Degree Project in Applied Biochemistry
KBKM05

Structure-based design of new antiviral substances against the Venezuelan equine encephalitis virus

A collaboration between
Lund University and the Swedish Defence Research Agency (FOI)

Spring 2022

LUND
UNIVERSITY

Supervisor at LTH:
Cedric Dicko

Author:
Elin Öberg

Supervisor at FOI:
Nina Forsgren

Examiner:
Johan Svensson Bonde

Abstract

The Venezuelan equine encephalitis virus (VEEV) is a highly infectious alphavirus responsible for severe outbreaks and epidemics in the Americas. The nonstructural protein 2 (nsP2) cysteine protease plays a fundamental role in the VEEV replication process, and is therefore, an attractive drug target for structure-based drug design using x-ray crystallography.

Generating protein crystals of sufficient quality for x-ray crystallography is challenging and time-consuming. This thesis aims to develop methods to obtain such crystals by optimizing crystallization conditions. Moreover, this thesis also investigates interactions between potential inhibitors and the nsP2 protease experimentally by conducting biochemical assays.

By screening numerous crystallization conditions with variations in pH, additives, precipitants and protein concentrations, a handful of conditions were found to generate crystals of high quality. Their quality was later confirmed by x-ray crystallography experiments. Furthermore, the overall collected data allowed structures with resolutions below 2 Å to be obtained. Some of these collected data sets are believed to comprise the nsP2 protease in complex with an inhibitor.

Biochemical assays provided information about steady-state kinetic parameters of the nsP2 and the potency of three potential inhibitors. Three-dimensional structures of the nsP2-inhibitor-complex together with information about interactions between potential inhibitors and the nsP2, will provide information to further design inhibitors based on structure.

Acknowledgements

This thesis was conducted during the spring semester of 2022 at the Swedish Defense Research Agency (FOI) in Umeå in collaboration with the Division of Pure and Applied Biochemistry, Department of Chemistry, Faculty of Engineering, Lund University in Lund.

I would like to begin to thank my primary supervisor at FOI Nina Forsgren, for her patience and help from day one. Nina, your constant support and encouragement has played an important role throughout this thesis, and I am grateful for everything I have learned from you. A special thanks to Fredrik Ekström, for contributing with your knowledge and for believing in me. Thanks to Daniel Wiktelius for your contribution regarding the probes and the chemical part of this thesis. My time at FOI would not have been the same without you three.

Finally, I am grateful to Cedric Dicko, my supervisor at Lund University, for helping me with all the administrative tasks regarding my master thesis, and for being easily accessible and helpful per mail.

Table Of Contents

Abbreviations

CFP	Cyan fluorescent protein
CV	Column volume
FRET	Fluorescence resonance energy transfer
IMAC	Immobilized metal affinity chromatography
LB	Lysogeny broth
LFS	Ligand friendly screen
MR	Molecular replacement
nsP	Nonstructural protein
PEG	Polyethylene glycol
SG1	Shot gun screen
TEV	Tobacco etch virus
T_m	Melting temperature
Trx	Thioredoxin
TSA	Thermal shift assay
VEEV	Venezuelan equine encephalitis virus
YFP	Yellow fluorescent protein

List of Figures

List of Tables

Chapter 1

Introduction

Viral diseases have been a threat to human health for thousands of years [33] and with traces of the Covid-19 pandemic still lingering after two years, the severeness of new and reemerging viral infections and their impact on human health and economic development, should not be understated [34]. Available treatments, such as vaccines and antivirals are essential tools to limit the spread of viral outbreaks.

Alphaviruses are small, positive, single-stranded RNA viruses that have caused severe infectious outbreaks in many parts of the world [48, 1, 37]. They can infect several types of host animals, including mosquitoes, which is one of the most common routes of transmitting the virus to humans [30]. The symptoms vary depending on what type of alphavirus is causing the infection. Sindbis virus causes fever, itching, and joint pain whereas the Venezuelan equine encephalitis virus (VEEV) causes influenza-type symptoms with fever, vomiting, myalgia, and in severe cases, fatal encephalitis [7]. There are no available antiviral treatments or vaccines against VEEV, however, there are several vaccine candidates at various stages of development [1, 42].

The RNA strand in alphaviruses has two open reading frames coding for two polyproteins. The first polyprotein is essential for the assembly of the viral particle, the other polyprotein consists of non-structural proteins (nsP), which are necessary for the transcription and translation of viral mRNA inside the cytoplasm of the host cell [1, 37]. The VEEV nsP2 protease is believed to have an essential role in RNA replication and propagation due to its proteolytic function [1, 25]. The nsP2 protein contains a cysteine-protease that is responsible for cleaving the polyprotein to mature proteins, which is necessary for the replicating machinery to become activated. Because the nsP2 plays a vital role in the virus's life cycle, it is an attractive target for inhibition and potential drugs. Today, only a few inhibitors inhibit both the protease activity of nsP2 and the replication of VEEV, however, with marginal effect [25, 11, 51].

X-ray crystallography is a powerful technique used for structure determination of biological macromolecules, and is thus, an important tool for structure-based drug design. The nsP2 protease is a drug target for developing antivirals against VEEV. To obtain information about its structure, protein crystals of high quality is needed. During x-ray crystallography, a crystal diffracts x-rays which provides information about the placement of the atoms in the crystal, and the three-dimensional structure can be determined. Obtaining high quality protein crystals suitable for x-ray crystallography can be challenging and time-consuming [21].

1.1 Aim

The three-dimensional structure of the VEEV nsP2 protease is known, but only one crystal structure of the protease in a complex with a bound inhibitor is published [6]. More structures of the enzyme-inhibitor-complex are needed to study how different inhibitors interact with the protease's active site. Together with information on how the inhibitors affect the protease's cleavage ability, the three-dimensional structure will provide information to further design inhibitors based on structure.

FOI has expressed, crystallized, and obtained high-resolution three dimensional structures of the VEEV nsP2 protease, but protein crystals of higher quality are needed to perform studies of the inhibitor-protease complex. This thesis aims to develop and refine further methods to study interactions between potential inhibitors and the nsP2 protease experimentally. Identifying what parameters (e.g. pH, precipitant concentration, and protein concentration) affect protein crystallization, and how these can be optimized to improve crystallization conditions with higher reproducibility.

Chapter 2

Background

Alphaviruses belong to the *Togaviridae* family and continuously circulate on all continents with the help of mosquitoes or other vertebrate hosts, where they can cause local outbreaks as well as human epidemics [26]. They are divided into two subgroups, old world and new world alphaviruses. The old world alphaviruses (for example the Sindbis virus) are endemic to Aisa, Europe, Australia and Africa, and commonly cause febrile illness. The new world alphaviruses, including the VEEV, are responsible for outbreaks in North, South and Central Americas, and are mostly associated with neurological disease [29, 1, 26]. VEEV infections in humans start with flu-like symptoms including fever, fatigue, headache, vomiting etc. In 4 - 14 % of cases, the short febrile illness progress into encephalitis with severe neurological complications, with approximately 1 % of cases being fatal [29, 42]. VEEV is easily aerosolized and was developed as a bioweapon by the USA and the Soviet Union during the cold war [15, 47]. The risk of large-scale bioterrorism using the VEEV is small, however, preparations ought to be made for a potential impact on public health [9]. As of today, there are no available antiviral treatments or vaccines against VEEV, however, there are several vaccine candidates at various stages of development [1, 42].

2.1 VEEV replication

VEEV is a single-stranded positive-sense RNA genome, containing two genes. The first encodes the viral replication machinery consisting of nonstructural proteins (nsPs). The second encodes the structural proteins that form the virus particle (capsid) [43, 1]. When the virus enters a host cell, the genomic RNA replicates in the cytoplasm and the nsPs are translated as polyproteins, either as P123 or P1234. The nsP2 protease (embedded in the polyprotein), cleaves the nsPs into their individual nsP1, nsP2, nsP3 and nsP4 proteins, see Figure 2.1.

The negative RNA strand is transcribed into a positive RNA strand, which then enters a virus replication complex. The nsPs aid the replication process and loading of newly synthesized positive RNA strands into capsids. Mature virus particles are then released from the cell ready to infect new ones [42].

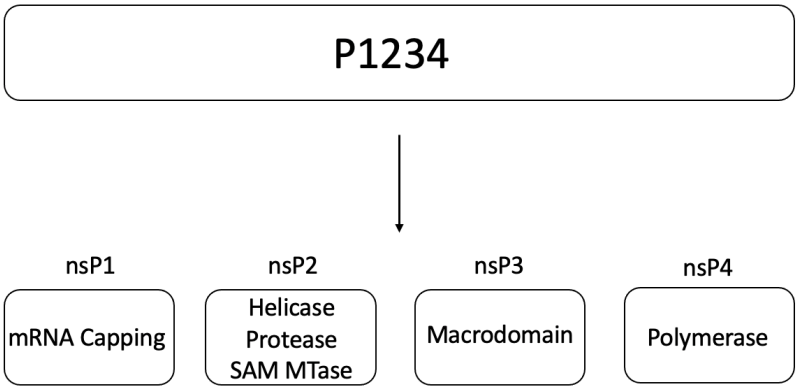


Figure 2.1: A schematic illustration of the nsPs processing by the nsP2 protease. Inspired by [1].

2.1.1 nsP2 - cysteine protease

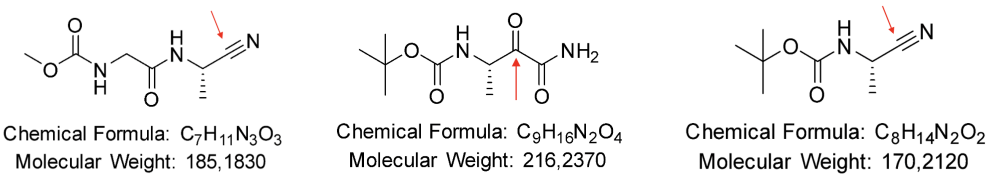
The nsP2 cysteine protease plays a fundamental role in the VEEV replication process and has multiple important functions and enzymatic activities. It cleaves the polyprotein into individual nsPs, it functions as a helicase to unwind RNA during replication and it is also believed to function as a nucleotide triphosphatase, fueling the RNA helicase activity [37, 1, 42].

The nsP2 is a cysteine protease, and cleaves its substrate at defined recognition sequences which makes it an attractive target for drug design. The active site is located between two domains, the cysteine protease domain and the SAM MTase domain. Both domains are involved in substrate recognition but the residues of the catalytic dyad, Cys477 and His546, are located in the protease domain. It is believed that when the enzyme is in its activated (ionized) state, the Cys and His exist as a thiolate-imidazolium ion pair. The deprotonated thiol then performs a nucleophilic attack on the carbonyl carbon on the substrate, and the catalytic cycle is initiated [38, 25, 39].

FOI has designed and synthesized three ligands (they will henceforth be referred to as probes), with the aim that they will bind to the nsP2 protease's active site, and subsequently, inhibit its proteolytic ability. The structure of the probes can be seen in Figure 2.2. The idea behind the design of the probes is that the catalytic sulfur atom forms a covalent bond with an electrophilic carbon on the probes. The red arrows in Figure 2.2 indicate what carbon atom is involved in the formation of a covalent bond.

2.2 X-ray crystallography

Structure-based drug design is a process where the first step is identifying a potential drug target, a protein. Structural information about the protein, and the protein bound to a compound, allows chemists to modify and improve the interactions between the compound and the protein, and thereby increasing its potency [31]. The



(a) DW0406 (b) DW0408 (c) DW0411

Figure 2.2: Chemical structures of the three probes, drawn by Daniel Wiktelius at FOI. The red arrows indicate what carbon atom is involved in the formation of a covalent bond.

drug target in this thesis is the VEEV nsP2 protease.

X-ray crystallography is a commonly used technique for structure determination of proteins and other macromolecules. The overall pipeline of x-ray crystallography is; identifying a drug target, protein expression and purification, protein crystallization screening and optimization (generating high-quality crystals), data collection and structure determination and lastly, drug development[12].

2.2.1 Protein purification

Yielding a high concentration of pure, homogeneous and soluble protein is a cornerstone in protein crystallography. Protein solutions with impurities may not be able to form crystals, or defects are incorporated into the crystal lattice during growth resulting in poor crystal quality [16]. Thus, adequate purification protocols and methods are required once the protein has been expressed [12, 44].

2.2.1.1 Immobilized metal affinity chromatography

Affinity chromatography is a purification method, where selective binding between two molecules is used to separate a target molecule, or a group of molecules, from a complex mixture. The selective binding between the two molecules is typically reversible, and purification is achieved where one of the molecules is immobilized on a surface, while the mixture containing the target molecule is in mobile phase [45]. One common strategy in immobilized metal affinity chromatography (IMAC) is to link a His-tag, a string of histidine residues, to the target molecule. The His-tag has a high affinity to metal ions such as Ni^{2+} and Co^{2+}. By immobilizing one of these metal ions on a resin matrix in a column, and allowing the mixture to flow through, the His-tag (bound to the target molecule) will bind to the column while the unwanted mixture flows through. The next step involves an elution procedure where a solution containing imidazole flows through the column. Imidazole competes with the His-tag for binding sites on the column to the metal ions, and at a sufficiently high concentration of imidazole, the His-tag releases from the column and the target molecule can be collected [27].

2.2.1.2 Dialysis

The collected target molecule usually exists in a solution containing a high concentration of imidazole after an IMAC procedure. To reduce the imidazole concentration, a dialysis procedure is conducted. The target molecule in solution is poured into a dialysis tube, a semi-permeable membrane of defined porosity, allowing only small particles to cross over the membrane. The tube is sealed and placed in a buffer solution or water, and through diffusion, the concentration of imidazole in the tube will equilibrate over time. Finally, the tube is opened and the solution with the target molecule is collected [40].

2.2.1.3 Ion exchange chromatography

Ion exchange chromatography is a separation technique based on charge, where negatively charged molecules bind to positively charged beads in a column, and vice versa. If a protein has a high pI, it will be negatively charged at physiological pH and a cation exchanger (positively charged column) will be used. The bound protein can then be eluted by increasing the concentration of NaCl. The positively charged sodium ions compete for the binding sites to the negatively charged column and the protein is released [4].

2.2.1.4 Gel-filtration chromatography

Gel-filtration chromatography is a separation technique based on size. The beads in the column are porous, thus, when a solution with molecules of different sizes is applied, small molecules will enter the porous openings in the beads. The larger molecules are too big to enter the openings and as a result, they flow through the column more rapidly than the smaller molecules. The elute is collected in fractions and the obtained protein is stored in a suitable buffer [4].

2.2.2 Protein crystallization

Growing protein crystals with high enough quality for structure determination is the rate-limiting step in x-ray crystallography work. The basic principle of crystallization is to take a solution of a sample, in high concentration, and induce it to come out of solution in such a way that well-ordered crystals will form [44].

Proteins are difficult to crystallize for many reasons. They are difficult to purify, and even with near 100 % purity, individual molecules can exist in different conformational states which antagonize the formation of an ordered crystal. Compared to small molecules, protein-protein contacts are comparable in binding energy, but the number of contacts in proportion to molecular weight are far fewer, and hence weaker. Proteins are relatively large and complex molecules allowing them to form many different intermolecular contacts. In an ordered crystal, only one or a few of those contacts can exist[23].

For a crystal to form, the interactions must be appropriate in geometric arrangement, degree of specificity and strength. The solution, in which the protein is stored, also influence these properties. Hence, by optimizing parameters such as pH, buffer, additives, temperature, protein concentration etc. the environment could be chosen to favor crystal formation [23].

One commonly used additive, or precipitant, is PEG (polyethylene glycol). This molecule acts as a space-filler and is utilized because of its "excluded volume effect". PEG excludes some of the solution volume from the protein molecules by taking up space and as a result, more effective attractions between the protein molecules are made. One can imagine the situation as restricting the proteins to a smaller container and thus increasing potential interactions and favoring association [23].

One other commonly used precipitant is different types of salts. Low concentrations of salts will decrease the amount of electrostatic interactions between the protein molecules, thus increasing the solubility and stability of the protein. At high salt concentrations, protein solubility decreases and starts to aggregate. The salt ions compete with the protein molecules in binding to water, resulting in an increase in electrostatic interactions between the protein molecules [2].

Protein crystallization requires two steps, nucleation (formation of a nucleus) and crystal growth. Protein crystallization heavily depends on entropy and an energy barrier has to be overcome for nucleation to occur [52, 19]. The Gibbs free energy for nucleation at thermodynamic equilibrium (ΔG) can be described as the sum of two energies. The negative term, the volume free energy (ΔG_v), is the energy required for bonds to form between molecules in the crystal. The positive term, the surface free energy (ΔG_s), represents the unsatisfactory bonds on the surface of the crystal [52, 23]. Supersaturation is the driving force behind nucleation. When a solution contains the maximum number of dissolved molecules at thermodynamic equilibrium, the solid state is more stable than the liquid state, resulting in a decrease in the Gibbs free energy and a spontaneous formation of crystal nucleation [52]. Once a nucleus has formed, crystal growth follows spontaneously [12].

