV9.1	Both

Buffer (pH 6.5), Precipitant Mixture 4, Alcohol Additives, and Galectin-8N variant V (1 week after) :

Table 8: Results of all wells containing galectin-8N variant V crystals crystallized with ligand MH-3-80, MH-1FH and NV9.1. The empty boxes indicate no crystals being produced.

V variant	32.5%	35%	37.5%	40%	42.5%
0.08M				MH-3-80 NV9.1	MH-3-80 NV9.1
0.12M				MH-3-80 NV9.1	All three
0.16M			MH-3-80 NV9.1	MH-3-80 NV9.1	All three

Observable in the two tables of ligand cocrystallization results above, the proteins presumably containing ligand NV9.1 had a higher probability of forming crystals over the remaining two ligands (if the ligand has actually been inserted) while MH-1FH was the most difficult and took the longest to crystallize for both variants of galectin-8N.

4.1.4 Data collection and refinement:

The following tables present the results of all crystals brought to the beamline (see table 9 , 10 and 11). Some crystal data are discarded due to various reasons due to; similar crystal data already existing, no crystals on the loop or not providing necessary data for the project.

For clarity the naming is constructed in a way where the small letter (m or v) indicates the variant of the human protein, big letter indicates the column, number in the middle indicates the row, the small letter a indicates subwell 1 of Morpheus, and lastly the number after _ is not so important only indicating the crystal sample which is in between 1-3.

Morpheus and Manually:

Table 9: Results of data of crystals extracted from the Morpheus and manual screens. All data were obtained from the BioMAX beamline at MAX IV.

Crystal	Overall Resolution Range (Å)	Outer Shell Resolution (Å)	Space Group	Multiplicity (Outer)	CC1/2 (%) (Outer)	Rmerge (%) (Outer)	I/Sigma (Outer)	Completeness (%) (Outer)
vE1a_1	50.1-1.8	1.9-1.8	P2 ₁ 2 ₁ 2 ₁	13.0 (8.4)	100 (60)	12.8 (152.0)	10.4 (1.3)	94.4 (67.0)
vE1a_2	50.2-1.7	1.8-1.7	P2 ₁ 2 ₁ 2 ₁	12.8 (10.3)	100 (70)	7.0 (130.6)	17.3 (1.3)	91.3 (47.0)
vH1a	50.2-1.6	1.7-1.6	P2 ₁ 2 ₁ 2 ₁	13.2 (12.6)	100 (60)	7.2 (181.6)	16.2 (1.4)	94.4 (62.9)
mD4a_1	50.1-1.4	1.6-1.4	P2 ₁ 2 ₁ 2 ₁	13.4 (13.8)	100 (70)	7.0 (175.1)	16.3 (1.5)	95.1 (73.2)
mH4a_1	50.2-1.4	1.5-1.4	P2 ₁ 2 ₁ 2 ₁	13.4 (11.5)	100 (50)	8.7 (180.6)	12.9 (1.6)	92.2 (58.9)
mH4a_2	42.5-1.6	1.7-1.6	P2 ₁ 2 ₁ 2 ₁	13.3 (13.8)	100 (70)	7.1 (179.7)	14.7 (1.4)	94.6 (67.5)

mH8a	50.1-1.5	1.6-1.5	P2 ₁ 2 ₁ 2 ₁	13.3 (13.9)	100 (70)	6.8 (156.2)	16.7 (1.6)	94.6 (67.8)
vC2_1	50.3-1.2	1.2-1.2	P2 ₁ 2 ₁ 2 ₁	12.9 (7.5)	100 (60)	6.7 (114.5)	13.3 (1.6)	95.1 (60.1)

vD1a did not have any important crystal for observing, while vE1b, vD1c, vH1b, vF4a, mD4a_2, vC2_2, vC2_3 were not required since already good data had been collected.

Ligand soaking and serendipity crystals (all variant V):

Table 10: Results of data of crystals of human galectin-8N variant V extracted from the ligand soaking screens and serendipity crystals, All data were obtained from the BioMAX beamline at MAX IV.

Crystal	Overall Resolution Range (Å)	Outer Shell Resolution (Å)	Space Group	Multiplicity (Outer)	CC1/2 (%) (Outer)	Rmerge (%) (Outer)	I/Sigma (Outer)	Completeness (%) (Outer)
A4_1	50.3-1.2	1.4-1.2	P2 ₁ 2 ₁ 2 ₁	12.6 (11.0)	100 (60)	16.1 (248.6)	13.2 (1.8)	94.1 (62.7)
A4_2	45.4-1.8	1.9-1.8	P222	12.3 (11.8)	100 (80)	14.7 (96.5)	12.9 (2.7)	99.8 (97.6)
B4_3	46.1-1.7	1.9-1.7	P2 ₁ 2 ₁ 2 ₁	13.1 (13.2)	100 (50)	20.1 (241.3)	11.9 (2.2)	94.1 (65.3)
A4_2_1	50.1-1.3	1.4-1.3	P2 ₁ 2 ₁ 2 ₁	13.3 (14.2)	100 (60)	32.0 (221.3)	10.7 (1.6)	89.8 (70.6)
A4_2_3	50.1-1.3	1.4-1.3	P2 ₁ 2 ₁ 2 ₁	13.3 (12.7)	100 (60)	12.4 (186.3)	13.8 (1.4)	95.0 (63.6)
V_2	42.0-0.9	1.0-0.9	P2 ₁ 2 ₁ 2 ₁	13.4 (13.3)	100 (70)	5.6 (162.2)	17.7 (1.6)	95.5 (64.7)

The data of B4_1, B4_2, A4_2_2, V_1, and V_3 were not gathered due to various factors such as: crystals were not found, Rmerge was too high, or similar data already gathered.

Ligand cocrystallization:

Table 11: Results of data of crystals of human galectin-8N variant M and V extracted from the ligand cocrystallization screens, All data were obtained from the BioMAX beamline at MAX IV. The naming of B, R and N indicates the inserted ligand where B is MH-1FH, R is NV9.1 and N is MH-3-80.

Crystal	Overall Resolution Range (Å)	Outer Shell Resolution (Å)	Space Group	Multiplicity (Outer)	CC1/2 (%) (Outer)	Rmerge (%) (Outer)	I/Sigma (Outer)	Completeness (%) (Outer)
vC5_B1	50.1-1.4	1.6-1.4	P2 ₁ 2 ₁ 2 ₁	13.1 (11.2)	100 (60)	8.9 (157.9)	14.1 (1.7)	88.8 (59.7)
vC5_B2	50.2-1.3	1.3-1.3	P2 ₁ 2 ₁ 2 ₁	13.2 (12.2)	100 (60)	7.0 (174.9)	16.9 (1.5)	94.7 (59.6)
vA5_R1	50.2-1.2	1.3-1.2	P2 ₁ 2 ₁ 2 ₁	13.4 (13.8)	100 (70)	9.8 (169.4)	15.4 (1.6)	93.3 (58.7)
vB4_R	50.2-1.3	1.4-1.3	P2 ₁ 2 ₁ 2 ₁	13.5 (13.8)	100 (70)	6.5 (158.8)	16.2 (1.5)	93.0 (59.7)
mB4_B1	50.2-1.2	1.3-1.2	P2 ₁ 2 ₁ 2 ₁	13.2 (11.2)	100 (70)	7.9 (151.8)	14.4 (1.6)	93.8 (58.8)
mB4_B2	50.1-1.3	1.3-1.3	P2 ₁ 2 ₁ 2 ₁	13.4 (13.3)	100 (50)	11.1 (239.8)	11.6 (1.4)	93.3 (52.5)
mB3_R2	50.2-1.5	1.6-1.5	P2 ₁ 2 ₁ 2 ₁	13.3 (13.6)	100 (50)	11.0 (202.1)	12.0 (1.5)	94.5 (55.7)

vA5_R2, vC5_B3, mC5_B, mB3_R1, mC4_N2, mA5_N1, and mA5_N2 were discarded due to not providing necessary data.

4.1.5 Data refinement and protein modeling:

From the all crystals formed, few of the best crystals suitable for further analysis were chosen then refined and modeled. The chosen crystals were as following (see table 12):

Table 12: A table of all crystals chosen for further observation. In addition to the data from the MAX IV the protein variant and ligand inserted are also noted down. The proteins noted down all have the space group P2₁2₁2₁ and an overall CC1/2 of 100 (see table 10 & 11 for outer values).

Crystal	Overall Resolution Range (Å)	Outer Shell Resolution (Å)	Multiplicity (Outer)	Rmerge (%) (Outer)	I/Sigma (Outer)	Completeness (%) (Outer)	Protein Variant	Ligand Inserted
A4_1	50.3-1.2	1.4-1.2	12.6 (11.0)	16.1 (248.6)	13.2 (1.8)	94.1 (62.7)	V	MH-3-80
A4_2_3	50.1-1.3	1.4-1.3	13.3 (12.7)	12.4 (186.3)	13.8 (1.4)	95.0 (63.6)	V	MH-3-80
V_2	42.0-0.9	1.0-0.9	13.4 (13.3)	5.6 (162.2)	17.7 (1.6)	95.5 (64.7)	V	*Nothing*
vC5_B2	50.2-1.3	1.3-1.3	13.2 (12.2)	7.0 (174.9)	16.9 (1.5)	94.7 (59.6)	V	MH-1FH
vA5_R1	50.2-1.2	1.3-1.2	13.4 (13.8)	9.8 (169.4)	15.4 (1.6)	93.3 (58.7)	V	NV9.1
mB4_B1	50.2-1.2	1.3-1.2	13.2 (11.2)	7.9 (151.8)	14.4 (1.6)	93.8 (58.8)	M	MH-1FH
mB3_R2	50.2-1.5	1.6-1.5	13.3 (13.6)	11.0 (202.1)	12.0 (1.5)	94.5 (55.7)	M	NV9.1

4.1.6 Ligand observed in the binding site:

From the seven models analyzed using phenix.refine and COOT it was possible to observe that most of them had the ligand inserted into the binding site. The green blob in the protein model indicates electron density missing which shows a shape of a ligand at the binding site and the following figure shows an example of a ligand possible to observe at the binding site (see figure 12):

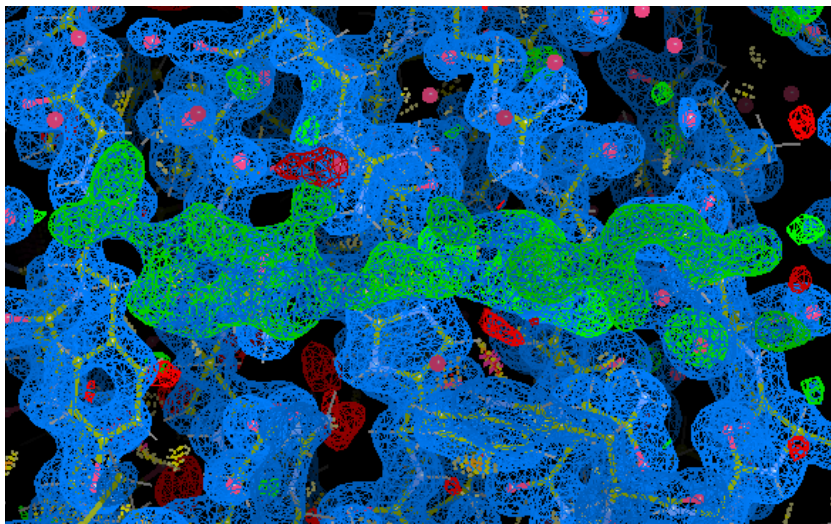


Figure 12: Picture of an electron density map taken from COOT of crystal vC5_B2. From the shape of the green blob, it is possible to observe that it contains ligand MH-1FH.

The protein crystal for variant V both short (A4_1) or long soaked (A4_2_3) in ligand MH-3-80 contained the wanted compound in the binding site. Both crystals of variant M and V cocrystallized with MH-1FH also both contained the ligand. On the other hand, the crystals of variant M and V cocrystallized with NV9.1 did not contain the ligand, and instead possessed lactose at the binding site. Lastly, the spontaneously crystallized V variant also contained lactose in the binding site, which indicates the fact that the protein was not an apoprotein to begin with. In the protocol received [5] it was mentioned that all proteins were likely apoproteins, however the models showed different results where all proteins contained lactose at the binding site.

4.1.7 Refinements and Models:

The following tables show all the refined models, refined several times until reaching the lowest possible R-free value. The figure below the table shows the model created, viewed in the computer program PyMOL.

Table 13: Final refinement results of the spontaneous variant V with all important variables to the right.

Validations:	Name: V_2
Resolution	30.57 - 0.89 Å
R-work	0.1346
R-free	0.1439
Bonds	0.010
Angles	1.150
Ramachandran favored	100.0%
Ramachandran outliers	0.0%

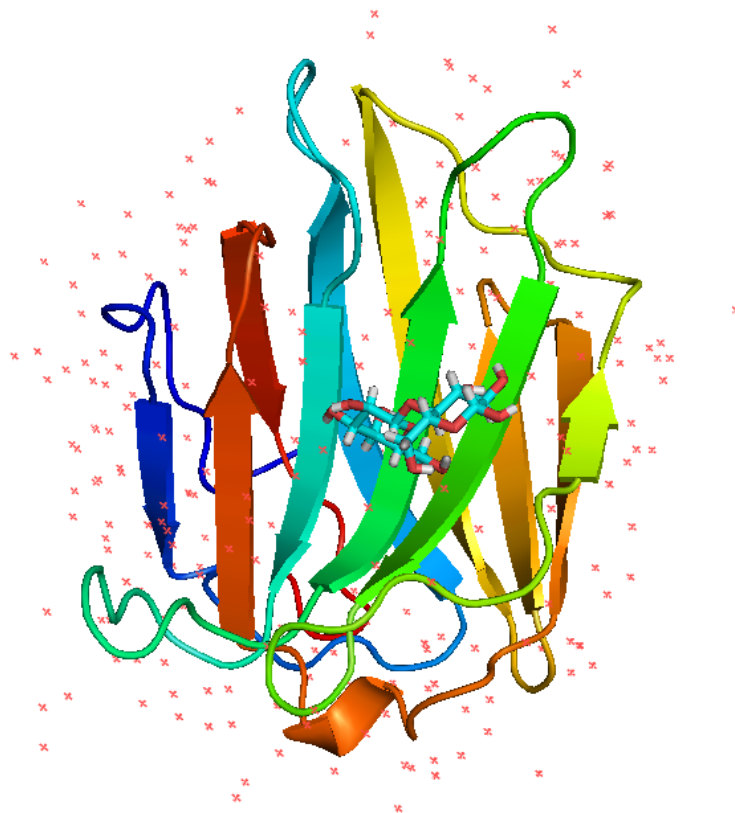


Figure 13: Model of chain A of human galectin-8N variant V with lactose in the binding site (V_2), viewed in PyMOL.

Table 14: Final refinement results of variant V soaked in ligand MH-3-80, with all important variables.

Validations:	Name: vA4_2_3
Resolution	42.50 - 1.31 Å
R-work	0.1515
R-free	0.1871
Bonds	0.009
Angles	1.158
Ramachandran favored	98.98%
Ramachandran outliers	0.0%

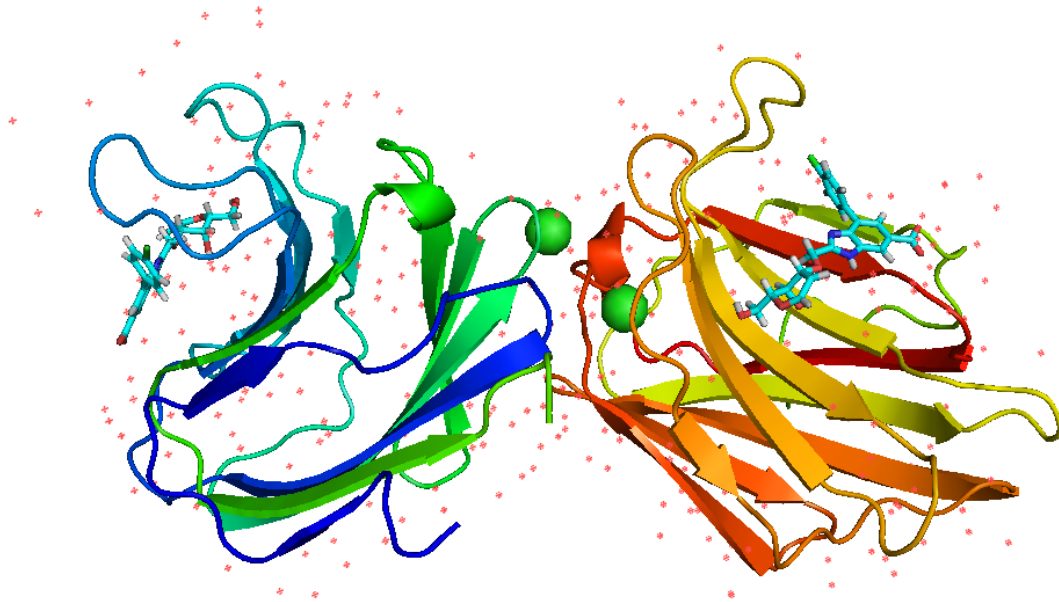


Figure 14: Model of human galectin-8N variant V with ligand MH-3-80 in the binding sites (vA4_2_3), viewed in PyMOL. Chain A is in the color of blue/cyan/green, while chain B is in red/orange/yellow.

Table 15: Final refinement results of variant M cocrystallized in ligand MH-1FH, with all important variables.

Validations:	Name: mB4_B1
Resolution	50.19 - 1.21 Å
R-work	0.1517
R-free	0.1834
Bonds	0.008
Angles	1.032
Ramachandran favored	98.98%
Ramachandran outliers	0.0%

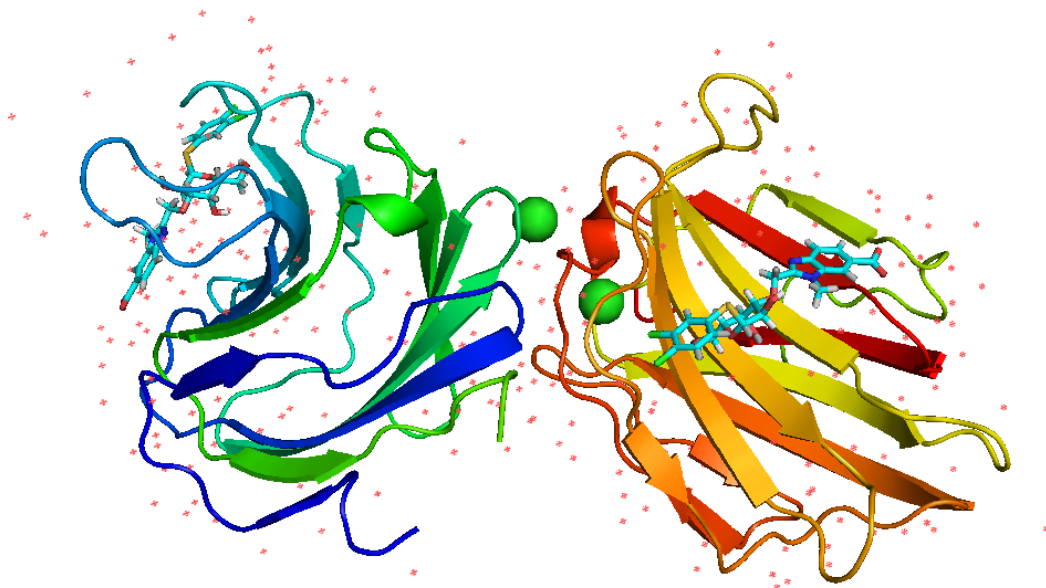


Figure 15: Model of human galectin-8N variant M with ligand MH-1FH in the binding sites (mB4_B1), viewed in PyMOL. Chain A is in the color of blue/cyan/green, while chain B is in red/orange/yellow.

Table 16: Final refinement results of variant V cocrystallized in ligand MH-1FH, with all important variables.