2.2.2.1 Hanging drop vapour diffusion

One of the most popular methods for protein crystallization is hanging drop vapour diffusion. The concept is quite straightforward. A drop containing equal parts of protein and crystallization condition (a solvent comprised of crystallization inducing components, for example PEG, salts, etc. Synonymous with crystallization cocktail), is sealed in a chamber with a solution of a high concentration of precipitant, see Figure 2.3. As a result, water vapor diffuses out of the drop until the osmolarity of the drop and the precipitant are equal. The decrease in water content in the drop causes the protein and precipitant concentration to increase which drives the system to the supersaturated phase. Hopefully, the nucleation zone is reached, where crystalline solids are formed. If not, an amorphous precipitant is formed instead [20]. As the crystal grows, more and more protein molecules come out of the solution and the protein concentration drops. The crystal will continue to grow until the protein concentration drops to the solubility limit, see Figure 2.4 [23, 12].

The aim of protein crystallography is to obtain high quality, three-dimensional crystals that can be exposed to the beam during x-ray crystallography. The crystals are usually required to be a minimum of 0.1 mm in the longest dimension [44].

2.2.2.2 Soaking

Structure-based drug design requires information about the protein in complex with ligands (potential drug candidates). Once high quality crystals have been generated

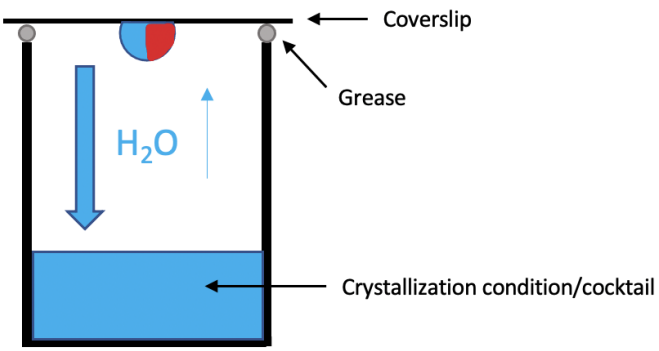


Figure 2.3: A schematic representation of a hanging drop set up. A drop of equal parts of protein (red) and crystallization condition (blue) is placed on a cover slip which has been sealed to a well containing a high concentration of precipitant solution (crystallization condition). Representation inspired by [8].

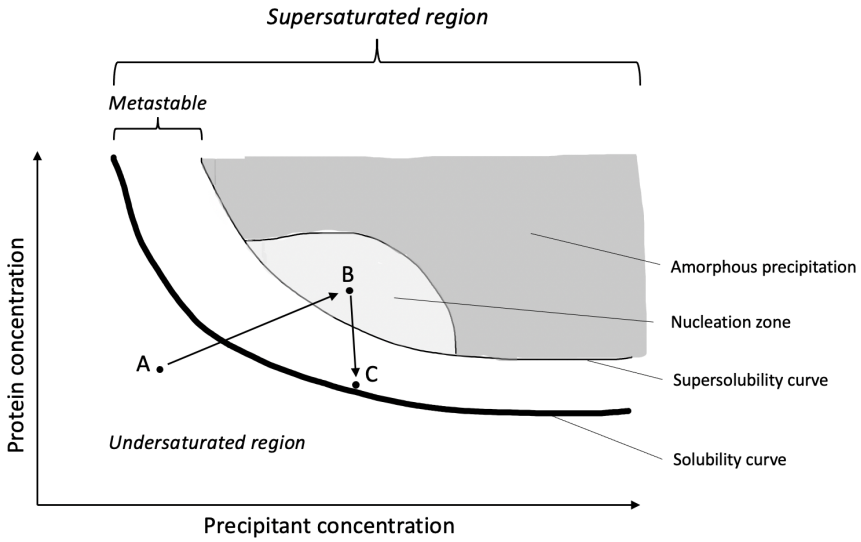


Figure 2.4: A schematic representation of a crystallization phase diagram. When the drop is set up, the system is in the unsaturated region (A). As water diffuses out of the drop, the protein and precipitant concentration increases which drive the system out of the undersaturated region up towards the supersaturated region. If the system reaches the nucleation zone (B), solid (crystalline) particles start to form. As the crystal starts to grow, more and more protein molecules come out of the solution, and the protein concentration drops (C). Illustration is inspired by [12].

they are soaked with ligands in soluble form. The ligand can access the active site through diffusion within the crystal lattice. The advantage of using soaking is the ability to gather many crystals of known structure and diffraction quality and them soak them in compounds of interest, allowing for rapid and reproducible procedures [16, 14].

2.2.2.3 Co-crystallization

If soaking is unsuccessful, i.e. the ligand does not bind to the active site or induces conformational changes (crystal cracking or dissolution) the crystal can not be used for x-ray crystallography. Co-crystallization is a method where the ligand of interest is added to the protein prior to a crystallization experiment. Thus, the ligand is free to bind to the active site in the solution allowing for more freedom in changes in lattice structure [14]. However, this method requires more resources in terms of time and material compared to soaking, as appropriate crystallization conditions must be found for each ligand [32].

2.2.3 Data collection and structure determination

When crystals of sufficient quality have been obtained, the next step is to analyze their diffraction pattern by conducting x-ray crystallography experiments. The crystal is mounted on a goniometer, a stand that rotates the crystal as the x-ray beam hits the crystal [44]. At specific angle intervals, one image of the diffraction pattern is taken (if a crystal is rotated 360 degrees and one image is taken per 0.1 degree rotation, 3600 images are taken in total). One data set contains all diffraction pattern images. It is possible to collect data from different places on the crystal. If the crystal is of high quality, a helical scan can be performed where two points on the crystal are chosen, and the beam moves along the crystal (while rotating) from the starting to the endpoint. It can be beneficial to utilize the full diffraction potential of a crystal by exposing its total volume. However, if there are visible damages to the crystal or if it is small in size, a single point can be chosen [17].

It is important to rotate the crystal during the experiment. As the x-rays hit planes of atoms in the crystal, they are diffracted. The diffracted waves create a unique pattern of spots (also called reflections) on a detector. The pattern arises due to the condition for constructive interference formulated by Bragg's law ($\lambda = 2sin\theta$). Numerous of reflections are generated during beam exposure, but only some reflections meet the condition of constructive interference, while most reflections are cancelled out due to destructive interference [17].

The diffraction pattern (the reciprocal lattice) represents the Fourier transform of the real space crystal lattice. Each reflection on the diffraction pattern contains information about the protein crystal. The first step in data processing is indexing where each reflection is assigned an index denoted with three integers (h,k,l). Information about crystal orientation, initial estimation of unit cell dimensions and crystallographic symmetry is obtained in this step [35, 44]. The second step is integration, where the intensities of the reflections are measured. The intensity, I, and the background value, Iσ, are saved. Scaling is done so the intensities of all the images in the data set can be related. The output from scaling is a computer file that contains the index of each reflection and its measured intensity[44].

As stated earlier, the diffraction pattern is the Fourier transform of the real space lattice. Information about the coordinates of the atoms in the unit cell (real space) are stored in the reflections on the diffraction pattern, where each reflection corresponds to a wave consisting of an amplitude and a phase, known as a structure factor. Hence, the inverse Fourier transform of the diffraction pattern will provide information about the atomic coordinates in the protein crystal [44, 49]. The primary product of

x-ray crystallography is an electron density map, which describes the distribution of electrons in the unit cell. The function of the electron density map (ρ) has to be solved at each point within the unit cell (x,y,z):

$$\rho(xyz) = \frac{1}{V} \sum_{\substack{hkl \\ -\infty}}^{+\infty} |F(hkl)| e^{-2\pi i[hx+ky+lz-\phi(hkl)]}$$

The equation above represents the Fourier transform between the real space and reciprocal space, where ρ(xyz) is the electron density at point x,y,x, |F(hkl)| is the amplitude and ϕ(hkl) is the phase of reflection hkl. V is the volume of the unit cell [49].

The amplitude is relatively easy to calculate, as it is the square root of the measured reflection intensities. The phases, however, are more difficult to calculate. Information about phases is not available from experimental data (the reflections) and thus, needs to be obtained elsewhere. A common method of obtaining information about phases is molecular replacement (MR). It is used when a closely related protein structure is available. Phases are "borrowed" from the already solved structure to calculate new structure factors, hence, there will always be an initial bias towards the solved structure [49, 44].

With information about amplitudes and phases, the electron density map can be calculated. The result is a three-dimensional grid within the unit cell, where the electron density is calculated at each of the grid points [44]. The resolution of the electron density map heavily relies on the number of measured reflections on the diffraction pattern. A true reconstruction of the crystal structure would require all F(hkl) terms (structure factors) in the Fourier summation which is impossible due to practical and theoretical reasons. Because the real space is inversely scaled to the reciprocal space, reflections scattered at the highest angle represent the smallest spacing between atoms [49].

The resolution is determined by statistics of the collected data from the outermost shell on the diffraction pattern. R_{merge} is a statistic for the precision of the measurements of each unique reflection, and should not exceed 5 % overall, but be less than 50 % in the outermost shell for high-quality data [22, 49]. The signal to noise ratio (I/σI) should not drop below 2 in the outermost shell. Completeness is a statistic defined by the number of measured reflections compared to the number of theoretically possible reflections unique to the given crystal symmetry. It is given as a percentage and should be near 100 % [3]. Improvements to the electron density map can be done by refinement, where computer programs perform sets of operations and calculations to improve phases [44].

When an electron density map of sufficient quality has been obtained, the model building can commence where sequences of amino acids are fitted to the electron density map. The electron density map and the structure model can be visualized in computer programs where manual adjustments and improvements to the model can be done [44]. Both files (electron density map and structure model), are run in refinement programs to improve phases and present statistics of how well the structure model agrees with the electron density map. Some of the important statistical parameters are (but not excluded to) the B-factor, the R-factor (R_{free} and

R_{work}) and the clashscore. The B factor reflect atomic displacement and quantifies the uncertainty of each atom. It ranges from 2-100 Å^2 and should be as low as possible [49]. The R-factor is used to assess the structure model quality, it is a measure of how well the structure model agrees with the crystallographic data and ranges from 0 - 1, the lower the better [46]. The clashscore is the number of serious clashes (overlapping atoms) per 1000 atoms and should be as low as possible [13]. When a sufficiently accurate structure model is obtained, supported by statistics, it is outputted as a PDB file and can be deposited to the Protein Data Bank [44].

2.3 Thermal shift assay

The foundation of protein crystallization is the protein's ability to organize itself into periodic arrangements. This is achieved when the protein forms moieties of the same kind and shape, within the asymmetric unit of the crystal. If such stable moieties can not form, crystallization will not occur [18]. The chances of finding successful crystallization conditions are greater when the protein exhibits conformational stability. Thus, finding an optimal storage buffer that keeps the protein in its most stable conformation is of great interest [36].

A thermal shift assay (TSA) utilizes a hydrophobic fluorophore, that only binds to hydrophobic amino acids in a protein. Ideally, at low temperatures, the protein is completely and correctly folded and no hydrophobic areas are exposed to the exterior. As temperature increases, the protein starts to unfold and hydrophobic areas become exposed, and allowing for the fluorophore to bind and fluorescence. The melting temperature (T_m) of a protein in a TSA, is approximated by the temperature at which the hydrophobic areas become exposed and is a reporter of the protein's thermal stability, which is correlated to the structural order and flexibility of the protein. To determine the T_m, the fluorescence is plotted against temperature, see Figure 2.5. The curve is ideally sigmoidal and the T_m corresponds to the midpoint of the transition curve [36].

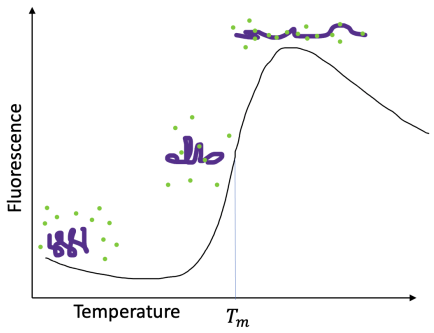


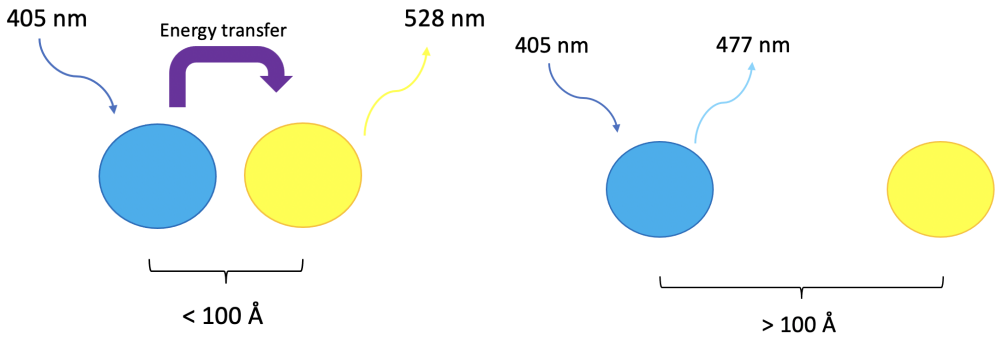
Figure 2.5: Schematic representation of an ideal TSA curve. At low temperatures, the protein (purple) is completely folded and fluorophores (light green) can not bind. As temperature increases, the protein unfolds, exposing hydrophobic areas and allowing the fluorophores to bind. At some point, a temperature is reached where the protein is fully denatured and maximum binding and fluorescence occur. At even higher temperatures the protein aggregates and the fluorophores dissociate. Representation inspired by [24].

2.4 SDS-PAGE

SDS-PAGE, short for sodium dodecyl sulfate–polyacrylamide gel electrophoresis, is a commonly used method for the separation of proteins or other macromolecules based on size and charge. SDS is a detergent that denatures the protein and coats it with a negative charge, proportional to the mass of the protein. The SDS-coated protein is then subjected to electrophoresis where the proteins migrate through a gel, driven by an electric field, from the negative to the positive electrode. Small molecules move more rapidly through the gel (less resistance), hence, molecules are separated based on size. The resulting protein bands in the gel are visualized by staining [5].

2.5 Fluorescence resonance energy transfer

Fluorescence resonance energy transfer (FRET) is a process where energy is transferred from an excited fluorophore (the donor) to another fluorophore (the acceptor). The energy transfer can only happen if, one, the donor and the acceptor are in close proximity to each other, typically 10 - 100 Å, see Figure 2.6 [41]. Two, the emission spectrum from the donor fluorophore must overlap with the excitation spectrum of the acceptor fluorophore [10]. By measuring the fluorescence over time, it is possible to detect binding or dissociation between molecules.



(a) (b)

Figure 2.6: A schematic illustration of how distance influences energy transfer. In (a) the distance between the donor and the acceptor is small enough for energy transfer to occur. In (b) the distance is to big resulting in fluorescence from the donor. Illustration inspired by [10].

2.6 Enzyme kinetics

The study of enzyme-catalyzed reactions and the rate at which it occurs (kinetics), provides information about the enzyme's function. A frequently used model for enzymatic reactions, under the assumption that the concentration of the enzyme remains constant, is the Michaelis-Menten equation [50].

The enzyme (E) binds to a substrate (S) to form an enzyme-substrate complex (ES). This leads to a formation of a product (P) which then leaves the enzyme [6]:

$$\mathrm{E} + \mathrm{S} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_2} \mathrm{P} + \mathrm{E}$$

The Michaelis-Menten constant K_M, relates the rate of catalysis to the concentrations of substrate and enzyme, and the rates of the individual steps (has the unit of concentration) [6]:

$$K_M = \frac{k_{-1} + k_2}{k_1}$$

After a number of subsequent simplifications, rearrangements and assumptions to the differential equations corresponding to the dynamics of these reactions, the Michaelis-Menten equation is obtained [6]:

$$V = V_{max} \frac{[S]}{[S] + K_M}$$

where the reaction velocity (V) is a function of substrate concentration. When the substrate concentration ([S]) is much less than K_M, $V \propto \frac{V_{max}}{K_M}$ and the reaction is first order with the rate of the reaction is directly proportional to the substrate concentration. When the substrate concentration is much greater than K_M, the reaction is zero order and independent of substrate concentration, as the rate is maximal and $V \propto V_{max}$. When the substrate concentration is equal to K_M, then $V \propto \frac{V_{max}}{2}$, hence, K_M is equal to the substrate concentration at which the reaction rate is half its maximal value. Therefore, the value of K_M provides a measure of the required substrate concentration for a significant catalysis to take place [6].

The turnover number k_{cat}, is the number of substrate molecules converted into product by an enzyme per unit time when the enzyme is fully saturated with substrate. It is obtained by dividing the maximal velocity (V_{max}) by the enzyme concentration present in the reaction [6].

Chapter 3

Materials and Methods

To find what variables positively influence protein and protein-inhibitor crystallization, several experiments were conducted, many in parallel. This project was mainly laboratory and constituted of two parts. First, screening and optimizing crystallization conditions by setting up crystallization drops and second, performing biochemical assays to get a better understanding of the probes' inhibitory effect on the nsP2's cleavage ability.

3.1 Protein expression

Frozen bacterial cells, *E.coli* strain BL21(DE3), containing the VEEV nsP2 protease gene had been prepared by FOI and were stored in glycerol stocks at - 80°C. The plasmid construct contained the protease gene fused to thioredoxin (Trx) and a hexahistidine tag with a *Tobacco etch virus* (TEV) cleaving site. Some of these cells were streaked onto agar plates and were stored overnight at 37°C and then placed in a fridge at 4°C over the weekend. A single bacterial colony was identified and used for overnight cultivation which was produced at 37°C with 100 ml 1 x Lysogeny Broth (LB) medium and 100 μl kanamycin (antibiotics) at 150 rpm.