Validations:	Name: vC5_B2
Resolution	50.21 - 1.25 Å
R-work	0.1483
R-free	0.1807
Bonds	0.008
Angles	1.127
Ramachandran favored	98.98%
Ramachandran outliers	0.75%

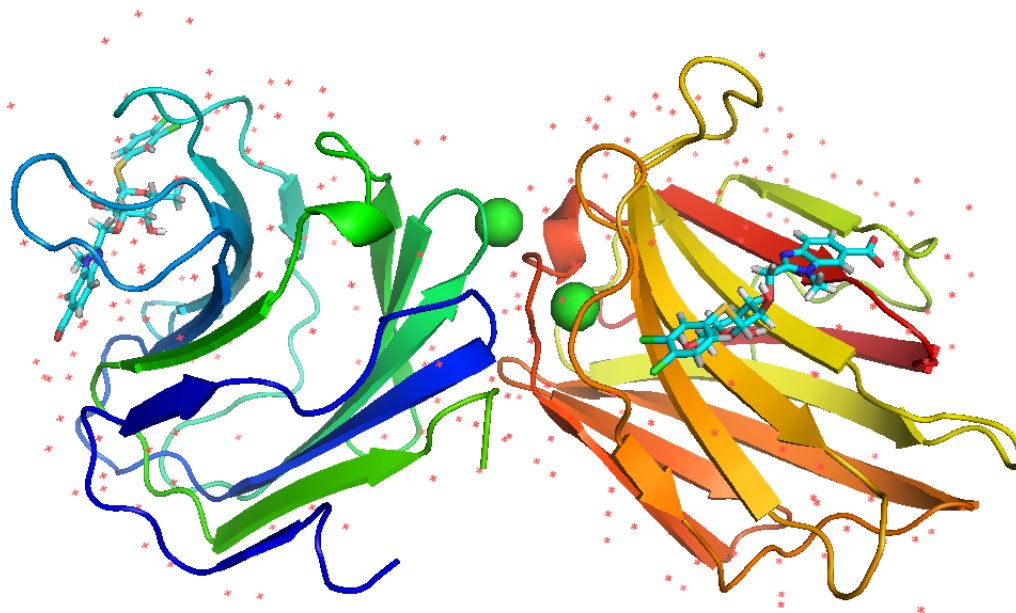


Figure 16: Model of human galectin-8N variant V with ligand MH-1FH in the binding sites (vC5_B2), viewed in PyMOL. Chain A is in the color of blue/cyan/green, while chain B is in red/orange/yellow.

Table 17: Final refinement results of variant V cocrystallized in ligand NV9.1, with all important variables.

Validations:	Name: vA5_R1
Resolution	50.21 - 1.20 Å
R-work	0.1522
R-free	0.1853
Bonds	0.012
Angles	1.327
Ramachandran favored	99.32%
Ramachandran outliers	0.0%

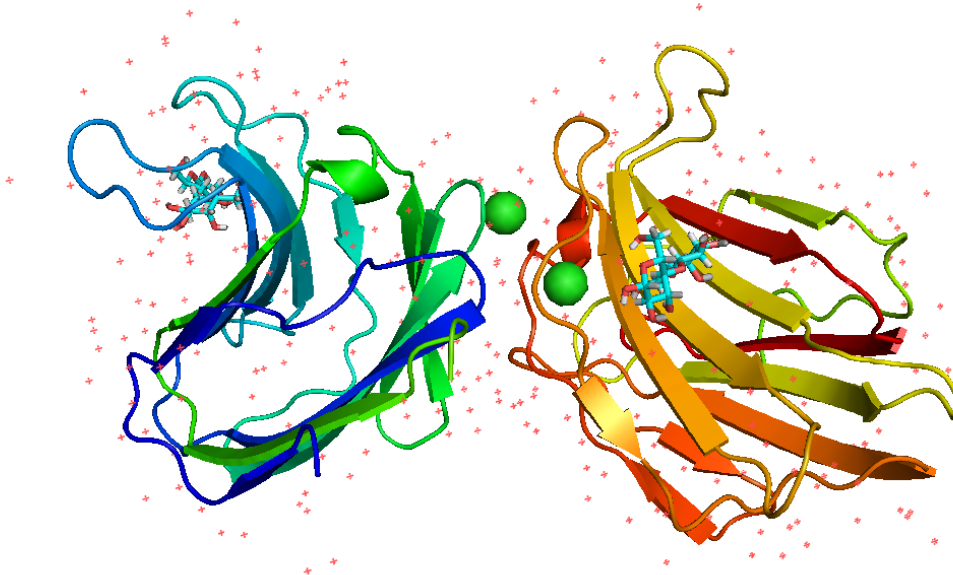


Figure 17: Model of human galectin-8N variant V with lactose in the binding sites (vA5_R1), viewed in PyMOL. Chain A is in the color of blue/cyan/green, while chain B is in red/orange/yellow.

4.1.8 Ligand Interaction in Binding Site:

The following table shows the interaction between inserted ligand and residues at the binding site (see table 18):

Table 18: Results of hydrogen bond interactions between inserted ligand and residues at the binding site. Note that interactions between ligand and water in the binding site have been excluded. The residues in parentheses are unusual/skeptical interactions.

Model	Ligand	Residues in the binding site
A4_2_3 (Chain A)	MH-3-80	Arg52, Arg66, His72, Arg76, Asn86, Glu96
A4_2_3 (Chain B)	MH-3-80	Arg52, (Gln54), Arg66, His72, Arg76, Asn86
mB4_B1 (Chain A)	MH-1FH	Arg52, Gln54, Arg66, His72, Asn74, Arg76, Asn86
mB4_B1 (Chain B)	MH-1FH	Arg52, Gln54, Arg66, His72, Arg76, Asn86
vC5_B2 (Chain A)	MH-1FH	Arg52, Gln54, Arg66, His72, Asn74, Arg76, Asn86
vC5_B2 (Chain B)	MH-1FH	Arg52, Gln54, Arg66, His72, Asn76, Asn86
vA5_R1 (Chain A)	Lactose (LAT)	Arg52, His72, Arg76, Asn86, Glu96
vA5_R1 (Chain B)	Lactose (LAT)	Arg52, His72, Arg76, Asn86, Glu96
V_2	Lactose (LAT)	Arg52, His72, (Asn74), Arg76, Asn86, Glu96

Possible to notice in the table above (table 18) the residues interacting with inserted ligands are quite similar for all proteins, if not the same. Regardless of protein variant or inserted ligand residues; Arg52,

His72, Arg76, and Asn86 will interact and form hydrogen bonds indicating that these residues are conserved in the site and are the main residues that must be concerned with the highest priority when synthesizing ligands. In addition, aside from the hydrogen bond interactions there are also hydrophobic interactions between phenyl-groups of the ligands and mainly aromatic residues such as Trp93.

Additionally, depending on the inserted ligand's structure, residues such as Arg66, Gln54 and Glu96 were found to interact. Arg66 interacts with ligands MH-3-80 and MH-1FH however not with lactose. Gln54 was found to interact only when ligand MH-1FH was inserted, while Gln96 was exclusively interacting with lactose (when only observing both chains, excluding Glu96 interactions in chain A with ligand MH-3-80).

Furthermore, there were two residues exclusive for one of the chains such as Glu96 interacting with MH-3-80 in chain A, and Asn74 interacting with MH-1FH in chain A for both variants of the human protein. Observing Ströhagen's structure of human galectin-8N [1], the interaction between Glu96 and MH-3-80 ligand is something exclusively in chain A, indicating the interaction not being serendipitous. The reason for this is due to how the protein is located in relation to other proteins, which can be observed using the "generate symmetry" function in PyMOL (see figure 18 and 19).

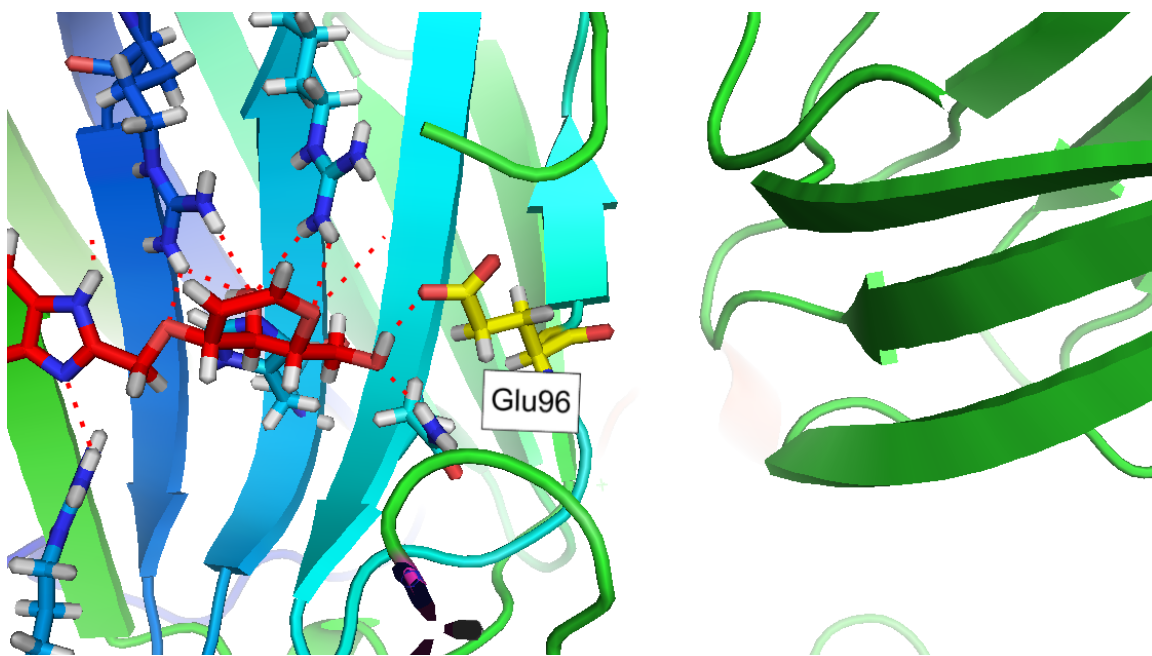


Figure 18: A picture of chain A of galectin-8N variant V when ligand MH-3-80 is inserted, taken from PyMOL. Focusing on Glu96 (in yellow) of chain A (in cyan), it is possible to observe that surrounding protein (green) to the left have an impact on affecting the interaction with the inserted ligand.

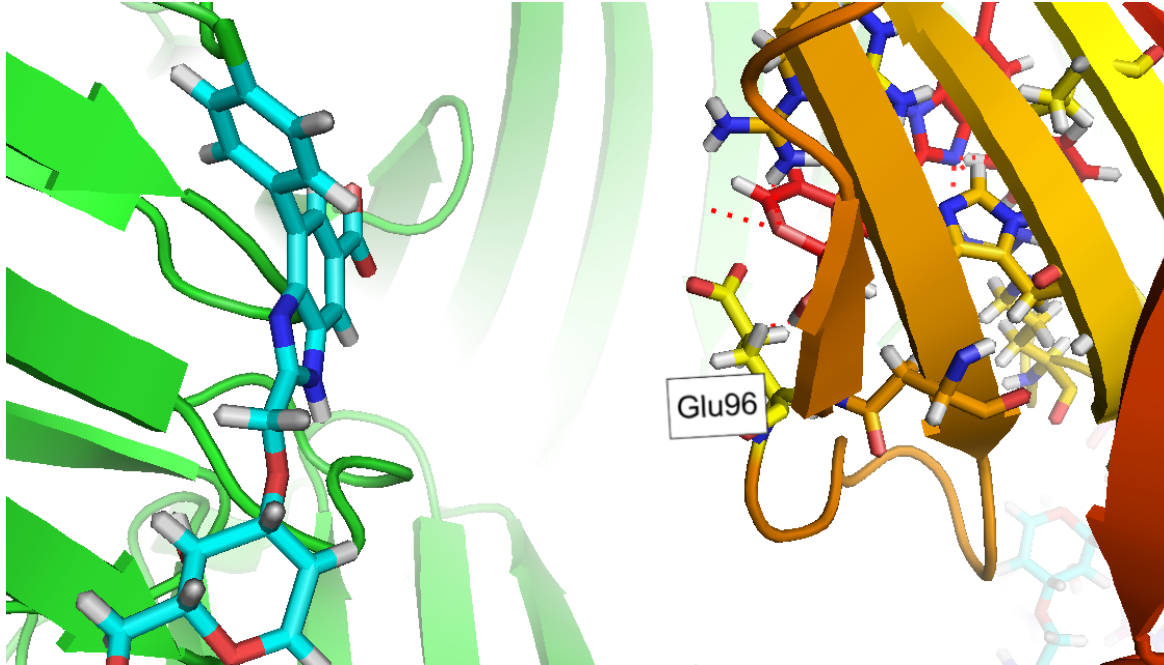


Figure 19: A picture of chain B of galectin-8N variant V when ligand MH-3-80 is inserted, taken from PyMOL. Focusing on Glu96 (in yellow) on chain B (in orange), it is possible to observe that the surrounding protein does not have an impact on affecting the interaction with the inserted ligand.

Possible to observe the glutamine 96 residue in chain A is located at the edge of the protein which makes it possible for the surrounding protein to “push” it to interact with the inserted ligand (see figure 18). On the other hand, when inspecting chain B the glutamine is facing the binding site of other surrounding proteins indicating no force is pushing the residue to interact with the ligand at the binding site (see figure 19).

For interaction between the ligand MH-1FH and asparagine 74 (in model mB4_B1) the interaction exclusively for chain A is due to conformation difference of the hydroxyl-group. When observing the available models, the hydroxyl-group has a difference in 180 degrees between chain A and B.

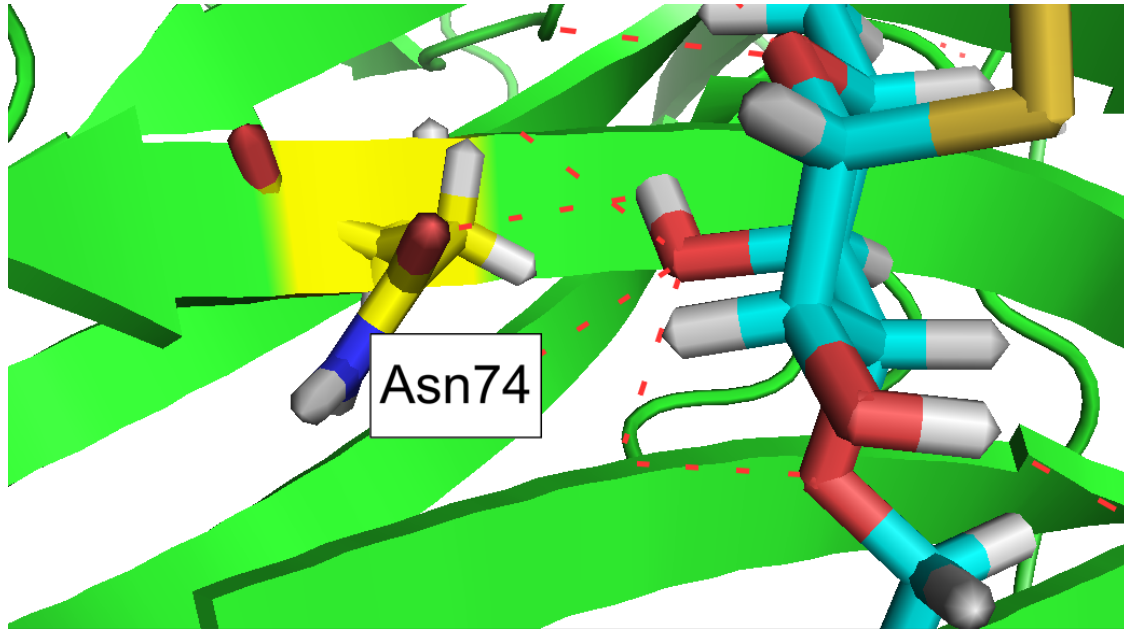


Figure 20: A picture of chain A of galectin-8N variant M when ligand MH-1FH is inserted, viewed in PyMOL. Focusing on Asn74 (in yellow) an interaction is possible to observe with the hydroxyl group of the ligand.

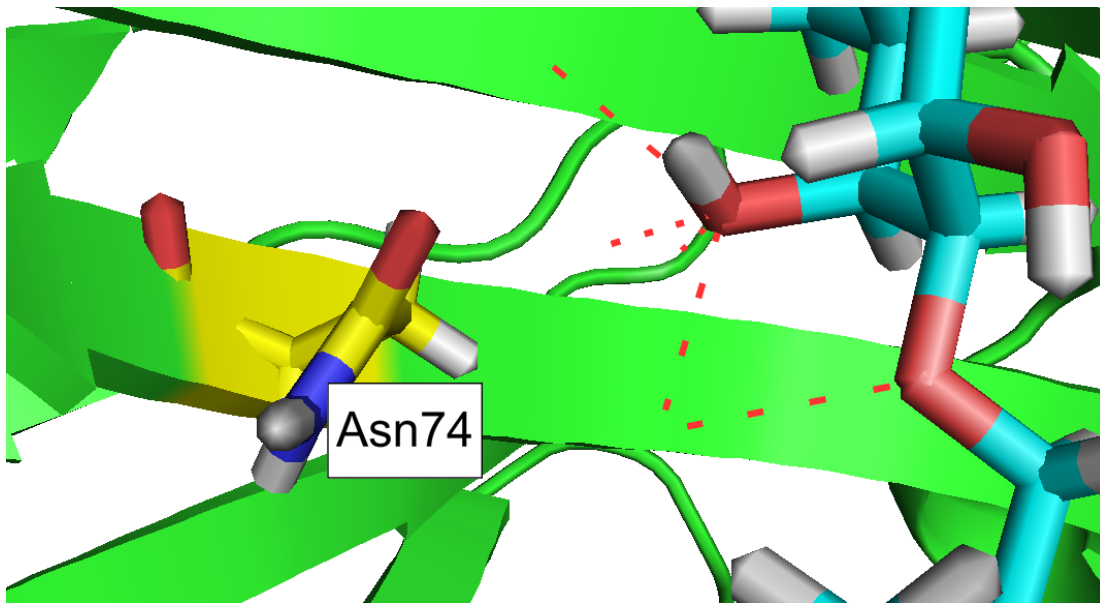


Figure 21: A picture of chain B of galectin-8N variant M when ligand MH-1FH is inserted, viewed in PyMOL. Focusing on Asn74 (in yellow) there is no interaction with the hydroxyl group of the ligand.

Whether this conformation of the hydroxyl group is a calculation error by phenix or not, the conformation difference between chain A and B can be observed in both variant M and V of galectin-8N. In order to find out whether this interaction exists or not, additional experiments have to be conducted with neutron crystallography to observe the correct conformation of hydrogen atoms.

There are two unusual interactions between ligands and residues such as Gln54 in model A4_2_3 or Asn74 in model V_2 exclusive to one chain in those specific models. When comparing the interaction of ligand MH-3-80 and Gln54 between A4_2_3 and Ströhagen's model of human galectin-8N variant M [1], it

is possible to notice a significant difference where the nitrogen group of glutamine residue hydrogen bonds with the nitrogen group of the ligand (see figure 22 and 23):

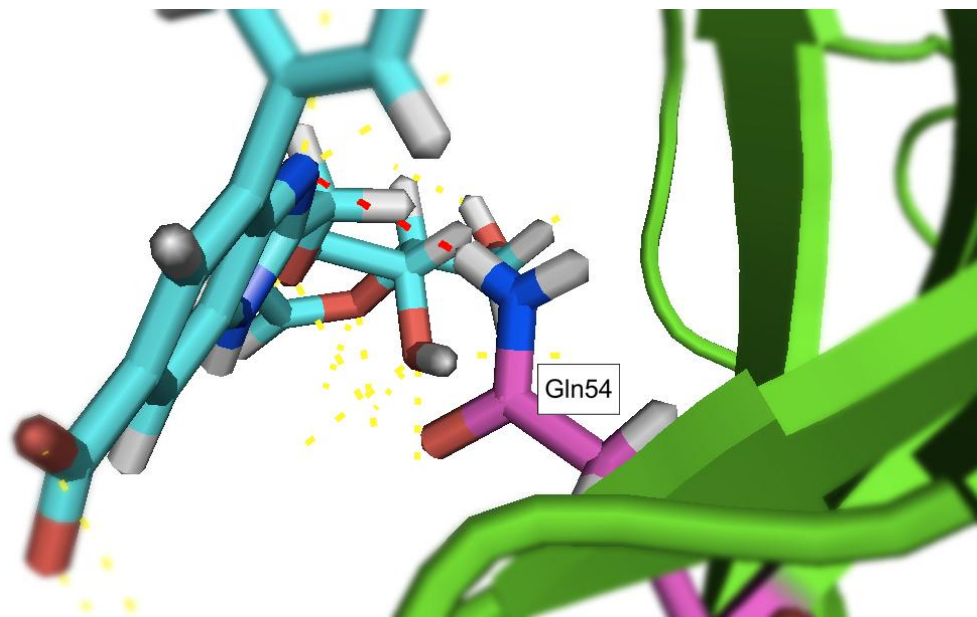


Figure 22: A picture of chain B of galectin-8N variant V (model: A4_2_3) when ligand MH-3-80 is inserted, viewed in PyMOL. Focusing on Gln54, an interaction (in red) between the residue and nitrogen group of the ligand is possible to observe.