Four shake flask cultivations, each containing 11 ml of overnight culture, 2000 ml 1 x LB and 2 ml kanamycin were incubated at 37°C and 140 rpm. When the bacterial cells had grown to an OD_{600} value of 0.581, the temperature was lowered to 20 °C. When the OD_{600} reached 0.894, 1 ml of IPTG (induces gene expression) was added to all four cultivations and they were left to incubate overnight. The cell culture was then centrifuged and the cell pellet was collected in falcon tubes and stored in -20°C.

3.2 Protein purification

The main purification methods were three chromatography steps, affinity, ion and gel permeation chromatography. These were performed using an ÄKTA™ pure chromatography hardware system together with a UNICORN™ 7 control software.

FOI had existing purification protocols and programs for VEEV nsP2, which were followed. Throughout the purification process, samples were saved and run in

SDS-PAGE gels to ensure that the nsP2 was not lost in any of the steps.

The following purification steps were performed twice, i.e. two batches of protein were produced. The first batch was produced in mid-February and the second in late March.

3.2.1 Sonication

One falcon tube, containing the cell pellet, was taken out from the freezer and left to thaw, 30 ml of lysis buffer was added, see Table A.1 for buffer ingredients. The cells were sonicated approximately 10 times for 60-90 seconds directly in the falcon tube, surrounded by ice to reduce the risk of overheating the cells. The lysate was then centrifuged at 18000 rpm for 60 min at 4 °C. The supernatant, expected to contain the protein, was transferred into a falcon tube.

3.2.2 Immobilized metal affinity chromatography

A 5 ml HisTrap Crude column, Ni-NTA, was attached and installed to the ÄKTA™ system. The air was pumped out from the pumps and the column and pumps were cleaned with MilliQ (ultrapure water). The column was equilibriated with 5 column volumes (CV) of wash buffer, 5 CV of elution buffer and lastly 5 CV of wash buffer, see Table A.1 for buffer ingredients. The lysate was loaded to the ÄKTA™ and the program for affinity chromatography was run. Based on the results from the chromatogram, the fractions expected to contain the protein were pooled together.

3.2.3 Dialysis and His-tag removal

The pooled elution and 1.3 ml of TEV-protease were poured into a dialysis tube, which was sealed and placed in 1000 ml dialysis buffer for one hour. After which, the buffer was replaced with new dialysis buffer (1000 ml) and left overnight under slow agitation. The next morning, the dialysate was transferred into a falcon tube. Samples were run in an SDS-PAGE where two separate bands were visible, the Trx-His-TEV complex and the nsP2 protease, confirming that the His-tag removal was successful.

3.2.4 Desalting

To exchange the elution buffer to low salt gradient buffer, a HiTrap-desalting column was used. It was attached to the ÄKTA™ and equilibrated with 5 CV of low salt gradient buffer and 5 CV of MilliQ, see Table A.1 for exact buffer ingredients. The dialysate was loaded onto the ÄKTA™ and the desalting program was run. The fractions expected to contain the protein, based on the chromatogram, were pooled together.

Due to human error, the first batch was not desalted which meant that the protein was kept in the elution buffer for the subsequent ion chromatography step. However, this error was believed to have a negligible effect on the resulting purification process.

3.2.5 Ion exchange chromatography

The pooled fractions were loaded onto the ÄKTA™ where a Resource S column (cation exchanger) had been installed and equilibrated with 5 CV of low salt buffer, 5 CV of salt gradient buffer and lastly, 5 CV of low salt buffer, see Table A.1 for exact buffer ingredients. The nsP2 protease has a high pI which means its net charge is negative at physiological pH and will bind to the positively charged column. The ion exchange chromatography program was run and the fractions expected to contain the protein, based on the chromatogram, were pooled together.

3.2.6 Gel-filtration chromatography and protein concentration determination

The pooled elute was concentrated in an amicon-tube together with gel-storage buffer until approximately 1 ml remained. A Superdex 200 column was installed and equilibrated with 0.5 CV MilliQ and 1 CV gel-storage buffer, see Table A.1 for exact buffer ingredients. The 1 ml concentrated sample was loaded onto the ÄKTA in two runs, approximately 500 μl at a time.

The fractions from the two runs expected to contain the purified protein were pooled together and concentrated in an amicon-tube until approximately 600-800 μl remained.

A spectrophotometer was used to measure the absorbance at $\lambda = 280$ and together with Beer-Lambert's Laws equation:

$$A = c \cdot l \cdot \epsilon \tag{3.1}$$

where A is the absorbance, c is the protein concentration, l is the length of the path light travels in the solution (cm) and ϵ is the molar extinction coefficient, the final protein concentration was calculated.

3.3 Initial crystallization screen

The screening of different crystallization conditions started by putting up drops using two commercial screens with premixed cocktails (Molecular Dimensions). These were the Ligand-Friendly screen (LFS) and the ShotGun screen (SG1), both containing 96 conditions each. The use of commercial screens for initial screening is commonly used since it is a convenient way of testing a variety of conditions without having to buy full sets of chemicals, saving both time and money. It is also a suitable starting point when there is limited knowledge about what parameters positively or negatively influence protein crystallization. Even though FOI had already obtained high-quality crystals, there was a desire to investigate as many conditions as possible.

3.3.1 Setting up hanging drops

300 μl of each condition was pipetted into the wells of a greiner 24 well combo plate (Molecular Dimensions). 1 μl of crystallization condition was pipetted onto a cover slip and then 1 μl of protein solution with a concentration of 15 mg/ml was pipetted

on top of the reservoir drop, allowing for the two solutions to diffuse and mix to form one drop of 2 μl solution. The coverslip was then flipped upside down with a set of tweezers, with the drop hanging down, and sealed on top of the well. The plates were stored in 4°C.

3.4 Thermal shift assay

The stability of a protein is an important variable in crystallography, it was, therefore essential to investigate in what storage buffer the nsP2 protease occupies the most stable conformation. A TSA was performed by screening 36 buffer solutions with variations in salt concentrations, additives, pH etc. see Table 3.1 for the full ingredients list. The overall design of the assay was inspired by Reinhard et al. [36] and some of the buffer solutions they tested were included in this assay. Additionally, buffer solutions that FOI was interested to test were also added, for example varying percentages of DMSO.

100 μl of sample volume was prepared for each buffer condition, it contained 15 μl of 2 mg/ml nsP2 + 10 μl 10 X GloMelt™ dye (Biotium) + 75 μl buffer solution. The final reaction volume was 20 μl, and was pipetted into two of the wells of a 96-well (A1 - H12) PCR plate (Bio-Rad Laboratories), as the assay was done in duplicates. When all wells had been filled with their respective buffer solution, the plate was sealed with sealing tape and centrifuged for approximately 30 seconds at 4000 rpm. The plate was placed in a CFX96™ real-time PCR detection system machine (Bio-Rad Laboratories) where it was gradually heated from 25 °C to 90 °C. Changes in fluorescence were monitored continuously and the data was analyzed with a CFX Manager™ software (Bio-Rad Laboratories). The software generated two sets of graphs, first, fluorescence vs temperature and second, the negative first derivative of the fluorescence vs temperature.

Table 3.1: Summary of buffer solutions screened in the TSA.

Well	**Buffer solution**
A1 - A2	50 mM Tris + 500 mM NaCl, pH 7.6
B1 - B2	50 mM Tris + 400 mM NaCl, pH 7.6
C1 - C2	50 mM Tris + 200 mM NaCl, pH 7.6
D1 - D2	50 mM Tris + 150 mM NaCl, pH 7.6
E1 - E2	50 mM Tris + 100 mM NaCl, pH 7.6
F1 - F2	50 mM Tris + 50 mM NaCl, pH 7.6
G1 - G2	50 mM Tris + 0 mM NaCl, pH 7.6
A3 - A4	50 mM HEPES + 500 mM NaCl, pH 7.4
B3 - B4	50 mM HEPES + 400 mM NaCl, pH 7.4
C3 - C4	50 mM HEPES + 200 mM NaCl, pH 7.4
D3 - D4	50 mM HEPES + 150 mM NaCl, pH 7.4
E3 - E4	50 mM HEPES + 100 mM NaCl, pH 7.4
F3 - F4	50 mM HEPES + 50 mM NaCl, pH 7.4
G3 - G4	50 mM HEPES + 0 mM NaCl, pH 7.4
A5 - A6	50 mM Tris + 150 mM NaCl + 10 % DMSO
B5 - B6	50 mM Tris + 150 mM NaCl + 5 % DMSO
C5 - C6	50 mM Tris + 150 mM NaCl + 3 % DMSO
D5 - D6	50 mM Tris + 150 mM NaCl + 2 % DMSO
E5 - D6	50 mM Tris + 150 mM NaCl + 1 % DMSO
A7 - A8	50 mM Tris, pH 8.5
B7 - B8	50 mM Tris, pH 8
C7 - C8	50 mM Tris, pH 7.5
D7 - D8	50 mM Bicine, pH 9
E7 - E8	50 mM Bicine, pH 8.5
F7 - F8	50 mM Bicine, pH 8
A9 - A10	50 mM HEPES, pH 8
B9 - B10	50 mM HEPES, pH 7.5
C9 - C10	50 mM HEPES, pH 7
D9 - D10	50 mM HEPES, pH 6.5
A11 - A12	50 mM Sodium Acetate, pH 5
B11 - B12	50 mM Bis-Tris Propane, pH 6.5
C11 - C12	50 mM Mops, pH 7
D11 - D12	50 mM Sodium Citrate, pH 5
E11 - E12	50 mM MES, pH 6.7
F11 - F12	50 mM Sodium Phosphate, pH 7
G11 - G12	H_2O (blank)

3.5 Optimizing crystallization conditions

Based on the results from the TSA (Figure 4.5), sodium phosphate buffer increases nsP2 stability compared to the other 35 buffer solutions that were screened. A new storage-gel buffer containing 50 mM sodium phosphate + 2 mM TCEP + 150 mM NaCl was therefore prepared. 200 μl of nsP2 (27 mg/ml) was concentrated in an amicon-tube together with the new storage buffer until the tris buffer had been replaced. The concentration was calculated to 16.9 mg/ml by measuring the absorbance at $\lambda = 280$ and using equation 3.1.

Based on the hits from the initial crystallization screen, listed in Table 4.1, new crystallization cocktails with varying concentrations of precipitants, pH, additives and protein concentration were developed. The optimization method of each condition followed one basic strategy, illustrated in Figure 3.1. The concentration of precipitant from a hit was kept the same in column 3 and 4, decreased in columns 1 and 2 and increased in columns 5 and 6. The pH was kept the same in row b and/or c, decreased in row a and increased in d. Most of the drops were set up with a protein concentration of 15 mg/ml but were in some cases decreased to 10 or 7.5 mg/ml. With this method, surrounding concentrations and variations of precipitant, pH etc. of the hit, could be tested to see if they yielded crystals with higher quality or not. The drops were set up in the same way as described in section 3.3.

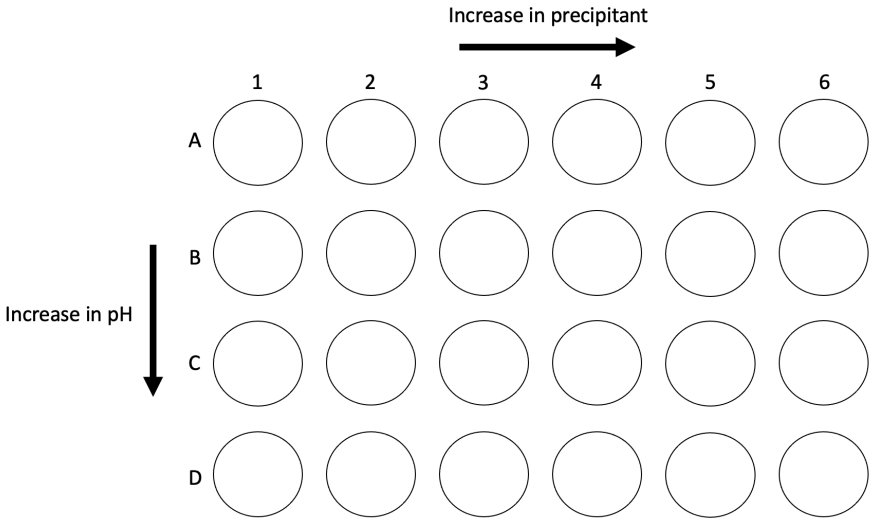


Figure 3.1: A schematic illustration of the optimization strategy where the concentration of precipitant and pH increases in each column/row.

For example, the optimization design based on hit SG1, tube 1-19 (Table 4.1) with 20 % w/v PEG 3350 + 0.2 M sodium formate, had 18 % w/v PEG 3350 in columns 1 and 2, 20 % w/v PEG 3350 in columns 3 and 4, and 22 % w/v PEG 3350 in columns 5 and 6. In this case, there was no change in pH since it was not specified in the SG1 screen. This meant that it was possible to screen two additives, sodium formate and ammonium formate, as they had identical crystallization cocktail composition (apart from the salt). The design of this screen can be seen in Figure 3.2 below.

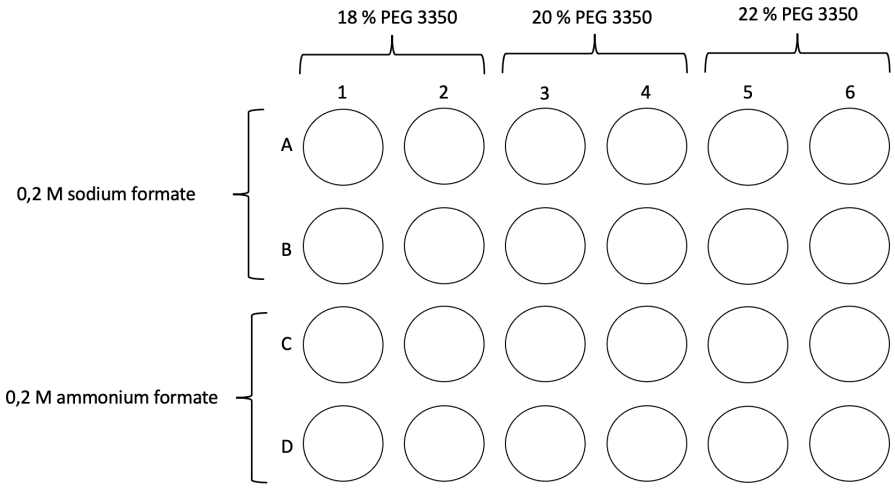


Figure 3.2: A schematic illustration of the optimization strategy based on hit SG1, tube 1-19.

One other example is the optimization of tube 1-14 from SG1 (Table 4.1). Here, the percentage of PEG and variation in pH follow the structure described in Figure 3.1. Wells C3 and C4 had the same crystallization cocktail as the SG1 screen and the surrounding wells either had increasing or decreasing percentages of PEG and variations in pH, as is depicted in Figure 3.3 below.

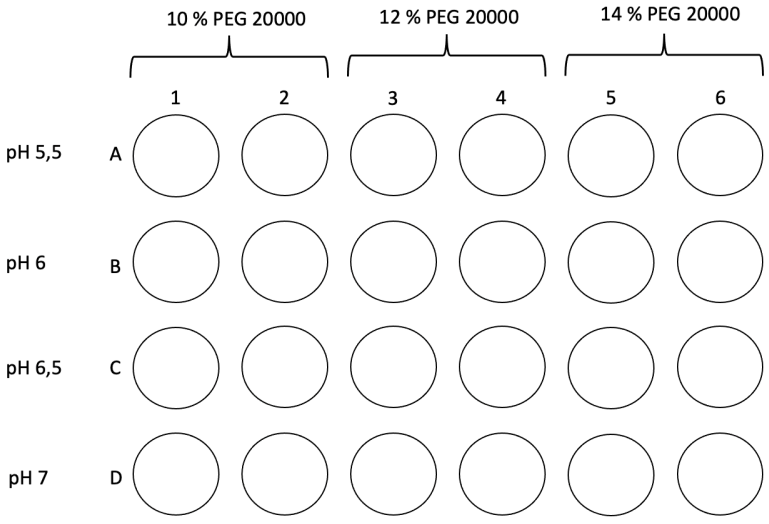
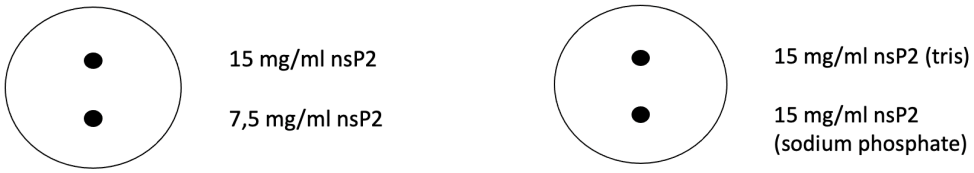


Figure 3.3: A schematic illustration of the optimization strategy based on hit SG1, tube 1-14.

It is possible to set up two drops on one coverslip, which allows screening of two nsP2 concentrations or storage buffers, per crystallization cocktail. Thus, a plate with 24 wells results in 48 drops. See Figure 3.4 for a schematic overview. This method reduces the use of chemicals, as less crystallization cocktail and time are needed, since the screen is reduced to one plate instead of two.



(a) (b)

Figure 3.4: A schematic of how to screen two protein concentrations (a) and two storage buffers (b) with one crystallization cocktail on one coverslip. A black dot represents one drop comprising of 1μl protein solution + 1μl crystallization cocktail.

It is worth noting that the chemicals used for these optimization screens sometimes differed from the chemicals used in the commercial screen, in the sense that they came from different manufactures. Most chemicals used in the optimization screens, except from PEG 20 000 which was ordered (Thermo Scientific), were already in storage at FOI and therefore used in the crystallization cocktails. Additionally, the hits from the LFS contained a second precipitant, 10 % v/v ethylene glycol (see Table 4.1), which was not in storage at FOI nor was it ordered, and was thus excluded from the crystallization cocktails. The same applied to the salts lithium chloride, magnesium formate and sodium bromide. Table 3.2 summarizes the conditions used for optimization, the rows colored in red represents the conditions that were not used due to absence of chemicals.

A total of 24 plates were set up to find optimal crystallization conditions, originally based on the hits from the LFS and SG1 screen. The result from each condition was evaluated by studying the drop in a light microscope (Leica MDG32) and then, depending on the results, new plates were set up according to the optimization strategy explained earlier. 12 of the 24 plates had two drops per coverslip, resulting in 864 drops.

Table 3.2: Summary of crystallization conditions that were used in the optimization process. The rows colored in red represent the conditions that were not used due to absence of chemicals.

Screen kit	**Tube**	**Precipitant**	**Buffer**	**pH**	**Salt**
SG1	1-14	12 % w/v PEG 20 000	0.1 M MES	6.5	None
SG1	1-19	20 % w/v PEG 3350	None	-	0.2 M Sodium formate
SG1	1-24	20 % w/v PEG 3350	None	-	0.2 M Magnesium formate dihydrate
SG1	2-37	20 % w/v PEG 3350	None	-	0.2 M Ammonium formate
LFS	1-20	20 % w/v PEG 6000	0.1 M MES	6	0.2 M Ammonium chloride
LFS	1-22	20 % w/v PEG 6000	0.1 M MES	6	0.1 M Magnesium chloride hexahydrate
LFS	1-23	20 % w/v PEG 6000	0.1 M MES	6	0.1 M Calcium chloride dihydrate
LFS	1-33	20 % w/v PEG 6000	0.1 M HEPES	7	0.2 M Lithium chloride
LFS	2-14	20 % w/v PEG 3350	0.1 M Bis Tris Propane	6.5	0.2 M Sodium bromide
LFS	2-18	20 % w/v PEG 3350	0.1 M Bis Tris Propane	6.5	0.2 M Sodium formate

3.6 Co-crystallization

After knowing under what crystallization conditions the nsP2 forms crystals, the next step was to find conditions where the protein would grow crystals together with one of the three probes in the active site. The addition of a probe in the hanging drop adds one more variable to the crystallization process, and so there was a desire to screen potential co-crystallization conditions.

To ensure that there was soluble probe in excess in the hanging drop, the final concentration ratio of soluble probe to soluble nsP2 was set to 6:1. Appropriate dilutions and combinations of nsP2 and probe were conducted to reach the concentration ratio of 6:1. The nsP2-probe mix was left at room temperature for at least one hour before the hanging drops were set up.

The LFS was used once more as an initial screen together with DW0408. Out of the 96 crystallization conditions, only one yielded crystalline structures, namely 0.1 M bis-tris propane + 0.1 M sodium formate + 20 % w/v PEG 3350 (+ 10 % v/v ethylene glycol). This condition was further optimized by varying the variables as described earlier in section 3.3. Apart from this, new drops containing the nsP2 and the three probes (DW0406, DW0408, and DW0411) were set up based on the hits from the optimization procedure performed in section 3.5.

3.7 Soaking

The day before freezing the crystals, all drops from the optimization procedures, sections 3.3 and 3.5, were screened and the best crystals were chosen to be soaked with inhibitor. These crystals appeared to be of high quality and suitable for x-ray crystallography. See Table 3.3 for a full list of the crystallization conditions of the drops containing the chosen crystals.

The probes (DW0406, DW0408, and DW0411) were diluted to a concentration of 30 mM, respectively. The coverslip containing the chosen drops was detached from the well, turned upside down (the drop facing upwards) and an initial 0.5 μl of soluble probe was pipetted on top of the drop. When the 0.5 μl had diffused into the drop, an additional 2 μl of soluble probe was added, 1 μl at a time. The coverslip was turned upside down (the drop hanging downwards) and sealed to the well again. Five drops were soaked with soluble probe the day before freezing, and the remaining drops were soaked in the morning of freezing.

Table 3.3: Summary of drops chosen for soaking

[nsP2]	Batch	Storage buffer	Precipitant	Buffer	pH	Salt	Soaking Time (hours)	Probe
15 mg/ml	1	Tris	12 % w/v PEG 20 000	0.1 M MES	6.5	None	4	DW0411
15 mg/ml	1	Sodium phosphate	10 % w/v PEG 20 000	0.1 M MES	5.5	None	4	DW0411
15 mg/ml	1	Sodium phosphate	10 % w/v PEG 20 000	0.1 M MES	5.5	None	4	DW0408
15 mg/ml	1	Sodium phosphate	10 % w/v PEG 20 000	0.1 M MES	6	None	4	DW0406
15 mg/ml	1	Sodium phosphate	10 % w/v PEG 20 000	0.1 M MES	6	None	4	DW0411
15 mg/ml	1	Sodium phosphate	12 % w/v PEG 20 000	0.1 M MES	6	None	5	DW0408
15 mg/ml	1	Sodium phosphate	12 % w/v PEG 20 000	0.1 M MES	6	None	5	DW0411
15 mg/ml	1	Sodium phosphate	14 % w/v PEG 20 000	0.1 M MES	6	None	24	DW0406
15 mg/ml	1	Sodium phosphate	14 % w/v PEG 20 000	0.1 M MES	6	None	5	DW0408
7.5 mg/ml	1	Sodium phosphate	10 % w/v PEG 20 000	0.1 M MES	5.5	None	24	DW0411
7.5 mg/ml	1	Sodium phosphate	12 % w/v PEG 20 000	0.1 M MES	6	None	24	DW0408
15 mg/ml	1	Tris	18 % w/v PEG 3350	None	-	0.2 M Sodium formate	6	DW0411
15 mg/ml	2	Tris	18 % w/v PEG 3350	0.1 M Bicine	8	None	6	DW0408
15 mg/ml	2	Tris	20 % w/v PEG 3350	0.1 M Bicine	8	None	24	DW0408
15 mg/ml	2	Tris	20 % w/v PEG 3350	0.1 M Bicine	8	None	24	DW0406

3.8 Freezing

Freezing the crystals is a crucial step, in the sense that the crystals are being moved and lifted out from the drops, and that the quality of the freezing severely affect the x-ray crystallography experiment. The cryoprotectants were composed, based on the crystallization cocktails for each drop that contained crystals, selected for x-ray crystallography. The cryoprotectants comprised of the chemicals in the crystallization cocktail + approximately 25 % glycerol. Because most crystallization cocktails were similar in composition with only small variations in quantity, only five sets of cryoprotectants were be prepared.

The coverslip, containing a drop with crystals selected for freezing, was detached from the well and turned upside down (the drop facing upwards) under the light microscope. With the drop and crystals visible in the microscope, 1 μl of cryoprotectant was pipetted near the drop. Loops (Molecular Dimensions) of varying sizes (0.02 mm x 0.1 - 0.3 mm), mounted on a holder were used to retrieve one crystal in the drop.

Once the crystal had been properly positioned in the loop along the major axis, see Figure 3.5, it was lifted out from the crystallization cocktail drop, into the cryoprotectant drop for a few seconds and then quickly submerged in a liquid nitrogen bath. The loop with the crystal was then detached from the handle and placed in one of the 16 sample positions in a uni-puck (Molecular Dimensions), already submerged in the liquid nitrogen bath.

Figure 3.5: A schematic illustration of a crystal (blue rectangle) retrieved and positioned along the major axis in the loop.

When all of the 16 sample positions in the uni-puck were filled, the puck was transferred to a puck-shelved shipping cane (Molecular Dimensions) which was stored in a thermos containing liquid nitrogen. When the shipping cane was filled with pucks, the thermos was sealed and locked. A total of 75 crystals were retrieved, stored in the pucks and shipped to BioMAX, the x-ray macromolecular crystallography beamline of MAX IV Laboratory in Lund [28].

3.9 X-ray crystallography

The x-ray crystallography experiments were conducted at BioMAX in Lund, but overseen and controlled by myself and my supervisors at FOI in Umeå. A computer at BioMAX was remotely controlled from a computer at FOI using MXCuBE3, a user interface for macromolecular crystallography experiments used at BioMAX. While connected to this software, the user can carry out x-ray crystallography experiments and make real-time decisions regarding sample alignments, data collection etc.

On the morning of the experiment, staff at BioMAX had loaded the pucks in a sample exchanger, a container filled with liquid nitrogen with positions for each puck. The pucks' positions were then entered in EXI (EXtended ISPyB), a laboratory information management system, ensuring that the soon to be obtained data was

given a unique name and stored under a directory, named after the samples. The MXCuBE3 software was then synchronized to the EXI database to obtain the sample information.

After the sample information had been loaded onto MXCuBE3, it was time to start the experiments. Selected samples (loops with crystals) were automatically mounted on the goniometer and placed in camera view, making it possible to see the loop and the crystal. In the "data collection" section in MXCuBE3, the position of where the beam was to hit the crystal, was manually determined by clicking on a position on the crystal three times. After one click, the crystal rotated 120 degrees and the second position was chosen, then the crystal rotated 120 degrees again and the third position was chosen. These three positions represented the coordinates of where the beam would hit the crystal, as it rotates 360 degrees during the experiment.

Once the data collection was finished, the diffraction pattern was visualized using ALBULA, a software program that displays and analyses image data from x-ray and electron detectors. An initial estimation of the resolution and quality of the crystal could be obtained by studying the diffraction pattern. All data collected from the experiment from each crystal was automatically saved and processed by different software programs accommodated by BioMAX. These files were then uploaded to EXI allowing for further data processing and refinements on computers at FOI.

3.9.1 Data processing

The initial data processing that was done by the software programs at BioMAX comprised of indexing and integration, which provided calculated and proposed data of the crystal systems, completeness and some statistical parameters. The most promising data sets were selected and downloaded. Scaling was done using the software program *Aimless*. Here, the resolution was determined by evaluating the statistics of the data collected from the outer shell. MR was done in the software program *PHENIX.phaser*, which required information from both the experimental data (the output file from *Aimless*), and information about phases from a related protein. In this case, the already solved VEEV nsP2 protease structure, available from PDB, was used as a template. The refinement process was done in the software program *PHENIX.refine*.

In the software program *Coot*, the structure model and the electron density map were displayed and further manual adjustments to the model were made. After improvements to the model had been done in *Coot*, the model and the experimental data were run again in *PHENIX.refine* for further refinements. The refinement process in *PHENIX.refine* and the improvements to the model made in *Coot*, were done repeatedly until the model fitted the diffraction data with high enough statistical certainty.

Although outside the scope of this project, the next step was to build the probe in the active site. A file with information about the probe (chemical composition, atomic coordinates etc.) was uploaded to *Coot* and adjusted to fit the electron density map. When the probe was placed in a suitable position in the active site, the structure model of the protein, the structure model of the probe and the experimental data were run in *PHENIX.refine* for refinement. Once again, the refinement process in *PHENIX.refine* and the improvements to the models made in *Coot*, were performed

repeatedly.

3.10 Gel-based assay

One way of investigating the potency of the probes, was to perform dose-response studies i.e., at what probe concentration does the nsP2 protease stop cleaving its substrate. The design of this dose-response study was inspired by Zhang et al. who conducted a similar study in 2020 [51].

The study was an in vitro study, where a range of probe concentrations were incubated together with the nsP2 protease and its substrate, V1/2, in Eppendorf tubes. The plasmid construct for the substrate encoded a cyan fluorescent protein (CFP), an nsP2 protease cleavage site motif and a yellow fluorescent protein (YFP). The substrate had previously been expressed and purified by FOI and was stored in - 80°. An assay volume of 100 μl was prepared for each tube, with 50 mM HEPES pH 7.4 as assay buffer. Two SDS-PAGE gels were run, the first after four hours and the second after approximately 24 hours. In Table 3.4, the concentrations of compounds in each eppendorf tube are listed.

Table 3.4: Summary of compounds and their concentrations used in the gel-based assay

Tube	**Content**
1	10 μM V1/2
2	10 μM V1/2 + 5 μM nsP2 protease
3	10 μM V1/2 + 5 μM nsP2 protease + 2.5 μM probe
4	10 μM V1/2 + 5 μM nsP2 protease + 1.25 μM probe
5	10 μM V1/2 + 5 μM nsP2 protease + 0.625 μM probe
6	10 μM V1/2 + 5 μM nsP2 protease + 0.3125 μM probe
7	10 μM V1/2 + 5 μM nsP2 protease + 0.1562 μM probe
8	10 μM V1/2 + 5 μM nsP2 protease + 0.0781 μM probe
9	10 μM V1/2 + 5 μM nsP2 protease + 0.039 μM probe

3.11 Continuous FRET assay

For the determination of steady-state kinetic parameters, K_M and k_{cat}, a continuous FRET assay was performed where the cleavage of the substrate was monitored continuously at room temperature. The design of this assay was inspired by Hu et al. who conducted a similar assay in 2016 [25]. The substrate, V1/2, is cleaved by the nsP2 protease into two products, CFP and YFP. With an excitation wavelength of 434 nm and emission wavelengths of 475 and 530 nm for the two products respectively, the cleavage of the substrate was continuously measured over time, and thus, the rate of hydrolysis and the kinetic parameters K_M and k_{cat}, could be calculated.

The assay was performed multiple times, as the first couple of experiments resulted in untrustworthy data. By the last experiment, there was a limited amount of substrate remaining, and the assay was therefore performed in duplicates instead of triplicates

as described by Hu et al. [25]. The last experiment produced the most reliable data and its methodology is described in the following paragraph.

The assay was performed in a black opaque 96 well plate (ThermoFisher), with column one containing substrate, and columns two and three containing substrate and eventually 3 μM nsP2 protease. The substrate concentration ranged from 0 - 100 μM, see Table 3.5 for overall layout of the plate. 50 mM HEPES pH 7.4 was used as assay buffer.

Table 3.5: The overall layout of the FRET assay plate

	0 μM nsP2	3 μM nsP2	3 μM nsP2
	1	2	3
A	100	100	100
B	80	80	80
C	60	60	60
D	40	40	40
E	20	20	20
F	10	10	10
G	5	5	5
H	0	0	0
	S μM	S μM	S μM

First, the substrate was added to all wells. The plate was placed in a FlexStation 3 Multi-Mode Microplate Reader (Molecular Devices) where the emission was measured using a SoftMax Pro 7 software (Molecular Devices) for 20 min at 2 min intervals and one endpoint measurement at T = 20 min. Second, 3 μM nsP2 protease was added to the wells in columns two and three. The emission was measured again for 120 min at 2 min intervals. Third, the plate was covered with film and left to incubate overnight at room temperature. In the morning of the next day, at approximately T = 20 h, the film was removed and an emission endpoint was measured.

The fraction of cleaved substrate, f, was calculated from the emission ratios at each time point using the following equation:

$$f = \frac{\frac{\frac{ex434}{em530}}{\frac{ex434}{em475}} - r_{uncut}}{r_{cut} - r_{uncut}}$$

where r_{uncut} is the emission ratio measured for the substrate in the absence of enzyme, and r_{cut} is the emission ratio measured when the substrate was fully cleaved. The nanomols of substrate was calculated by multiplying f by the nanomols of substrate at time zero. For each substrate concentration, the amount of cut substrate (nanomoles) was plotted against time (min) using GraphPad. The initial velocity for each substrate concentration was calculated by fitting the data to a linear regression line (y = kx + m). The slope (value of k) is equivalent to the reaction rate i.e. the amount of product (nanomols) produced per minute. The rate of spontaneous

hydrolysis for each substrate concentration, measured from T = 0 - 3 hours, was subtracted from the reaction rates.

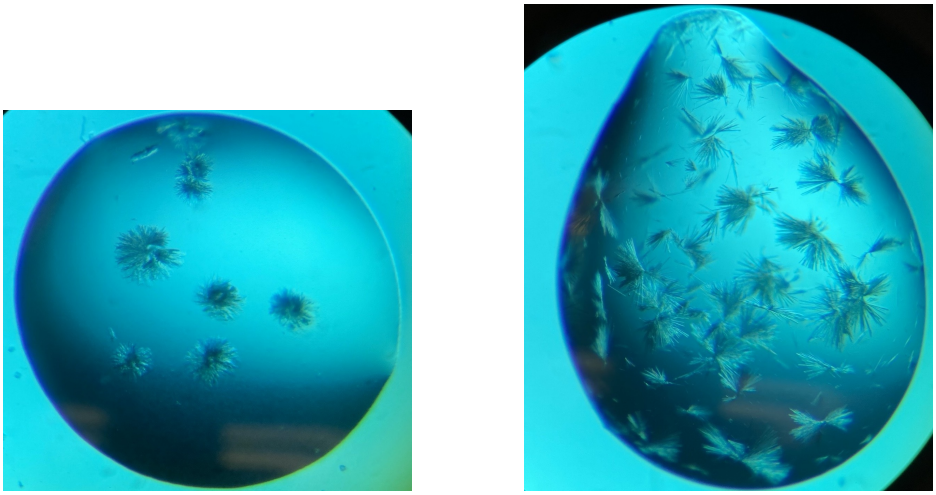
The reaction rates were plotted against the substrate concentrations, and the data were fitted to the Michaelis-Menten equation to obtain K_M and V_{max}. k_{cat} was calculated by dividing V_{max} with the nsP2 concentration present in each well.