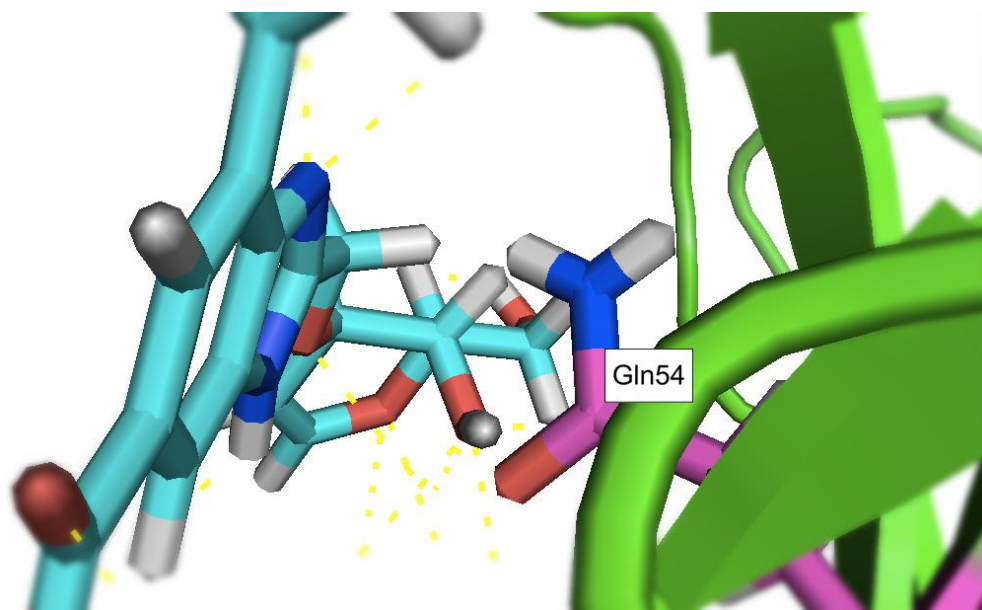


Figure 23: A picture of chain B of galectin-8N variant M (model: Ströhagen's [1]) when ligand MH-3-80 is inserted, viewed in PyMOL. Focusing on Gln54, there is no interaction between the residue and nitrogen group of the ligand unlike model A4_2_3.

Nitrogen tends to be an electron donor which means that a hydrogen bond formed by two donors is unlikely to occur. If the glutamine had an oxygen group instead of a nitrogen group such interaction would be possible due to oxygen being an electron acceptor. However, analyzing the conformation and planes

of the ligand and residues it is unlikely that interaction between two donors takes place, thus we may conclude that the interaction between MH-3-80 and Gln54 in model A4_2_3 is likely an error made by phenix.

Lastly, when comparing the interaction between lactose and Asn74 between model vA5_R1 and V_2 the conformation of the ligand itself differs, where an interaction can be found in model V_2, however not in vA5_R1 (see figure 24 and 25) .

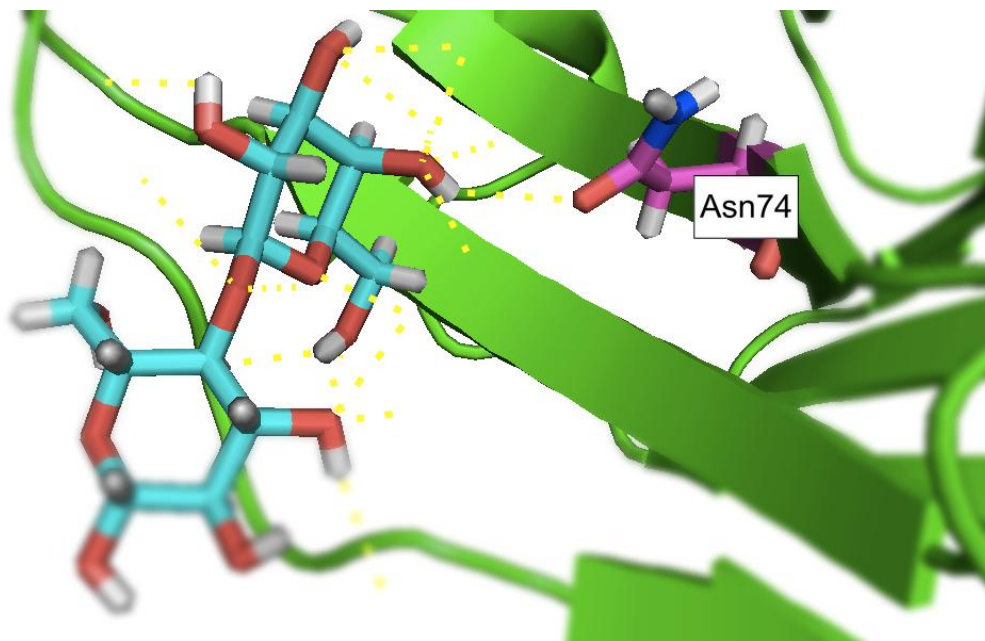


Figure 24: A picture of galectin-8N variant V (model: V_2) when lactose is inserted, viewed in PyMOL. Focusing on Asn74 an interaction between hydroxyl-groups is possible to observe.

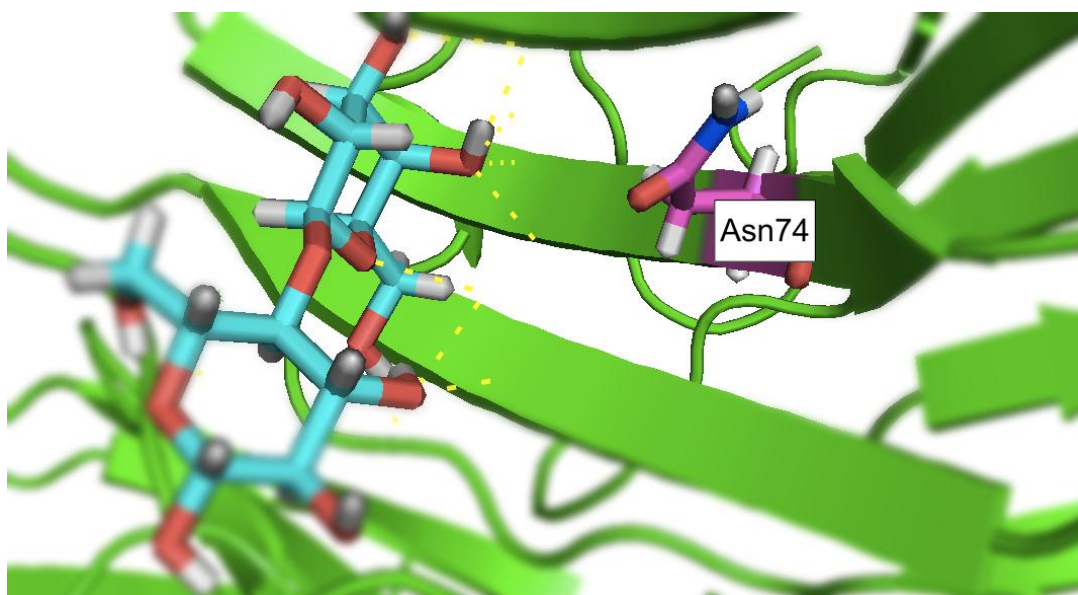


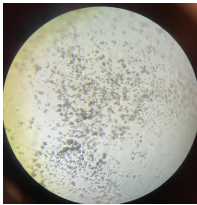
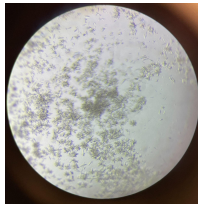
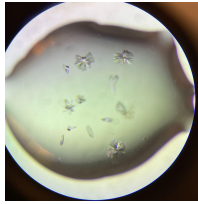
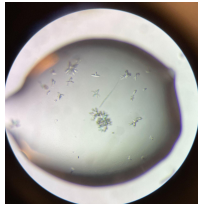
Figure 25: A picture of galectin-8N variant V (model: A5_R1) when lactose is inserted, viewed in PyMOL. Focusing on Asn74 there are no visible interactions between the two hydroxyl-groups, unlike model V_2.

The conformation of the lactose in the binding site is true and valid, but the hydrogens presented on the models are difficult to tell. In X-ray crystallography it is not possible to observe the hydrogens of residues thus phenix.refine calculates and adds hydrogen onto the model at places where it would make the most sense to place them with the most probable rotamer. The hydrogens on the model created in phenix and COOT are theoretical values which is an estimation of placement of the atoms, therefore in order to confirm whether there is an interaction or not experiments where hydrogens are visible such as Neutron Crystallography has to be carried out, just like Asn74 interaction exclusively for chain B when MH-1FH ligand is inserted mentioned above.

4.1.9 Manual Seed Beads Results:

Protein crystals of human galectin-8N variant M were observable in drops containing high concentration of precipitant and pipetted with seed stock/dilution solution. The following table shows the result of the crystals formed during the Seed Bead experiment (see table 19):

Table 19: The results of the Seed Beads obtained from the EasyXtal plate. The wells with crystals observed were photographed through a microscope.

Seed Solution	Mixture 4 (75%)	25.0%	27.5%	30.0%	32.5%
1 (Stock) & 2 (Dilution 1)	Alc Conc. 0.12M	1. Nothing 2. Nothing	1. Nothing 2. One tiny crystal	1. Nothing 2. Clutter of tiny crystals 	1. Nothing 2. Clutter of tiny crystals. 
3 (Dilution 2) & 4 (Dilution 3)		3. Nothing 4. Nothing	3. Nothing 4. Nothing	3. Nothing 4. Nothing	3. Many big spiky crystals.  4. Big spiky crystals. 

The result above proved that stroking the protein with the fiber string tool did not produce any crystals. On the other hand when pipetting the seed solution directly into the drop crystals formed in presence of high

concentration of precipitant. This may be due to the received fiber string tool being a prototype model, thus the required amount of seed solution may have not been attached to the fiber.

4.2 Mouse Galectin-8N Results:

4.2.1 nanoDSF Melting Point Analysis and Buffer Screens:

From the results obtained from the nanoDSF machine and PR.ThermControl software, the melting point of mouse galectin-8N was measured to be 56.6°C (see figure 26).