Chapter 4

Results and Discussion

4.1 Initial crystallization screen

Out of the 192 screening conditions that were used in the initial screen, 10 conditions were considered to be hits. A hit is when the protein forms some type of crystalline structure. From this, optimization of different variables can be done to further facilitate crystallization formation of higher quality. The components of these conditions can be seen in Table 4.1. The majority of drops had, from what could be determined by studying the crystals in a light microscope, crystals of poor quality. The crystals in Figures 4.1 and 4.2 were thin and needle-like structures, deemed unfit for x-ray crystallography. Nevertheless, the protein reached the nucleation zone and started forming crystalline structures, which served as a good starting point for further optimization.



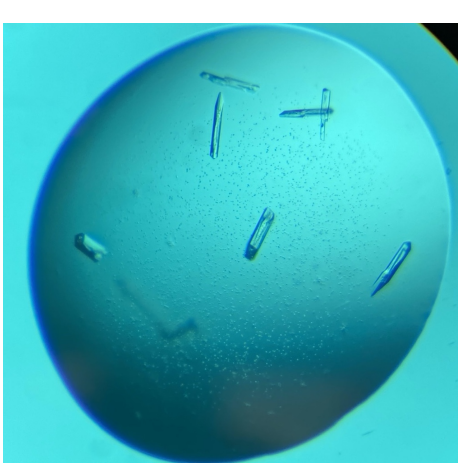
(a) LFS 1-23 (b) LFS 2-18

Figure 4.1: Two drops from the ligand-friendly screen. The protein formed crystalline structures but not solid, high-quality crystals.

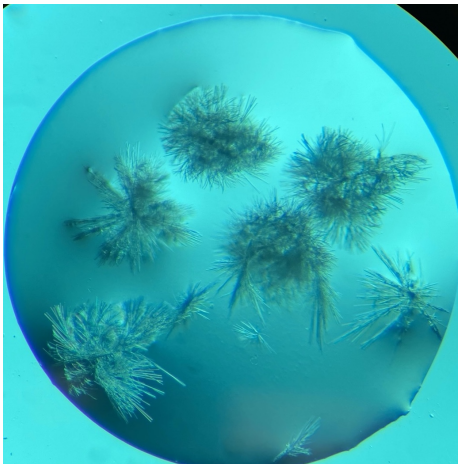
Table 4.1: Summary of crystallization conditions that yielded hits in the initial screen.

Screen kit	**Tube**	**Precipitant 1**	**Precipitant 2**	**Buffer**	**pH**	**Salt**
SG1	1-14	12 % w/v PEG 20 000	None	0.1 M MES	6.5	None
SG1	1-19	20 % w/v PEG 3350	None	None	-	0.2 M Sodium formate
SG1	1-24	20 % w/v PEG 3350	None	None	-	0.2 M Magnesium formate dihydrate
SG1	2-37	20 % w/v PEG 3350	None	None	-	0.2 M Ammonium formate
LFS	1-20	20 % w/v PEG 6000	10 % v/v Ethylene glycol	0.1 M MES	6	0.2 M Ammonium chloride
LFS	1-22	20 % w/v PEG 6000	10 % v/v Ethylene glycol	0.1 M MES	6	0.1 M Magnesium chloride hexahydrate
LFS	1-23	20 % w/v PEG 6000	10 % v/v Ethylene glycol	0.1 M MES	6	0.1 M Calcium chloride dihydrate
LFS	1-33	20 % w/v PEG 6000	10 % v/v Ethylene glycol	0.1 M HEPES	7	0.2 M Lithium chloride
LFS	2-14	20 % w/v PEG 3350	10 % v/v Ethylene glycol	0.1 M Bis Tris Propane	6.5	0.2 M Sodium bromide
LFS	2-18	20 % w/v PEG 3350	10 % v/v Ethylene glycol	0.1 M Bis Tris Propane	6.5	0.2 M Sodium formate

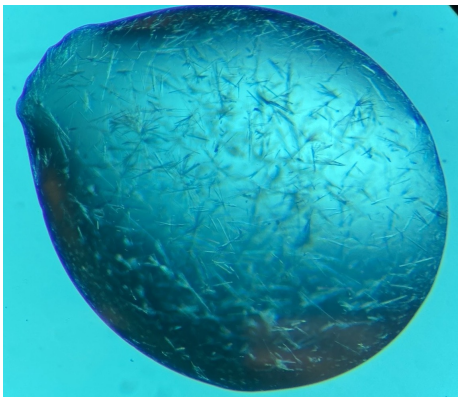
One condition, SG1-14, generated thick, rod-like 3D crystals and can be seen in Figure 4.2a. Here, the crystals appear to have grown in all directions after initiating nucleation. This condition served as a primary lead for further optimization. The components in this cocktail were 12 % w/v PEG 20 000 + 0.1 M MES buffer pH 6.5.



(a) SG1 1-14



(b) SG1 1-19



(c) SG1 1-24

Figure 4.2: Three drops from the ShotGun screen. The quality of the crystals are evidently different in these drops, with thick rod-like 3D structures in (a) to thin needle-shaped crystals in (b) and (c).

The other two images in Figure 4.2 are, again, examples of crystals of poorer quality. However, they are still considered hits and served as starting points for further optimization.

4.2 Thermal shift sssay

The stability investigation of the nsP2 in different buffer solutions was performed in a thermal shift assay, which generated two sets of plots, the fluorescence intensity as a function of temperature, and the negative derivative of fluorescence as a function of temperature. From these plots, the protein's melting temperature in different buffers (T_m) could be estimated. The fluorescence vs temperature plot, seen in Figure 4.3, shows the fluorescence intensity of all wells. The ideal shape of a melting curve is sigmoidal, with little or no fluorescence intensity before protein unfolding is initiated,

followed by a sharp transition, as demonstrated in Figure 2.5. Most of the wells follow the sigmoidal shape but approximately 1/3 of the wells have initial fluorescence intensities above 4000 units, indicating that the protein is already unfolded or partly unfolded at the beginning of the experiment. Evidently, no meaningful estimation of T_m could be done for these conditions. However, it could imply that they made the protein relatively unstable even at lower temperatures. The buffer solutions yielding fluorescence intensities above 4000 units, were 50 mM HEPES + 200 - 0 mM NaCl and 50 mM HEPES pH 6.5 - 8.

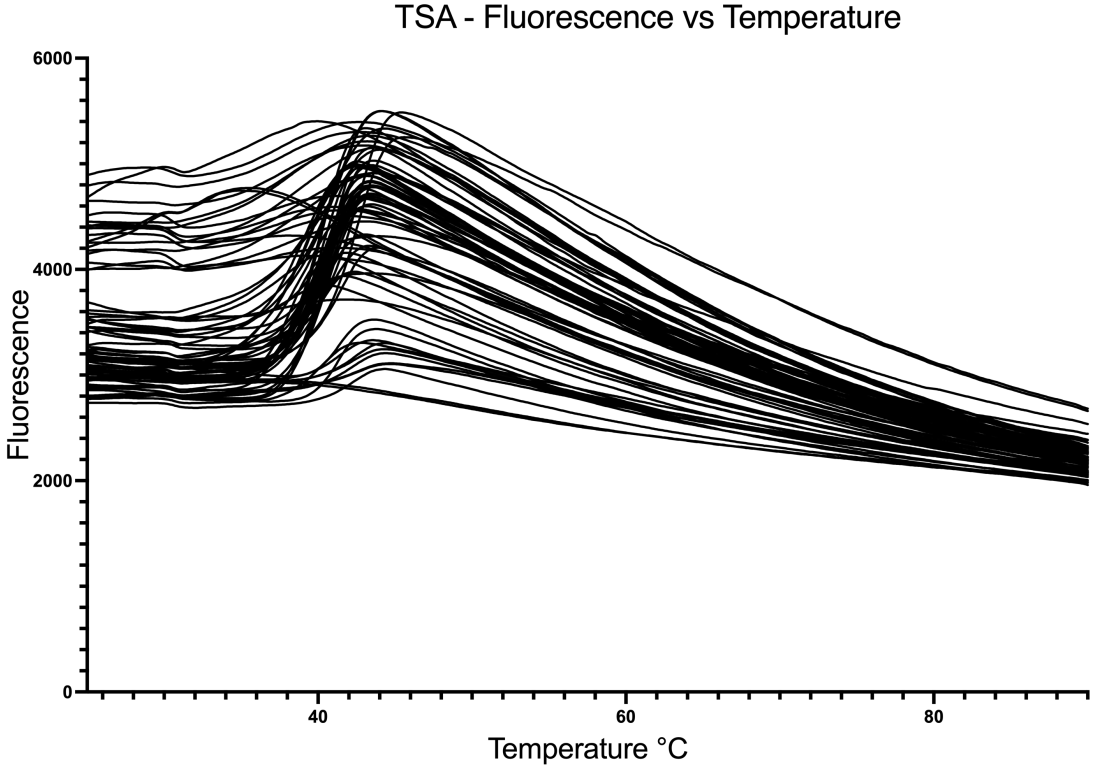


Figure 4.3: All wells are plotted (fluorescence vs temperature). Approximately one third of all wells emitted fluorescence intensities above 4000 units, indicating that the protein is already unfolded or partly unfolded at the beginning of the experiment. Plot generated in GraphPad.

By plotting the negative first derivative of the fluorescence as a function of temperature, an easier estimation of the T_m was obtained. Here, the T_m corresponds to the temperature at the lowest -dF/dT value of the curve, compared to the midpoint of the transition curve in Figure 4.3. In Figure 4.4, the negative first derivative of all wells was plotted vs temperature, generating distinct negative peaks. Most curves peak at roughly the same temperature, approximately 40 °C.

In Figure 4.5, only a few selected curves were plotted. These were curves with some of the lowest peak values, but the plot also includes one curve with a high peak value for reference. The red peak, corresponding to 50 mM HEPES + 0 mM NaCl, had an initial fluorescence intensity above 4000 units in Figure 4.3, and when studying its behavior in Figure 4.5, it does not have a defined negative peak, as expected. The other curves in Figure 4.5 show low peak values, and the curve corresponding to 50 mM sodium phosphate peaks at approximately 43 °C, nearly 3 °C higher than the

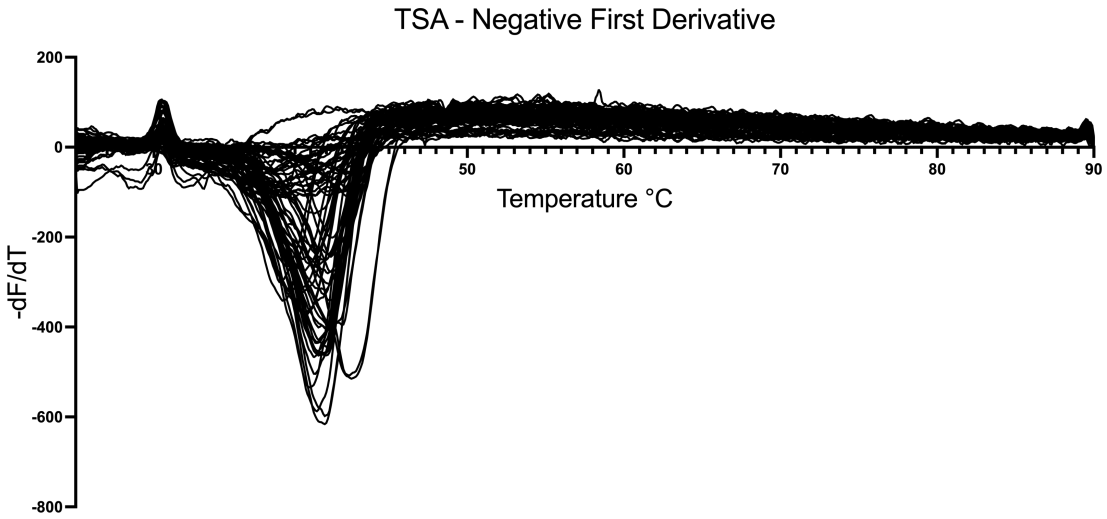


Figure 4.4: The negative first derivative of all wells are plotted against temperature. Distinct negative peaks are noticeable. T_m is equal to the the lowest -dF/dT value of the curves. Plot generated in GraphPad.

majority of the other peaks. This is a substantial increase in T_m, indicating that sodium phosphate buffer increases the stability of the nsP2.

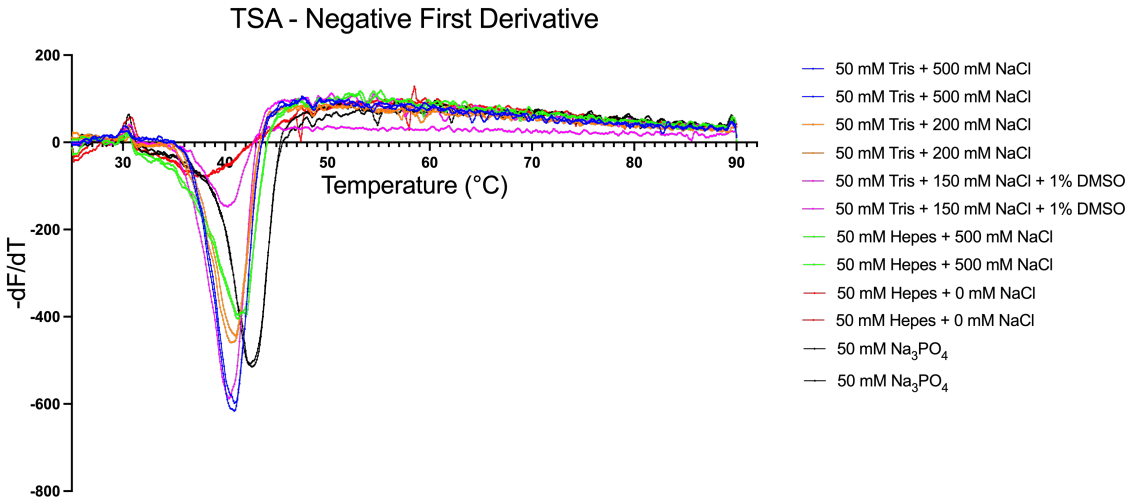


Figure 4.5: The negative first derivative of selected wells, plotted against temperature. The buffer that increased the stability of the nsP2 the most was 50 mM sodium phosphate (black line). Plot generated in GraphPad.

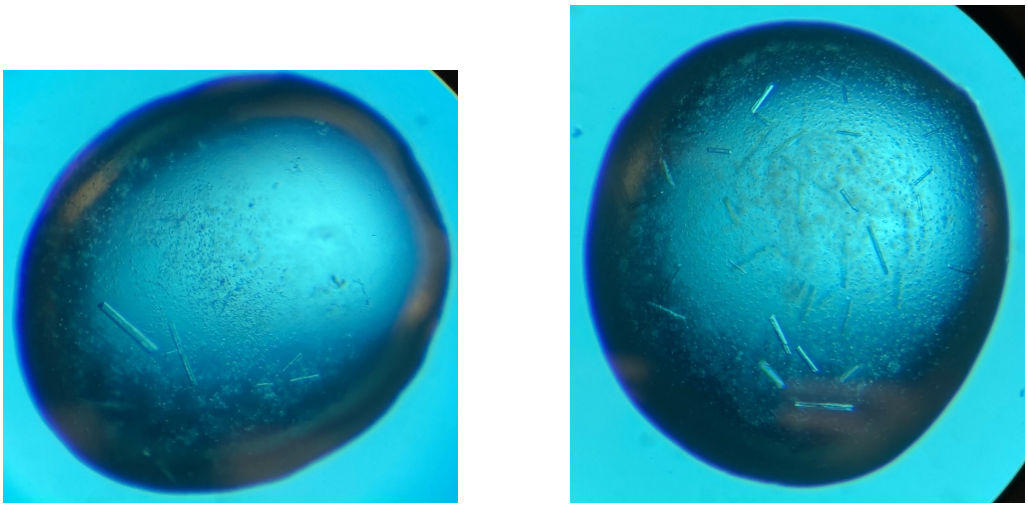
4.3 Optimizing crystallization conditions

Investigating in what crystallization conditions the nsP2 forms crystals, repeatedly, has been the main objective of this project. Finding one, or a few, cocktail conditions that produce high-quality crystals allows for a straightforward process of putting up many drops resulting in numerous crystals that can be soaked with a chosen inhibitor.

Based on the hits from the initial screen, a total of 864 drops were set up during the optimization process. As some of the chemicals that were used in the initial screen were not in storage at FOI or came from different manufacturers, some limitations and adjustments to the crystallization cocktails had to be done. They were therefore comprised of the chemicals available at FOI, except from PEG 20 000 which was ordered (Thermo Scientific).