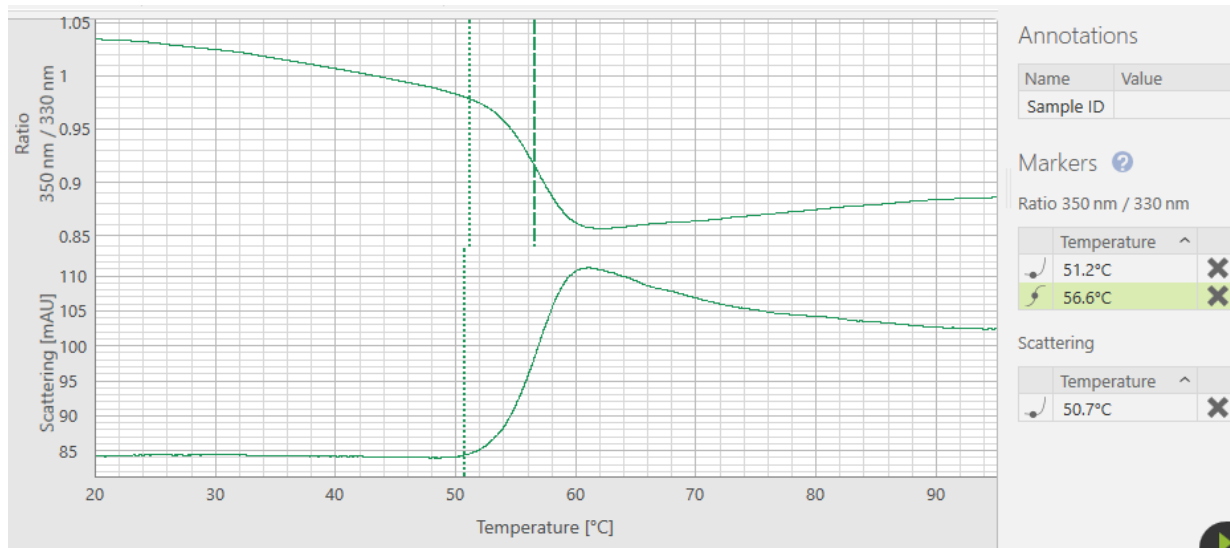


Figure 26: The graph was obtained from the software PR.ThermControl of NanoDS, showing the melting point to be 56.6°C.

For the mouse galectin-8N with MH-3-80 ligand inserted the fluorescence was immeasurable due to the ligand having similar fluorescent wavelengths as tyrosine and tryptophan.

4.2.1.2 RUBIC Screen:

The following tables shows all the melting points and the change in temperature below (see figure 12):

	1	2	3	4	5	6	7	8	9	10	11	12
A	55.3	41	38.9	52.1	52.6	56.7	58.8	55.3	55.1	58.5	58.4	56.5
B	55.9	56.1	56	58.9	54.5	57.1	56.7	56.7	56.5	56	56.6	51.6
C	57.4	44.6	32.6	48.5	51.6	56	58.3	56.1	56.1	58.7	58.5	57.1
D	57.3	56.7	57.3	58.9	55.7	58.1	57.5	58.1	57.8	57.7	/	54
E	37.2	41.3	50.3	54.8	56.7	57.5	58	57.7	57.7	57.3	57.1	55.6
F	55.8	56.7	57.5	58.5	56.5	57.5	58.9	60.4	56	56.1	56.8	56.6
G	56.5	57.3	57.7	58.3	58.5	58.5	56.6	56.7	57.5	58	58.2	58.1
H	53.1	55.3	55.1	55.7	57.6	56.1	55.4	55.9	55.2	53.6	51.9	49.5
							Melting=	56.6				All in unit: °C
A	-1.3	-15.6	-17.7	-4.5	-4	0.1	2.2	-1.3	-1.5	1.9	1.8	-0.1
B	-0.7	-0.5	-0.6	2.3	-2.1	0.5	0.1	0.1	-0.1	-0.6	0	-5
C	0.8	-12	-24	-8.1	-5	-0.6	1.7	-0.5	-0.5	2.1	1.9	0.5
D	0.7	0.1	0.7	2.3	-0.9	1.5	0.9	1.5	1.2	1.1		-2.6
E	-19.4	-15.3	-6.3	-1.8	0.1	0.9	1.4	1.1	1.1	0.7	0.5	-1
F	-0.8	0.1	0.9	1.9	-0.1	0.9	2.3	3.8	-0.6	-0.5	0.2	0
G	-0.1	0.7	1.1	1.7	1.9	1.9	0	0.1	0.9	1.4	1.6	1.5
H	-3.5	-1.3	-1.5	-0.9	1	-0.5	-1.2	-0.7	-1.4	-3	-4.7	-7.1

Figure 27: A figure of a table showing all the melting points obtained from RUBIC buffer screens, in addition the change in temperature. For well D11 the capillary tube broke during the experiment making it not possible to obtain data from.

The result obtained from RUBIC shows 13 buffers that affect the mouse galectin-8N to increase its melting point with over 1°C. Most common buffers to affect the melting point positively were: phosphate monobasic, citrate, Tris-HCl, HEPES, SPG, or BICINE, all within pH range of 6.0-8.0.

4.2.1.3 Durham Screen:

The following tables shows all the melting points and the change in temperature below (see figure 28):

Tm	1	2	3	4	5	6	7	8	9	10	11	12
A	55.1	55	43.8	42.6	42.9	42.9	38.8	44.1	49.7	34.3	45.5	51.7
B	31.4	43.9	52.4	33.3	45.8	52.6	48.9	51.6	53.1	52.4	56.4	57.7
C	56.5	58.4	60	54.3	57.3	56.6	49.1	53	55.2	55.2	55.2	56.6
D	52.5	55.1	56.5	55.6	58.5	58.4	53.4	55.1	55.8	56.6	56	56.7
E	57.1	59.2	59.7	56	57.7	57.8	54.8	55	54.7	56	55.9	56.2
F	55.4	56	57.1	56.4	57	57	56.5	56.4	56.4	50.5	56.4	55.9
G	56.5	56.3	54.8	56.6	56.5	55.1	56.6	56	56.2	51.5	53.9	52.2
H	52.7	51.9	49.9	53.1	51.6	49.2	55	55	55	55.1	46.3	44.7
												Tm = 56.6
Delta Tm	1	2	3	4	5	6	7	8	9	10	11	12
A	-1.5	-1.6	-12.8	-14	-13.7	-13.7	-17.8	-12.5	-6.9	-22.3	-11.1	-4.9
B	-25.2	-12.7	-4.2	-23.3	-10.8	-4	-7.7	-5	-3.5	-4.2	-0.2	1.1
C	-0.1	1.8	3.4	-2.3	0.7	0	-7.5	-3.6	-1.4	-1.4	-1.4	0
D	-4.1	-1.5	-0.1	-1	1.9	1.8	-3.2	-1.5	-0.8	0	-0.6	0.1
E	0.5	2.6	3.1	-0.6	1.1	1.2	-1.8	-1.6	-1.9	-0.6	-0.7	-0.4
F	-1.2	-0.6	0.5	-0.2	0.4	0.4	-0.1	-0.2	-0.2	-6.1	-0.2	-0.7
G	-0.1	-0.3	-1.8	0	-0.1	-1.5	0	-0.6	-0.4	-5.1	-2.7	-4.4
H	-3.9	-4.7	-6.7	-3.5	-5	-7.4	-1.6	-1.6	-1.6	-1.5	-10.3	-11.9

Figure 28: A figure of a table showing all the melting points obtained from Durham buffer screens, in addition the change in temperature.

The result obtained from Durham shows 7 buffers that affect the mouse galectin-8N to increase its melting point with over 1°C. Most common buffers to affect the melting point positively were: citrate, phosphate, ADA, PIPES, or malonic acid, all within pH range of 6.5-7.3.

From the results obtained from the two buffer screens it was concluded that 100 mM of citrate buffer at pH 6.5 would be the most suitable buffer for mouse galectin-8N for further crystallization attempts, due to phosphate buffer having a risk of producing salt crystals.

4.2.2 Mouse Galectin-8N Commercial Screen (Stock in PBS buffer):

The results obtained from JCSG+, PACT Premier, and Morpheus were as follows; for JCSG+ 5/96 wells produced protein lumps where the remaining 91 wells had nothing but unaffected drops or precipitate. PACT premier and Morpheus did not contain anything of interest in any of the 96 wells. The proteins were observed over a span of over 50 days.

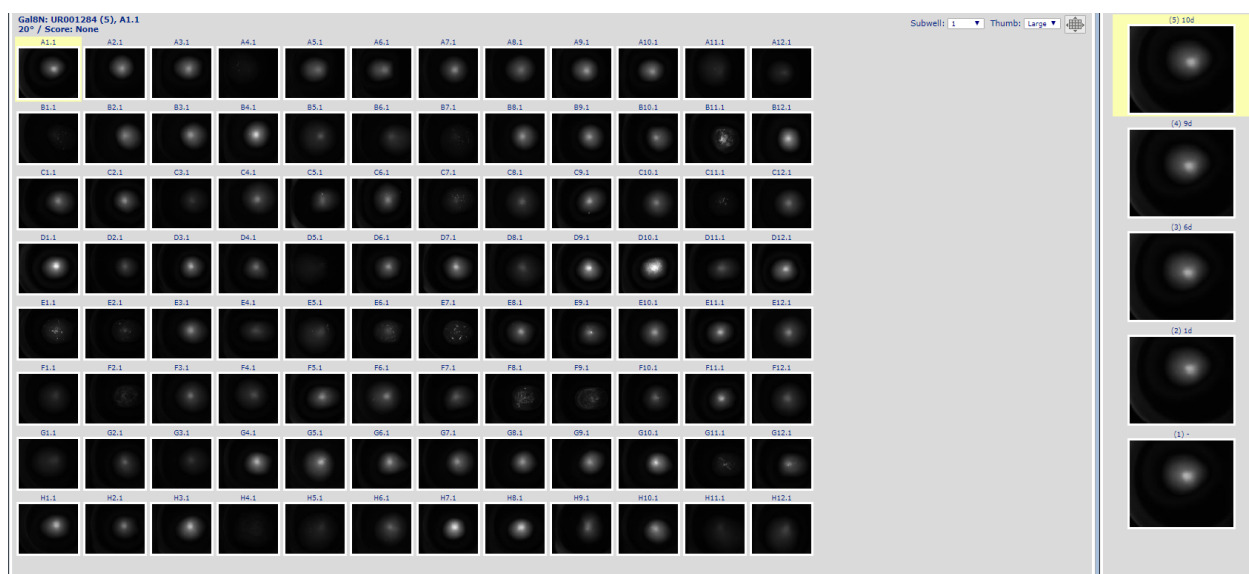


Figure 29: An example of the results of JCSG+ commercial screen, showing no signs of crystals but instead: nothing, nucleation, or precipitate. The figure was obtained from LP3.