Initial verification of the quality of the crystals was done throughout the optimization process by studying the drops in a light microscope. It took approximately three to six days for the crystals to grow, hence, if there was no sign of crystal growth after one week, that condition was assumed to negatively influence crystal growth. Correspondingly, the conditions that positively influenced crystal growth after approximately one week, were used to design new plates with that condition as a lead.

The result from the TSA performed in section 3.4, indicated that sodium phosphate, instead of tris, in the storage buffer, would keep the protein in a more stable conformation. Therefore, a plate with 48 drops, 24 with nsP2 from the sodium phosphate buffer and the other 24 from the tris buffer, was set up with 0.1 M MES + 10 - 14 % PEG 20000. Ten out of the 24 drops with buffer containing sodium phosphate, produced high-quality crystals when observed in the microscope. These drops were later chosen to be soaked with probes. Two of these drops can be seen in Figure 4.6 below.



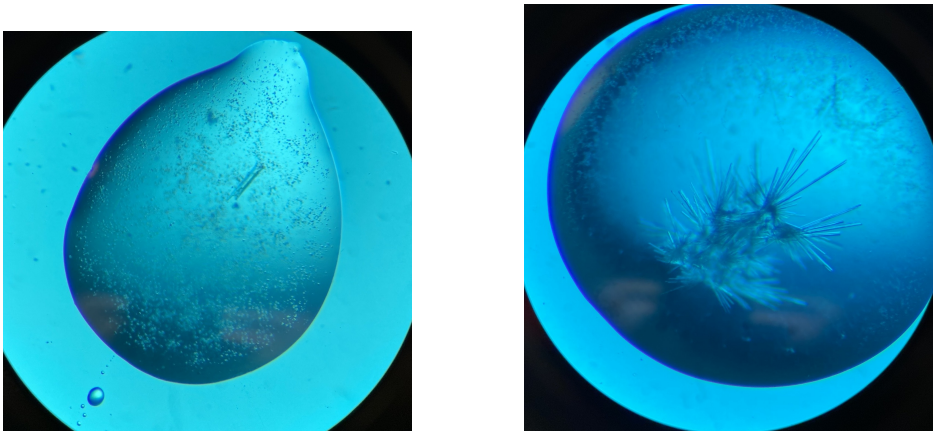
(a) (b)

Figure 4.6: Drops containing crystals, with sodium phosphate as storage buffer. Images taken with a mobile phone.

The drop in Figure 4.6a contains one long, thick, rod-shaped crystal similar to the crystals in Figure 4.2a. It is even possible to see that the tip of the rod is shaped like a triangle, which indicates one solid crystal instead of several thinner layers stacked on top of each other. The crystals in Figure 4.6b are somewhat smaller than the crystal in Figure 4.6a, but they are, nevertheless, rod-shaped which suggests high quality crystals. Unfortunately, the drops with Tris as storage buffer did not produce any high-quality crystals. A potential explanation could be the lack of 10 % ethylene glycol in the crystallization cocktail. However, the drops with sodium phosphate in

the storage buffer formed crystals, despite the lack of 10 % ethylene glycol.

Throughout the optimization process, three more plates generated crystals that were considered suitable to eventually be soaked with inhibitor. Figure 4.7 showcases the varying crystal qualities of these drops, where the drop in Figure 4.7a contains one thick, solid crystal and the drop in Figure 4.7b contains a crystal looking like a stack of needles growing in all directions. There was a possibility that the ends of these "needles" were good enough to diffract, and so, they were also chosen to eventually be soaked.



(a) (b)

Figure 4.7: Drops with crystals of varying quality chosen to be soaked. (a) containing 7.5 mg/ml nsP2 (sodium phosphate buffer) + 0.1 M MES pH 5.5 + 10% PEG 20000 and (b) 15 mg/ml nsP2 (tris buffer) + 0.1 Bicine pH 8 + 18 % PEG 3350. Images taken with a mobile phone.

The methodology of this optimization process was, in essence, trial and error based. Several crystallization conditions were tested, evaluated and improved, again and again, to obtain crystals of high quality. Some of the many challenges of this method were the fragility of the crystallization process and the difficulty to account for, and fine-tuning, all parameters involved. For example, PEG is a very viscous liquid and it is therefore extremely difficult to pipette exact amounts when doing it by hand. The uncertainty of pipetting was a general error occurring with all chemicals, but even more so with viscous liquids such as PEG and glycerol. Moreover, in some cases, there was not enough grease on top of the wells of the plates which meant that the coverslip was not sealed properly and air was let in. As a consequence, some drops were completely dried out and some formed a skin-like layer on the surface of the drop. This undoubtedly affects crystallization formation, as the drop and crystallization cocktail no longer exist in an enclosed system. Furthermore, temperature, humidity and vibrations to plate also affect crystallization.

The freshness of the protein was also a major influential parameter. The first 17 plates (out of 24) were set up with protein from batch one. A portion of that protein was stored on ice and the rest was frozen at - 80°C. There was a notable aggravation in crystal formation when the protein had been stored on ice for too long, approximately two to three weeks. An adequate amount of protein was therefore thawed from the -80°C freezer before setting up new drops. A second batch was

purified to be used for the remaining optimization and co-crystallization process.

Out of the 24 plates that were set up with drops during this optimization process, four had drops containing crystals selected for soaking. The crystallization conditions of these drops have already been summarised in Table 3.2.

4.4 Co-crystallization

A total of ten plates were set up during the co-crystallization optimization process. Six of those generated drops with crystals, however, on average, only two drops per plate generated crystals with high enough quality to be suitable for x-ray crystallography. The majority of these crystals were needle-shaped, one example is shown in Figure 4.8 below.

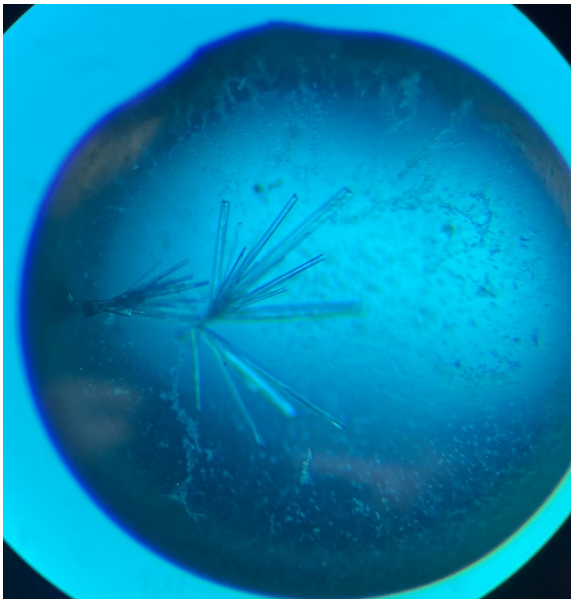


Figure 4.8: A needle-shaped crystal grown together with probe DW0406 (crystallization condition: 15 mg/ml nsP2 (Tris) + 18% PEG 3350 + 0.1 M Bicine). Image taken with a mobile phone.

The quality of the crystals was overall worse than the quality of the crystals produced in section 3.5, in the sense that no rod-shaped crystals were generated. It is feasible to believe that the addition of a ligand in the active site alters the crystal packing, i.e. how the atoms/molecules arrange themselves in a crystal. As a consequence, it could potentially be even more difficult to generate high-quality crystals with a bound ligand in the active site.

The LFS was used to screen for conditions that positively influenced crystal formation with the nsP2 together with a probe, which resulted in one hit that acted as a lead for further optimization. However, only one probe, DW0408, was used in this screen due to time limitations. Because the three probes are relatively similar in chemical composition, it was assumed that the result from the screen would be the similar regardless of which of the three probes was used. However, for future x-ray crystallography experiments, each probe should be tested and screened individually to find the optimal crystallization conditions.

The co-crystallization conditions that generated crystals chosen for x-ray crystallography are listed in Table 4.2.

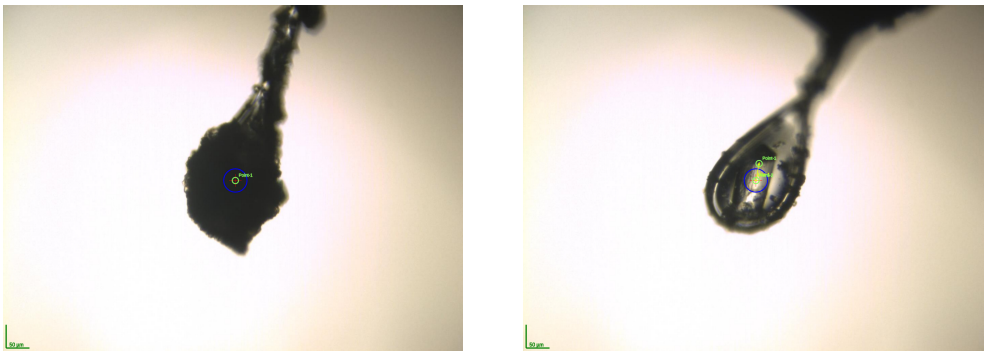
Table 4.2: Summary of co-crystallization drops chosen for freezing

[nsP2]	Batch	Storage buffer	Precipitant	Buffer	pH	Salt	Probe
15 mg/ml	2	Tris	14 % w/v PEG 3350	0.1 M Bis-Tris propane	7	0.2 M Ammonium formate	DW0406
15 mg/ml	2	Tris	16 % w/v PEG 3350	0.1 M Bis-Tris propane	7	0.2 M Sodium formate	DW0411
15 mg/ml	2	Tris	15 % w/v PEG 3350	0.1 M Bis-Tris propane	7	0.2 M Sodium formate	DW0411
15 mg/ml	2	Tris	18 % w/v PEG 3350	0.1 M Bicine	8.5	-	DW0406
15 mg/ml	2	Tris	20 % w/v PEG 3350	0.1 M Bicine	8.5	-	DW0406

4.5 X-ray crystallography

75 crystals were sent to BioMAX in Lund. 73 data sets were collected, 13 of which had promising data sets (promising statistics) and five of these seemed to have extra electron density in the active site.

During data collection, when the position of where the beam would hit the crystal was to be decided, one unexpected problem arose. In the majority of cases, the crystal in the loop was not visible. In fact, the whole loop was covered with a black, clutter-like substance making it impossible to locate the crystal. In Figure 4.9a, one example of this problem is shown. A single point in the middle of the loop was chosen, as the best guess of where the crystal could be located. In Figure 4.9b, the crystal is clearly visible in the loop and two positions on the crystal were chosen for a helical scan.



(a) (b)

Figure 4.9: Loops with crystals in camera view during data collection. The loop and crystal is not visible behind the black clutter in (a), but clearly visible in (b). Green scale bar = 50 μm.

The formation of black clutter covering a majority of the loops, indisputably, negatively affected the outcome of the collected data. Without the ability to locate the crystal and decide where to position the beam, most times the crystal was missed, or only portions of the crystal were hit with the beam, resulting in data sets with very few or no reflections. This is why only 13 out of the 75 crystals produced promising statistics. It is poorly understood why some loops were completely covered and some were not. An estimated guess is that there could have been flakes of clutter or junk in the liquid nitrogen bath container that got stuck to the loop as it was submerged in the nitrogen during freezing (section 3.8).

4.5.1 Refinement and model building

Out of the five data sets that had promising statistics and showed extra electron density in the active site, one was chosen for further data processing and refinement. This data set came from a crystal generated in a drop from the SG1 screen, tube 1-14, as seen in Figure 4.2a and soaked with DW0411.

After several refinement and model building cycles, a relatively good model of the nsP2 protease was obtained. The resolution of this data set was chosen to 1.4 Å

which provided enough details in the electron density map to visualize the rings in the aromatic amino acids, see Figure 4.10.

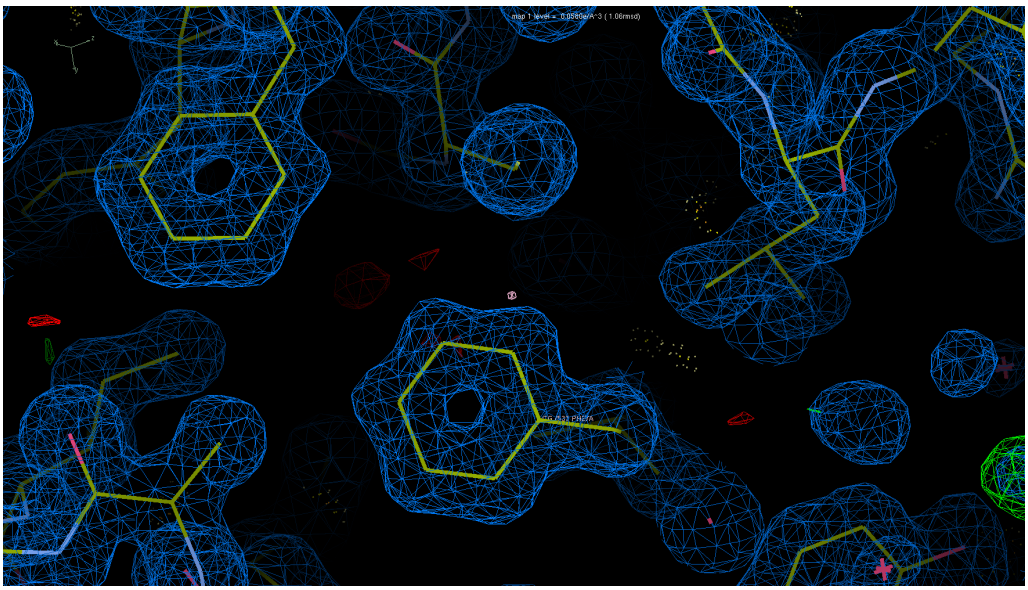


Figure 4.10: Details of the electron density map surrounding the structure model. The blue grid, i.e. the electron density map, represents areas where the model has the right amount of electrons present compared to the experimental data. The resolution of 1.4 Å provides enough details to almost see circles around each atom and holes in the middle of the aromatic rings. Visualized in Coot.

In Figure 4.11, the nsP2 protease's active site is in focus where the cysteine, one of the two important amino acids of the catalytic diad, and its sulfur atom (depicted as bright yellow at the end of the stick model) are visible. The blue grid represents areas where the model has the right amount of electrons present compared to the experimental data (the electron density). The green grid represents areas where the program has calculated there is a deficiency in electron density, and the red grid represents areas where the electron density is too high.

The electron density map continues out from the sulfur atom in Figure 4.11 to then be surrounded by a green grid, indicating there is a deficiency in electron density. This suggests that something has bound to the sulfur atom when the data was collected, and because the stick model only includes the amino acids in the nsP2 protease and not a bound probe, the program highlights this area in green. The red area surrounding the sulfur atom, indicates that this area has too many electrons. This could indicate a covalent bond between this atom and potentially an atom on the probe, as a covalent bond involves sharing of electron pairs between two atoms and a decrease in electron density in that area.

The building of the probe to the structure model of the nsP2 protease is outside the scope of this project. However, one brief attempt was made which can be seen in Figure 4.12. The dashed line represents a covalent bond between the sulfur atom of the cysteine and the imine carbon atom of the probe. There are several red areas surrounding the probe, which indicates that the experimental data from the reflections do not support the placement of the probe. This is by no means a finished model, but rather an initial suggestion of where the probe could be positioned in the

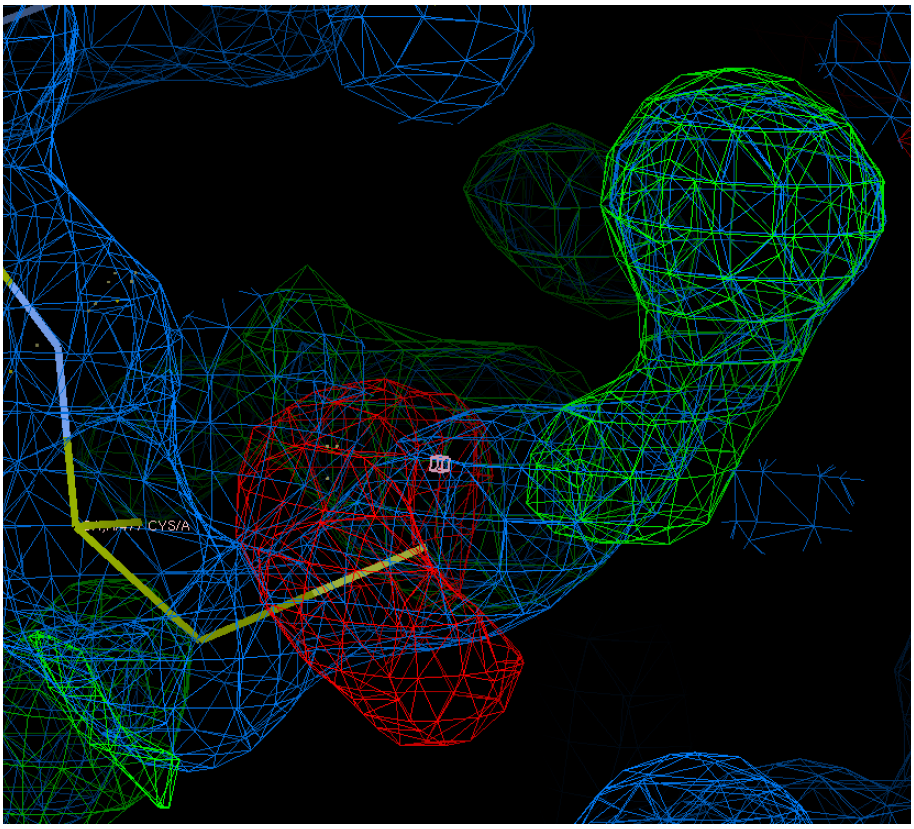


Figure 4.11: Details of the nsP2 protease's active site. The yellow sticks represents the amino acid cysteine, with its sulfur atom at the end (in bright yellow). The green grid represents an area where there are not enough electrons present and could indicate a bond to a molecule. Visualised in Coot.

active site.