Moreover, the ligand MH-3-80 inserted into the mouse galectin-8N also fluorescent wavelength at 330-350 nm which made the UV-light observation close to non feasible. The following figure shows an example of UV-light illuminated on to the mouse protein with ligand MH-3-80 in presence (see figure 30):

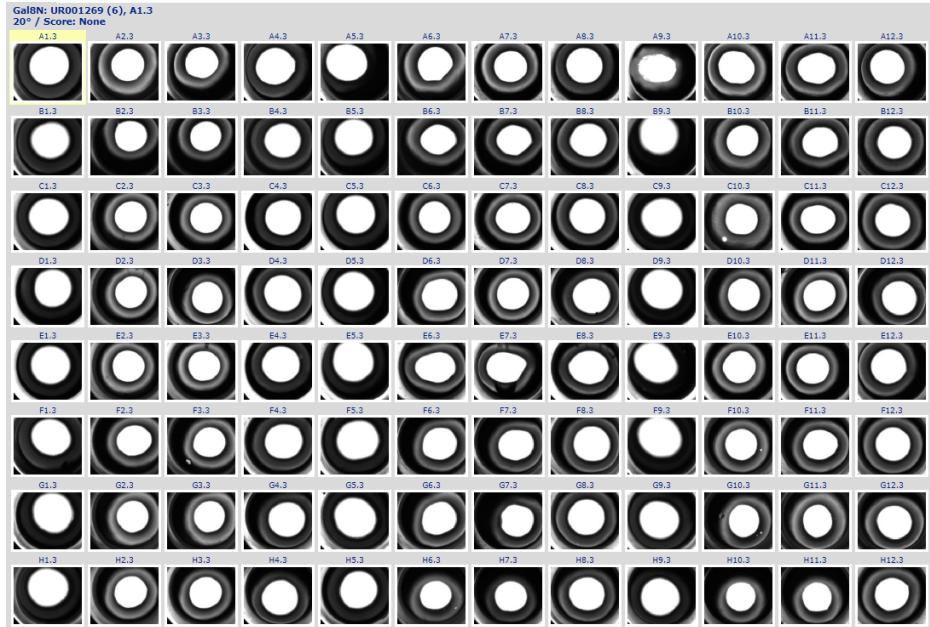


Figure 30: Example of ligand MH-3-80 added into the mouse protein drop, possible to observe that the ligand has traits of fluorescence. The figure was obtained from LP3.

Though on day 21 in well C11 subwell 3 of the JCSG+ screen big well shaped protein lumps were visible and between day 21-30 it was possible to observe presumably circular crystals (see figure 31). However, the crystals deteriorated and vanished between day 30-40.

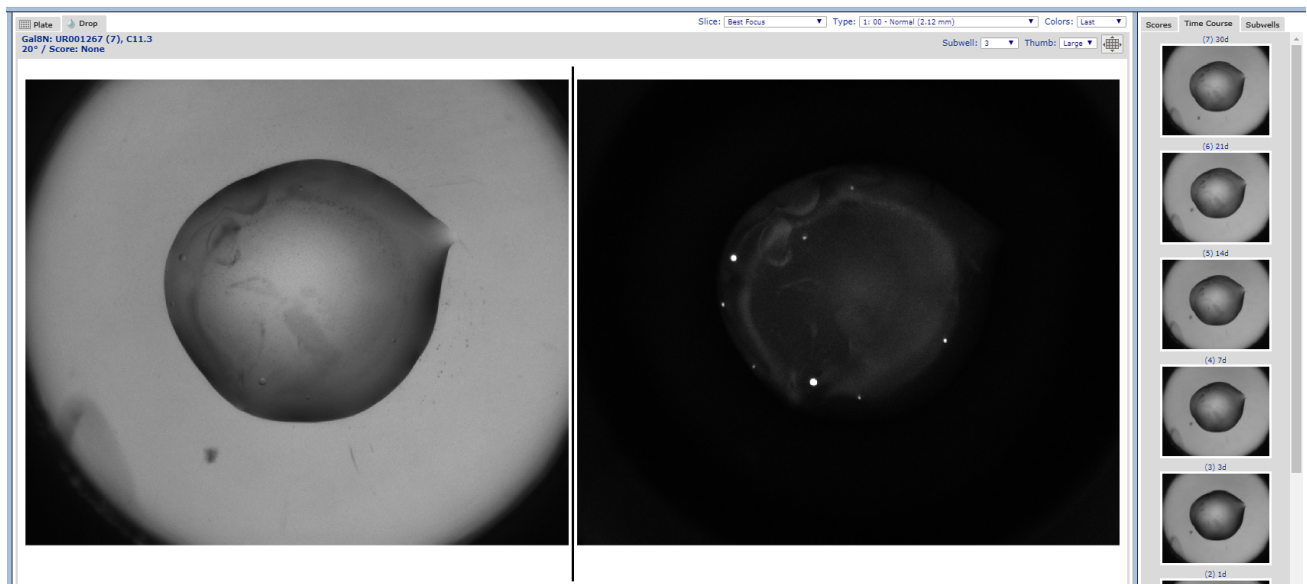


Figure 31: Well C11 subwell 3 with MH-3-80 inserted showing signs of well-shaped circular protein lumps or crystals on day 30. Picture of the protein drop to the left and UV-light illuminated to the right. The figure was obtained from LP3.

However, the plate was once again scanned at day 54 and the presumed crystals that vanished between day 30-40 were back on place, in addition the overall amount and the fluorescence-intensity of crystals had increased (see figure 32). Though, considering the “crystals” vanishing and appearing randomly and

the fluorescence-intensity increasing it is skeptical whether these round objects are crystals or just well shaped lumps of protein. The lumps/crystals still existed on day 62.

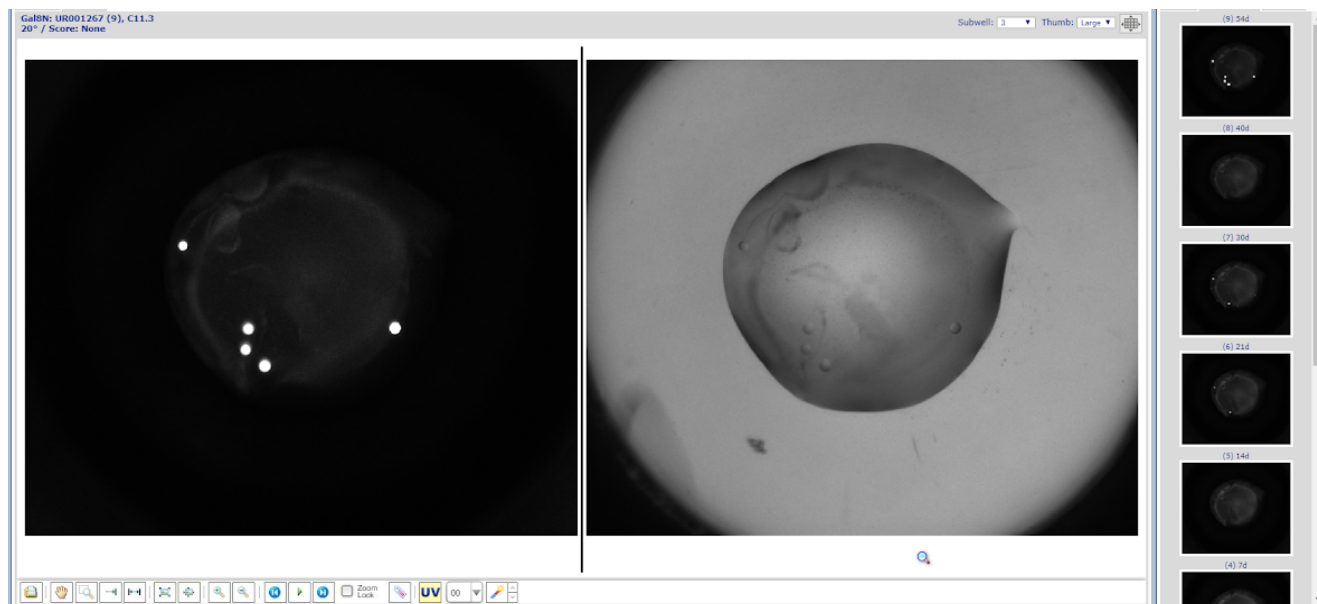


Figure 32: Well C11 subwell 3 with MH-3-80 inserted showing signs of well-shaped circular protein lumps or crystals on day 54. Picture of the protein drop to the right and UV-light illuminated to the left. The figure was obtained from LP3.

4.2.3 Mouse galectin-8N Commercial Screen (Stock in Citrate buffer):

Results showed that no visible crystals were observed. Instead all five commercial screens had in common to produce more light or heavy precipitate and sometimes even protein lumps. BCS produced light precipitate in about 76% of the wells and protein lumps were observable in two wells, while ProPlex had about 44% of the wells containing light precipitate. Morpheus had precipitate in about 78%, JCSG+ had light/heavy precipitate in about 66% of the wells while three wells contained protein lumps, and lastly PACT premier had precipitate in about 98% of the wells.

Not to mention, the Morpheus screen either had none or really weak precipitate in wells containing precipitant mixture 1 (aka column 1, 5, 9), which means that mouse galectin-8N does not find a surrounding containing mixture 1 (see table 3 of mixtures in methods) optimal.

4.2.4 Seed Bead with Commercial Screen Results:

After observing the Morpheus screen together with seed beads the results were that; nothing interesting happened. In all 96 wells nothing special occurred. However, it was noticed that nucleation of the protein occurred in most of the wells but column 1, 5, and 9, which are the wells containing precipitant mixture 1.

5.0 Discussion:

To crystallize human galectin-8N variant M and V the optimal conditions with high success rate is done by utilizing Morpheus commercial screen or imitating the conditions of Morpheus in the manual optimization screens. The protein generally has a high preference for buffer 1 with pH 6.5 and precipitant mixture 4. When it comes to preference for additives the two variants they both prefer in conditions of either ethylene

glycols and monosaccharides, then alcohols for second place. In presence of alcohol mixture in the reservoir solution variant M does not have problem crystallizing since all 15 wells in the manual screen had crystals (see table 6). However, for variant V in low concentration of precipitant and alcohols has a troublesome time to crystallize, so in conclusion for increased reproducibility with well-shaped crystals use; precipitant mixture 4 at concentration in between 35-42.5% and 0.12-0.16 M of alcohol mixture, together with buffer 1.