A substantial amount of time, effort and skill are required to obtain a good model, supported by statistics, with the probe bound to the active site. The structure model and the placement of the probe presented in this thesis are far from ideal. This is evident when studying the refinement statistics for the nsP2 protease bound to the probe presented in Table 4.3. The B factor for the ligand reaches almost a value of 90 Å^2 which indicates high uncertainty of its position. The clashscore value without the probe bound to the active site was 0.59 (see Table B.1) but reached a value of 2.72 with the probe, again indicating the unfavorable position of the probe. However, the statistics for the data collection (before refinement) indicate high-quality data. The mean I/sigma(I) for the outer shell is just over six and completeness is near 100 %, which supports the chosen resolution of 1.4 Å.

All of the 13 data sets that had promising statistics, also turned out to have high-quality data resulting in high resolution below 2 Å. This confirms that the initial screening for high-quality crystals, performed throughout the optimization process when the crystals were studied in the light microscope, were in fact, crystals of high quality. The conditions that were developed throughout the optimization process generated crystals of high quality which resulted in excellent crystallographic data. Only one of the 13 data sets came from a crystal generated by co-crystallization, the other came from soaked crystals.

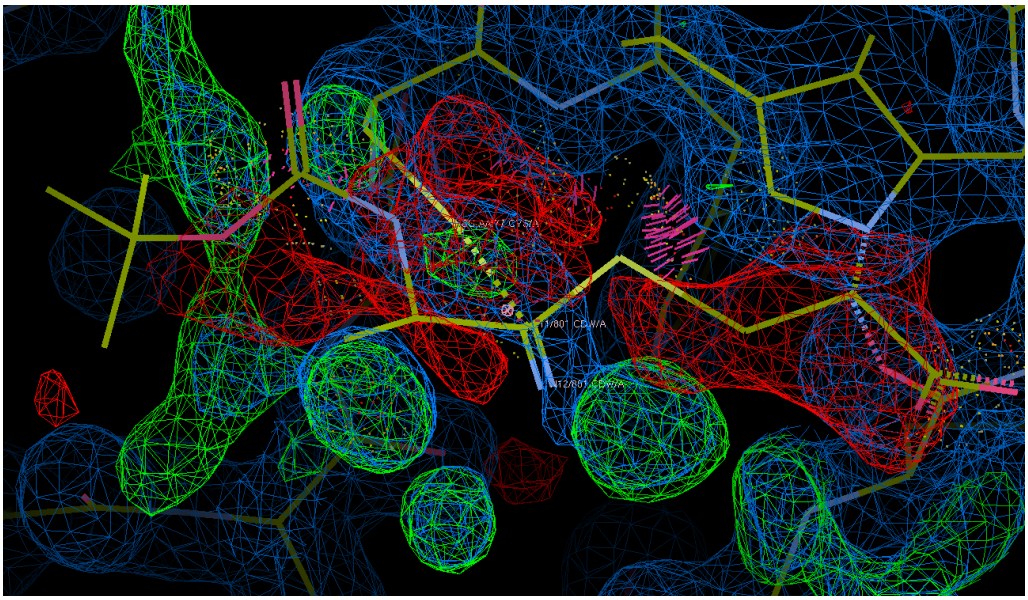


Figure 4.12: Details of the nsP2 protease's active site with the probe (DW04011) built in to the model. The dashed line represents a covalent bond between the sulfur atom (bright yellow) of the cysteine and the imine carbon atom (yellow) of the probe. Visualised in Coot.

The other four data sets that indicated extra electron density in the active site, were all crystals that were soaked with inhibitor. If further refinement cycles and model building with a probe shows that the extra density is there because the thiol had formed a covalent bond to a probe, then the soaking procedure will be considered a success. However, if the statistics do not support the placement of a probe, but rather a connection to a water, PEG or sodium phosphate molecule, then the soaking will have to be considered unsuccessful.

Table 4.3: Data collection and refinement statistics for the nsP2 protease bound to DW0411. Refinement statistics for just the nsP2 protease are presented in section B, Table B.1. Statistics for the highest-resolution shell are shown in parentheses.

Data collection	
Wavelength (Å)	0.976
Resolution range (Å)	51.05 - 1.4 (1.45 - 1.4)
Space group	P 21 21 21
Unit cell dimensions (Å/degrees)	60.888 63.361 86.199 / 90 90 90
Total reflections	833340 (56412)
Unique reflections	66334 (6562)
Multiplicity	12.6 (8.6)
Completeness (%)	99.96 (99.98)
Mean I/sigma(I)	29.50 (6.09)
Wilson B-factor	14.97
R_{merge}	0.048 (0.3523)
R_{meas}	0.05005 (0.3745)
R_{pim}	0.014 (0.125)
CC1/2	0.999 (0.97)
CC*	1 (0.992)
Refinement	
Reflections used in refinement	66319 (6561)
Reflections used for R_{free}	3297 (326)
R_{work}	0.1781 (0.2127)
R_{free}	0.1983 (0.2379)
CC(work)	0.963 (0.924)
CC(free)	0.952 (0.886)
Number of non-hydrogen atoms	2905
Number of macromolecules	2565
Number of ligands	19
Number of solvent	321
Protein residues	320
RMS - bond lengths (Å)	0.006
RMS - bond angles (degrees)	0.81
Ramachandran favored (%)	96.54
Ramachandran allowed (%)	3.46
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.00
Clashscore	2.72
Average B-factor (Å^2)	22.18
B-factor (macromolecules) (Å^2)	20.37
B-factor (ligands) (Å^2)	88.94
B-factor (solvent) (Å^2)	32.73
Number of TLS groups	7

4.6 Gel-based assay

The theory and practicality of this assay are quite straightforward, however, after the fifth time attempting this experiment, no reliable results have been obtained. In the first two attempts, there were problems with the equipment and the gels had to be discarded. One gel had to be run to investigate at what nsP2 concentration the bands would be visible, as they were very weak in the discarded gels. Based on that gel, 5μM of nsP2 was required.

In the fourth attempt, two gels were run in accordance with the method described in section 3.10, with an nsP2 concentration of 5μM and the probe DW0408. The result of the first gel, which was run after four hours of incubation time, can be seen in Figure 4.13. The bands are weak but it is possible to see the uncut substrate in well one and two. However, it is important to note that the nsP2 was added to the tube (corresponding to well number 2) with the substrate shortly before the gel was run. Thus, that tube was only incubated for approximately 15 min, while the tubes corresponding to wells 3 to 9 were incubated for four hours.

When comparing the uncut substrate band at 58 kDa in wells 3 to 9 in Figure 4.13, it seems to become less and less visible, indicating that at lower concentrations of probe, more substrate is cleaved. The uncut substrate band in well 3 seems to be nearly as thick and visible as in wells 1 and 2, indicating that a probe concentration of 2,5 μM inhibits the nsP2 protease from cleaving its substrate. Even though the bands are quite weak, it is still possible to acknowledge the dose-dependent manner in which the probe inhibits the nsP2's proteolytic effect.

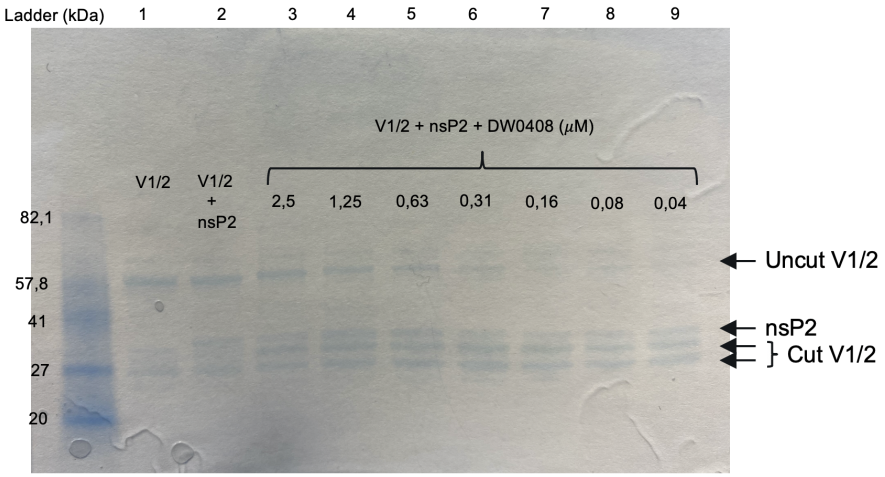


Figure 4.13: Gel after four hours of incubation time. As the probe concentration decreases, more substrate is cut. The uncut substrate is visible at 58 kDa, the nsP2 at 38.3 kDa, and the two substrate products, CFP and YFP, at 30.9 KDa and 27.4 kDa respectively.

A second gel was run after 24 hours which can be seen in Figure 4.14. This time, the gel was stained twice which generated more visible bands. Unfortunately, the dose-dependent manner demonstrated in Figure 4.13 is not displayed here. More alarmingly, the substrate, which is supposed to be uncut in well one, due to the absence of enzyme, seems to have broken down by itself. It appears that after a

notably long time, the substrate becomes unstable and breaks down, which is likely to have happened in all wells. This makes it difficult to draw any conclusions because the substrate will break down into its cleavage products regardless of how effectively the probe inhibits the nsP2 protease.

There ought to be two cleavage products in Figure 4.14, at 27.4 and 30.9 kDa in wells one and two. But only one band at 27.4 kDa is visible which is a mystery.

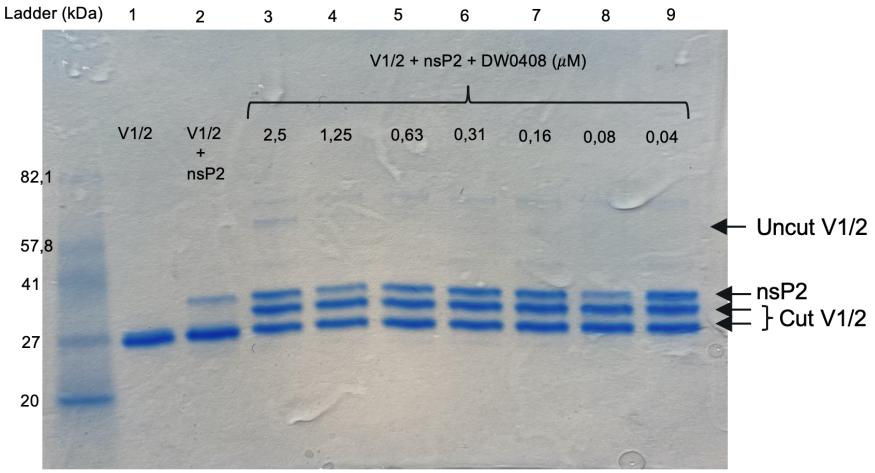


Figure 4.14: After 24 hours of incubation time, the substrate in well 1 has broken down. Only one band in well 1 is visible at 27.4 kDa.

In the fifth attempt, all three probes were used in separate experiments resulting in six gels. The results from the six gels were pretty much the same without any considerable differences. Figure 4.15 and 4.16 show the results from the gels containing probe DW0408. In both gels, the uncut substrate band is visible in well 1 at 58.3 kDa and in all other wells, three bands are visible at 27.4, 30.9 and 38.3 kDa. No other bands are visible, indicating that the substrate was cleaved in all wells, except in well 1.

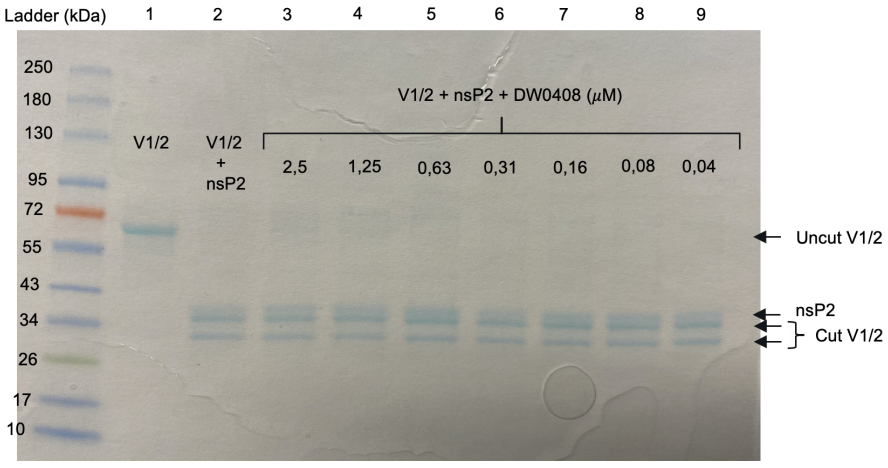


Figure 4.15: No dose-dependent manner is noticeable in this gel. The substrate is still intact in well 1 after 4 hours of incubation time.

As the substrate had broken down by itself in the previous attempt (Figure 4.14), it could also have happened here in Figure 4.15, explaining why there is no detectable dose-dependent manner. However, the band in well 1 discards that hypothesis, as it is clearly intact in its uncut form. In Figure 4.13, 2.5 μM of probe was enough to inhibit nsP2, resulting in a visible uncut substrate. That probe concentration was evidently not enough to inhibit the nsP2 in this case. The band in well one in Figure 4.16 representing the uncut substrate at 58.3 kDa, shows that it was stable enough to remain in its uncut form even after 24 hours.

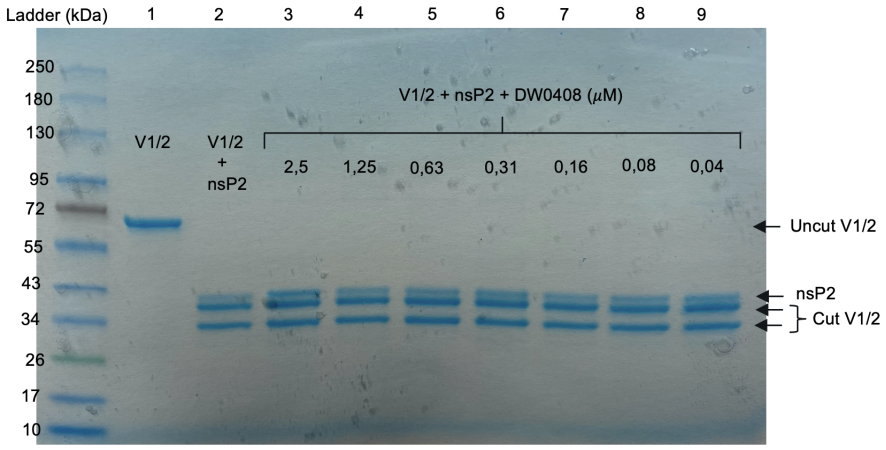


Figure 4.16: No dose-dependent manner is noticeable in this gel. The substrate is still intact in well 1, even after 24 hours of incubation time.

The main uncertainty in this study was the potential instability of the substrate. To investigate whether or not the nsP2 is inhibited by the probes in a dose-dependent manner, a reliable and stable substrate is a necessity. The last six gels that were run showed no difference in inhibition dependent on dose. More disappointingly, they did not show a tendency for inhibition at all, which proposes that the probe had very little or no affinity to the nsP2. Further development and refinements to this method are required to obtain reliable results.

4.7 Fret assay

The amount of cleaved substrate (nmol) for each substrate concentration was plotted against time (min) in Figure 4.17, only including the linear range T = 2 - 16 min. The values at T = 0 were abnormally high due to the movement of the plate when it was placed in the reader and thus excluded. The initial velocity values range from 0.00963 - 0.06751 nmol/min, see the legends in Figure 4.17 for exact values.

In Figure 4.18, the initial velocities were plotted against the substrate concentration and fitted to the Michaelis-Menten equation from which K_M and V_{max} were calculated and presented by the software program. k_{cat} was calculated by dividing V_{max} with the nsP2 concentration present in the wells. The values of the kinetic parameters are presented in table 4.4.

As stated in section 3.11, the assay was performed multiple times due to untrustworthy data of the initial experiments. Consequently, there was only enough substrate to

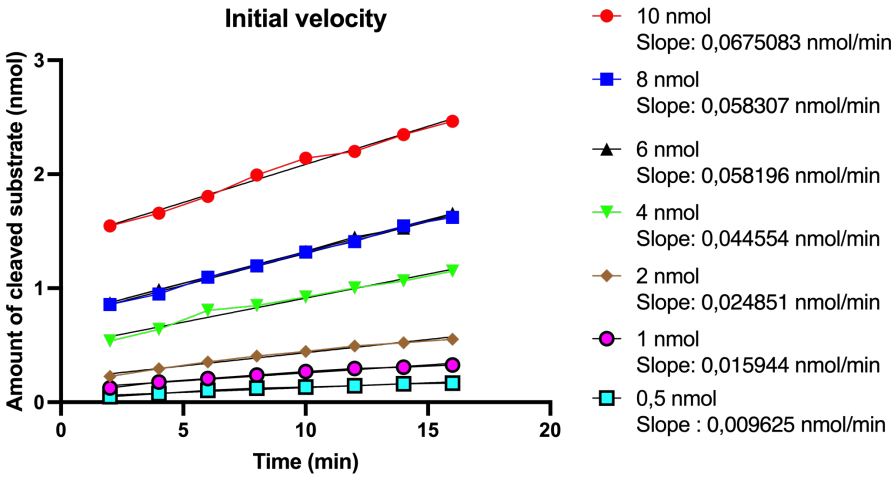


Figure 4.17: The initial velocities for each substrate concentration was obtained by fitting a linear regression line and calculate the slope value.

perform this assay in duplicates. The results from this assay are therefore not as statistically significant as an assay performed in triplicates. Nevertheless, the methodology of this assay is a starting point to developing and improving further assays to determine the nsP2's proteolytic ability. But also, perform assays with the probes present to obtain IC_{50} values.