When it comes to the use of Seed Beads kit it will produce similar shaped spiky crystals or several tiny crystals depending on the seed stock concentration, and is a good tool to experiment with for increased reproducibility of human galectin-8N. In the commercial and manual screens the shape and size of the crystals were uncontrollable resulting in being long and big, which makes Seed Beads an excellent strategy if you require numerous crystal samples with similar shapes. Just keep in mind that pipetting the seed stock solution directly into the protein drop will function better than stroking through the drop with a fiber string.

For the interactions observed between inserted ligands and the residues at the binding site of galectin-8N variant M and V several interactions were observable and noted down. All three ligands have interactions with mainly the residues: Arg52, His72, Arg76, Asn86, and Trp93. Some interactions exist depending on the inserted ligand where MH-3-80 and MH-1FH interact with Arg66, MH-1FH with Gln54, and lastly lactose's interaction with Glu96. When synthesizing ligands to insert to the binding site of the protein the first four residues are the most important ones to be concerned about. Then there are some one chain exclusive interactions between ligands and residues however in order to observe whether there are interactions or not, further experiments have to be conducted where hydrogen atoms are visible.

The reason for why ligand NV9.1 did not bind to the binding site of the human proteins were due to the low dissociation constant (K_d). NV9.1 has the lowest binding affinity with only 290 μM , thus the concentration was increased from 50 mM to 100 mM during cocrystallization. However, lactose has an affinity of 90 μM for human galectin-8N which means that NV9.1 has to have a higher concentration in order to bind. This explains why ligands MH-3-80 and MH-1FH were able to bind due to their low K_d -value, since lactose was already in the binding site of all the received proteins. To bind ligand NV9.1 to human galectin-8N, either the concentration of the ligand has to be further increased to around 200 mM which builds the question whether the ligand dissolves in such low volume of DMSO, or prepare apoproteins with no ligand in the binding site.

In the end, no crystals were producible for mouse galectin-8N (unless well C11 in JCSG+ contains crystals), which explains why no prior crystallization data existed during the research about the protein. The closest results obtained were in well C11 of JCSG+ screen when MH-3-80 ligand was inserted and the protein was in PBS buffer with pH 7.5, where well C11 contains: ammonium sulfate, sodium-acetate at pH 4.6. What made it difficult to conclude the most successful result in all eight mouse commercial screens is that all parameters affecting the crystallization were not consistent and no patterns were discoverable. Most of the successful results that produced protein lumps often had a pH in between 6.5-7.5, however the lowest and highest pH to form protein lumps were 4.0 and 10.8. The salts, additives, and precipitant mix in most commercial screens had no consistency except for Morpheus.

Something noticed on Morpheus screen during the crystallization of the mouse protein was that column 1, 5, and 9 had no results in every experiment regardless of the buffer, indicating that the mouse protein has a low to no preference for precipitant mixture 1. Though, the remaining three precipitant mixtures (2, 3 and 4) had similar results of nucleation or precipitation. In addition, low preference was also observable for additives such as: divalents, halogens, and NPS. Most of the protein lumps and nucleation effects

were noticed to occur in the JCSG+ screen in both PBS and citrate buffer, thus it is possible to conclude that for further experimentations of crystallization the mouse protein it is highly recommended to test the conditions of JCSG+ commercial screen in a higher protein concentration than 2.6 or 2.2 mg/ml.

This project can be further expanded to discover even more efficient strategies to increase the crystallization and reproducibility of human galectin-8N. The methods and amounts described during this project are sufficient and enough to produce good crystals, but improvement of methods can always be explored. The models and insertion of sample ligands into the binding site can also be further explored by experimenting with new ligands with different side branches to find other new interactions, which can be useful for medical purposes in the future. In addition, to increase the accuracy and precision of the models techniques where hydrogen atom rotamers are visible such as neutron crystallography would be beneficial to explore with. Lastly, new crystallization attempts to crystallize mouse galectin-8N can be further explored. Best commercial screens to explore the trial-and-errors of crystallization would be JCSG+, however try to avoid inserting ligand MH-3-80 due to its fluorescent natures making it difficult to observe in UV-light. Manual screens with similar conditions imitating well C11 in JCSG+ (protein in PBS buffer with MH-3-80 ligand) can also be experimented with, to then get the protein modeled and be published in data banks for further observations and research.

6.0 References:

- [1] Ströhagen N. “*Crystal structure of lactose and a novel ligand in complex with galectin-8N*”, Bachelor thesis, Department of Biochemistry and Structural Biochemistry, Lund University, Sweden, 2021.
- [2] Johannes L, Jacob R, Leffler H, “*Galectins at a glance*”, Journal of Cell Science, Volume 131, Issue 9, 2018.
- [3] Chang W, Tsai M, Kuo P, Hung J, “*Role of galectins in lung cancer*”, Oncology Letters, Nov; 14(5): 5077–5084, 2017
- [4] Hadari Y.R, Arbel-Goren R, Levy Y, Amsterdam A, Alon R, Zakut R, Zick Y, “*Galectin-8 binding to integrins inhibits cell adhesion and induces apoptosis*”, Journal of Cell Science, Volume 113, Issue 13, 2000
- [5] Lund Protein Production Platform, “*Production and purification of human Gal8N aa4-158 M/V-version / Mouse Gal8N 3-157, batch 1, performed by LP3 September 2021*”, Protocol, received 25-03-2022
- [6] Malvern Panalytical, “*Binding Affinity*”, link: <https://www.malvernpanalytical.com/en/products/measurement-type/binding-affinity>, taken 2022
- [7] Phenix (Python-based Hierarchical ENvironment for Integrated Xtallography), “*Structure refinement in Phenix, phenix.refine - general purpose crystallographic structure refinement program*”, link <https://phenix-online.org/documentation/reference/refinement.html>, taken 2022
- [8] Hampton Research, “*Crystal Growth Techniques*”, link: <https://hamptonresearch.com>, 2001
- [9] Molecular Dimensions, “*JCSG Plus*”, “*PACT Premier*”, “*Morpheus*”, “*The BCS Screen*”, “*ProPlex*”, from link: <https://www.moleculardimensions.com/>, taken 2022.

[10] Harvard Medical School, “*Differential Scanning Fluorimetry (DSF)*”, link: [‘https://cmi.hms.harvard.edu/differential-scanning-fluorimetry’](https://cmi.hms.harvard.edu/differential-scanning-fluorimetry) , Center for Macromolecular Interactions, taken 2022.

[11] Thermo Fisher Scientific, “*NanoDrop Product Guide*”, link: [‘https://www.thermofisher.com/se/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/ultraviolet-visible-visible-spectrophotometry-uv-vis-vis/uv-vis-vis-instruments/nanodrop-microvolume-spectrophotometers/nanodrop-products-guide.html’](https://www.thermofisher.com/se/en/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/ultraviolet-visible-visible-spectrophotometry-uv-vis-vis/uv-vis-vis-instruments/nanodrop-microvolume-spectrophotometers/nanodrop-products-guide.html) , taken 2022.

[12] Hampton Research, “*Seed Bead™ Kits*”, link: [‘https://hamptonresearch.com/product-Seed-Bead-Kits-42.html’](https://hamptonresearch.com/product-Seed-Bead-Kits-42.html) , taken 2022.

7.0 Appendix:

7.1 Method:

3.3 Screen Optimization (Hanging Drop):

Appendix Table 1: Set up for the manual screen of human galectin-8N with precipitant mixture 1 and alcohol mixture as additives. The volume of each content is noted down and each reservoir well has about 1 ml solution.

Mixture 1 (60%)	26%	28%	30%	32%	34%	Contents
0.08M	100µl 433µl 67µl 400µl 2µl	100µl 467µl 67µl 367µl 2µl	100µl 500µl 67µl 333µl 2µl	100µl 533µl 67µl 300µl 2µl	100µl 567µl 67µl 267µl 2µl	Buffer Precipitant Additive Water 0.5M TCEP
0.12M	100µl 433µl 100µl 367µl 2µl	100µl 467µl 100µl 333µl 2µl	100µl 500µl 100µl 300µl 2µl	100µl 533µl 100µl 267µl 2µl	100µl 567µl 100µl 233µl 2µl	
0.16M	100µl 433µl 133µl 333µl 2µl	100µl 467µl 133µl 300µl 2µl	100µl 500µl 133µl 267µl 2µl	100µl 533µl 133µl 233µl 2µl	100µl 567µl 133µl 200µl 2µl	

Appendix Table 2: Set up for the manual screen of human galectin-8N with precipitant mixture 4 and alcohol mixture as additives. The volume of each content is noted down and each reservoir well has about 1 ml solution.

Mixture 4 (75%)	32.5%	35%	37.5%	40%	42.5%	Contents
0.08M	100µl 433µl 67µl 400µl 2µl	100µl 467µl 67µl 367µl 2µl	100µl 500µl 67µl 333µl 2µl	100µl 533µl 67µl 300µl 2µl	100µl 567µl 67µl 267µl 2µl	Buffer Precipitant Additive Water 0.5M TCEP
0.12M	100µl 433µl 100µl 367µl 2µl	100µl 467µl 100µl 333µl 2µl	100µl 500µl 100µl 300µl 2µl	100µl 533µl 100µl 267µl 2µl	100µl 567µl 100µl 233µl 2µl	
0.16M	100µl 433µl 133µl 333µl 2µl	100µl 467µl 133µl 300µl 2µl	100µl 500µl 133µl 267µl 2µl	100µl 533µl 133µl 233µl 2µl	100µl 567µl 133µl 200µl 2µl	

3.7 Seed Beads (human galectin-8N variant M):

Appendix Table 3: Set up for the manual screen for Seed Beads of human galectin-8N variant M with precipitant mixture 4 and alcohol mixture as additives. The volume of each content is noted down and each reservoir well has about 1 ml solution.

Seed Solution	Mixture 4 (75%)	25.0%	27.5%	30.0%	32.5%	Contents
1 (Stock) & 2 (Dilution 1)	Alc Conc. 0.12M	100µl 357µl 100µl 443µl 2µl	100µl 393µl 100µl 407µl 2µl	100µl 429µl 100µl 371µl 2µl	100µl 464µl 100µl 336µl 2µl	Buffer Precipitant Additive Water 0.5M TCEP
3 (Dilution 2) & 4 (Dilution 3)		100µl 357µl 100µl 443µl 2µl	100µl 393µl 100µl 407µl 2µl	100µl 429µl 100µl 371µl 2µl	100µl 464µl 100µl 336µl 2µl	