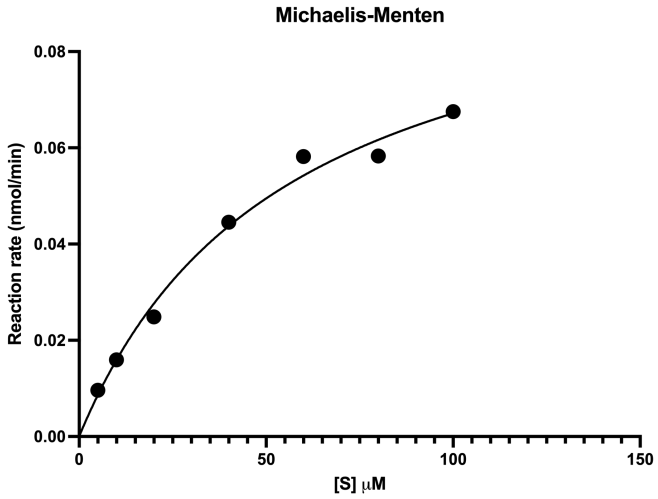


Figure 4.18: To obtain values of K_M and V_{max}, the initial velocities were plotted against the substrate concentration and fitted to the Michaelis-Menten equation.

Table 4.4: Summary of kinetic parameters

V_{max}	K_M	k_{cat}
0,1046 nmol/min	55,71 μmol	0,35 min^{-1}

Chapter 5

Conclusions

Throughout this project, hundreds of drops with varying crystallization conditions were set up in the hopes of finding conditions that would generate high-quality crystals of the nsP2 protease. Several conditions were found to produce such crystals, which were also confirmed by x-ray crystallographic data statistics.

It can be concluded that the nsP2 protease is occupying a more stable confirmation when stored in a buffer containing sodium phosphate, rather than Tris. Overall, the drops containing the nsP2 protease in sodium phosphate buffer generated crystals more frequently than the drops with the nsP2 in Tris buffer, and with higher quality. A fresh batch of protein also generated better crystals compared to a protein that had been stored on ice for a long period of time, approximately two to three weeks.

Second, crystals generated from co-crystallization had in general, lower quality than crystals grown without a probe. The needle-shaped structures, growing in all directions appeared to be thinner and more fragile than the thick, rod-shaped crystals that were obtained without a probe. The x-ray crystallographic data statistics showed that the crystals, grown without a probe were, in fact of high quality and produced data that allowed high resolution below 2 Å.

Third, a tendency of the inhibitory effect of probe DW0408 could be detectable in one gel in the gel-based assay, when a protein concentration of 5 μM was used. This indicates that the mechanism of action of this probe, to an extent, does form a covalent bond, and thus, inhibits the proteolytic effect of the nsP2 protease. However, due to the inability of reproducing these results, they should not be overstated.

Finally, the trial and error-based method of finding crystallization conditions proved to be successful and generated several crystals of high quality. If the soaking method is successful, these conditions will provide FOI with a reproducible protocol to generate more crystals of the nsP2 protease-inhibitor complex, which will aid their pursuit of designing inhibitors based on structure.

5.1 Future work

During the optimization process, adjustments to crystallization conditions had to be made as most of the chemicals at FOI came from different manufacturers, or were not in storage. For further optimization processes or investigations, the exact chemicals

that were used in the commercial screens could be ordered and used. Perhaps the addition of the second additive, 10% ethylene glycol, could generate high-quality crystals with or without sodium phosphate in the storage buffer.

If the soaking method is proven to be unsuccessful for the nsP2 protease, a more extensive co-crystallization optimization process for each of the three probes ought to be performed. Although it will require a large amount of protein and chemicals for the crystallization conditions, it could be proven fruitful if the right conditions are found.

One of the most urgent matters to address for future x-ray crystallography experiments is the formation of black clutter surrounding the loops during freezing. Without the ability to see the crystals in the loops, it is near impossible to locate a position on the crystal for the beam to hit, and no useful data will be obtained. Perhaps new improved cryoprotectants can be developed, as well as investigating the tanks and containers where the liquid nitrogen is stored. The success rate will undoubtedly increase if this problem is addressed and solved.

One structure model has been presented in this thesis. The other four data sets with extra electron density in the active site, need to be processed and refined to obtain more information about the probes' binding to the active site. When all five structure models have been generated, they might provide information that together, gives clues about the binding interactions between the probes and the nsP2 protease.

The gel-based assay should be further developed to investigate the probes' ability to inhibit the nsP2's proteolytic effect. One major problem with this assay was the instability of the substrate. Even so, in some of the gels, the substrate did not seem to break down by itself as it was intact in the first well, but there was no visible inhibitory effect at all in the other wells with the substrate, nsP2, and probe. If this is due to the probes' low affinity to the nsP2, or if the substrate broke down on its own, will not be clarified unless new experiments are performed with a stable and reliable substrate.

Further FRET assays need to be conducted in order to obtain reliable values of the nsP2's steady-state kinetic parameters, and they ought to be performed in triplicates. The methodology presented in this thesis, acts as a starting point for further development, and eventually, expanding the assay to include the determination of IC_{50} values of the probes.

Bibliography

[1] Farhana Abu Bakar and Lisa FP Ng. "Nonstructural proteins of alphavirus—potential targets for drug development". In: *Viruses* 10.2 (2018), p. 71.

[2] Tsutomu Arakawa and Serge N Timasheff. "Mechanism of protein salting in and salting out by divalent cation salts: balance between hydration and salt binding". In: *Biochemistry* 23.25 (1984), pp. 5912–5923.

[3] Valentina Arkhipova, Albert Guskov, and Dirk-Jan Slotboom. "Analysis of the quality of crystallographic data and the limitations of structural models". In: *Journal of General Physiology* 149.12 (2017), pp. 1091–1103.

[4] Jeremy M. Berg et al. *Biochemistry*. 8th ed. W. H. Freeman and Company, 2015, pp. 69–70.

[5] Jeremy M. Berg et al. *Biochemistry*. 8th ed. W. H. Freeman and Company, 2015, pp. 71–73.

[6] Jeremy M. Berg et al. *Biochemistry*. 8th ed. W. H. Freeman and Company, 2015, pp. 225–230.

[7] Nishank Bhalla et al. "Host translation shutoff mediated by non-structural protein 2 is a critical factor in the antiviral state resistance of Venezuelan equine encephalitis virus". In: *Virology* 496 (2016), pp. 147–165.

[8] Aleksandar Bijelic and Annette Rompel. "Polyoxometalates: more than a phasing tool in protein crystallography". In: *ChemTexts* 4.3 (2018), pp. 1–27.

[9] Michael S Bronze et al. "Viral agents as biological weapons and agents of bioterrorism". In: *The American journal of the medical sciences* 323.6 (2002), pp. 316–325.

[10] Joshua A Broussard et al. "Fluorescence resonance energy transfer microscopy as demonstrated by measuring the activation of the serine/threonine kinase Akt". In: *Nature protocols* 8.2 (2013), p. 265.

[11] Javier Campos-Gomez et al. "A novel cell-based assay to measure activity of Venezuelan equine encephalitis virus nsP2 protease". In: *Virology* 496 (2016), pp. 77–89.

[12] Naomi E Chayen and Emmanuel Saridakis. "Protein crystallization: from purified protein to diffraction-quality crystal". In: *Nature methods* 5.2 (2008), pp. 147–153.

[13] Vincent B Chen et al. "MolProbity: all-atom structure validation for macromolecular crystallography". In: *Acta Crystallographica Section D: Biological Crystallography* 66.1 (2010), pp. 12–21.

[14] Patrick M Collins et al. "Gentle, fast and effective crystal soaking by acoustic dispensing". In: *Acta Crystallographica Section D: Structural Biology* 73.3 (2017), pp. 246–255.

[15] Eric Croddy, Clarisa Perez-Armendariz, and John Hart. *Chemical and biological warfare: a comprehensive survey for the concerned citizen*. Vol. 22. Springer, 2002.
[16] Dennis E Danley. "Crystallization to obtain protein–ligand complexes for structure-aided drug design". In: *Acta Crystallographica Section D: Biological Crystallography* 62.6 (2006), pp. 569–575.
[17] Zbigniew Dauter. "Collection of X-ray diffraction data from macromolecular crystals". In: *Protein Crystallography* (2017), pp. 165–184.
[18] Marc C Deller, Leopold Kong, and Bernhard Rupp. "Protein stability: a crystallographer's perspective". In: *Acta Crystallographica Section F: Structural Biology Communications* 72.2 (2016), pp. 72–95.
[19] Zygmunt S Derewenda and Peter G Vekilov. "Entropy and surface engineering in protein crystallization". In: *Acta Crystallographica Section D: Biological Crystallography* 62.1 (2006), pp. 116–124.
[20] Moshe A Dessau and Yorgo Modis. "Protein crystallization for X-ray crystallography". In: *JoVE (Journal of Visualized Experiments)* 47 (2011), e2285.
[21] Jan Drenth. *Principles of protein X-ray crystallography*. Springer Science & Business Media, 2007.
[22] Victor RA Dubach and Albert Guskov. "The resolution in X-ray crystallography and single-particle cryogenic electron microscopy". In: *Crystals* 10.7 (2020), p. 580.
[23] SD Durbin and G Feher. "Protein crystallization". In: *Annual review of physical chemistry* 47.1 (1996), pp. 171–204.
[24] *GloMelt™ Thermal Shift Protein Stability Kit*. 2019.
[25] Xin Hu et al. "Kinetic, mutational, and structural studies of the Venezuelan equine encephalitis virus nonstructural protein 2 cysteine protease". In: *Biochemistry* 55.21 (2016), pp. 3007–3019.
[26] Dal Young Kim et al. "Venezuelan equine encephalitis virus nsP2 protein regulates packaging of the viral genome into infectious virions". In: *Journal of virology* 87.8 (2013), pp. 4202–4213.
[27] Bio-Rad Laboratories. *His-Tag Purification*. URL: `https://www.bio-rad.com/featured/en/his-tag-purification.html`. (accessed: 29.05.2022).
[28] MAX IV Laboratory. *BioMAX*. URL: `https://www.maxiv.lu.se/accelerators-beamlines/beamlines/biomax/`. (accessed: 08.05.2022).
[29] Lindsay Lundberg, Brian Carey, and Kylene Kehn-Hall. "Venezuelan equine encephalitis virus capsid—the clever caper". In: *Viruses* 9.10 (2017), p. 279.
[30] Caroline K Martin and Margaret Kielian. *Identification of human and mosquito receptors for alphaviruses*. 2022.
[31] Laurent Maveyraud and Lionel Mourey. "Protein X-ray crystallography and drug discovery". In: *Molecules* 25.5 (2020), p. 1030.
[32] Ilka Müller. "Guidelines for the successful generation of protein–ligand complex crystals". In: *Acta Crystallographica Section D: Structural Biology* 73.2 (2017), pp. 79–92.
[33] YL Nene et al. "A glimpse at viral diseases in the ancient period 1". In: *Asian Agri-Hist* 11 (2007), pp. 33–46.
[34] World Health Organization et al. *Combating emerging infectious diseases in the South-East Asia region*. Tech. rep. WHO Regional Office for South-East Asia, 2005.

[35] Harold R Powell. "X-ray data processing". In: *Bioscience reports* 37.5 (2017).
[36] Linda Reinhard et al. "Optimization of protein buffer cocktails using Thermofluor". In: *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* 69.2 (2013), pp. 209–214.
[37] Jonathan C Rupp et al. "Alphavirus RNA synthesis and non-structural protein functions". In: *The Journal of general virology* 96.Pt 9 (2015), p. 2483.
[38] Andrew T Russo, Mark A White, and Stanley J Watowich. "The crystal structure of the Venezuelan equine encephalitis alphavirus nsP2 protease". In: *Structure* 14.9 (2006), pp. 1449–1458.
[39] Andrew T Russo et al. "Structural basis for substrate specificity of alphavirus nsP2 proteases". In: *Journal of Molecular Graphics and Modelling* 29.1 (2010), pp. 46–53.
[40] Larry Schwartz. "Desalting and Buffer Exchange by Dialysis, Gel Filtration or Diafiltration". In: *Pall Scientific & Technical Report* (2003).
[41] Rajesh Babu Sekar and Ammasi Periasamy. "Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations". In: *The Journal of cell biology* 160.5 (2003), p. 629.
[42] Anuj Sharma and Barbara Knollmann-Ritschel. "Current understanding of the molecular basis of Venezuelan equine encephalitis virus pathogenesis and vaccine development". In: *Viruses* 11.2 (2019), p. 164.
[43] Gyehwa Shin et al. "Structural and functional insights into alphavirus polyprotein processing and pathogenesis". In: *Proceedings of the National Academy of Sciences* 109.41 (2012), pp. 16534–16539.
[44] MS Smyth and JHJ Martin. "x Ray crystallography". In: *Molecular Pathology* 53.1 (2000), p. 8.
[45] Marjeta Urh, Dan Simpson, and Kate Zhao. "Affinity chromatography: general methods". In: *Methods in enzymology* 463 (2009), pp. 417–438.
[46] Jimin Wang. "Estimation of the quality of refined protein crystal structures". In: *Protein Science* 24.5 (2015), pp. 661–669.
[47] Scott C Weaver et al. "Venezuelan equine encephalitis". In: *Annual Reviews in Entomology* 49.1 (2004), pp. 141–174.
[48] Lwande Wesula Olivia et al. "Global emergence of Alphaviruses that cause arthritis in humans". In: *Infection ecology & epidemiology* 5.1 (2015), p. 29853.
[49] Alexander Wlodawer et al. "Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination". In: *The FEBS journal* 280.22 (2013), pp. 5705–5736.
[50] Ruoting Yang et al. *Systems Biology*. 2nd ed. 2014, pp. 159–187. URL: `https://doi.org/10.1016/B978-0-12-411557-6.00008-2`..
[51] Huaisheng Zhang et al. "Vinyl Sulfone-Based Inhibitors of Nonstructural Protein 2 Block the Replication of Venezuelan Equine Encephalitis Virus". In: *ACS Medicinal Chemistry Letters* 11.11 (2020), pp. 2139–2145.
[52] Ren-Bin Zhou et al. "A review on recent advances for nucleants and nucleation in protein crystallization". In: *CrystEngComm* 19.8 (2017), pp. 1143–1155.

Appendix A

Buffers

During the nsP2 protease purification procedure, several buffers were used. The components of each of those are listed in Table A.1 below.

Table A.1: Buffers used for different stages in protein purification. All buffers were equilibriated to pH 7.6

Tris Buffer	**Lysis Buffer**	**Wash Buffer**	**Saltgradient Elution**
121 g Tris 1000 ml dH_2O	50 mM Tris 500 mM NaCl 1 mM TCEP 5 % Glycerol 1 spatula Lysozym	50 mM Tris 500 mM NaCl 1 mM TCEP 5 % Glycerol 60 mM Imidazole	50 mM Tris 1.25 M NaCl 1 mM TCEP 5 % Glycerol
Elution Buffer	**Dialysis Buffer**	**Low-Salt Buffer**	**Gel-Storage Buffer**
50 mM Tris 500 mM NaCl 1 mM TCEP 5 % Glycerol 300 mM Imidazole	50 mM Tris 250 mM NaCl 1 mM TCEP 5 % Glycerol 1 mM EDTA	50 mM Tris 150 mM NaCl 1 mM TCEP 5 % Glycerol	50 mM Tris 150 mM NaCl 2 mM TCEP

Appendix B

Diffraction statistics

The statistics obtained for refinement of the nsP2 protease structure model, without a ligand built in in the active site, are presented in Table B.1 below.

Table B.1: Refinement statistics for the nsP2 protease. The data collection statistics are the same as presented in Table 4.3. Statistics for the highest-resolution shell are shown in parentheses.

Refinement	
Reflections used in refinement	66319 (6561)
Reflections used for R_{free}	3297 (326)
R_{work}	0.1773 (0.2150)
R_{free}	0.1972 (0.2399)
CC(work)	0.964 (0.922)
CC(free)	0.953 (0.886)
Number of non-hydrogen atoms	2886
Number of macromolecules	2565
Number of solvent	321
Protein residues	320
RMS - bond lengths (Å)	0.005
RMS - bond angles (degrees)	0.74
Ramachandran favored (%)	96.54
Ramachandran allowed (%)	3.46
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.00
Clashscore	0.59
Average B-factor (Å^2)	21.74
B-factor (macromolecules) (Å^2)	20.39
B-factor (solvent) (Å^2)	32.59
Number of TLS groups	7